frontiers | Research Topics

Perception, cognition, and working memory: Interactions, technology, and applied research

Edited by
Hong Xu and Edwin James Burns

Published in
Frontiers in Psychology
Frontiers in Neuroscience

ISSN 1664-8714
ISBN 978-2-83251-979-0
DOI 10.3389/978-2-83251-979-0

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Perception, cognition, and working memory: Interactions, technology, and applied research

Topic editors
Hong Xu — Nanyang Technological University, Singapore
Edwin James Burns — Edge Hill University, United Kingdom

Citation
Xu, H., Burns, E. J., eds. (2023). *Perception, cognition, and working memory: Interactions, technology, and applied research.* Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83251-979-0

Table of contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 17 April 2023
DOI 10.3389/fpsyg.2023.1190194

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Bernhard Hommel,
University Hospital Carl Gustav Carus, Germany

*CORRESPONDENCE
Hong Xu
✉ xuhong@ntu.edu.sg

SPECIALTY SECTION
This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

RECEIVED 20 March 2023
ACCEPTED 31 March 2023
PUBLISHED 17 April 2023

CITATION
Xu H and Burns EJ (2023) Editorial: Perception, cognition, and working memory: interactions, technology, and applied research.
Front. Psychol. 14:1190194.
doi: 10.3389/fpsyg.2023.1190194

Editorial: Perception, cognition, and working memory: interactions, technology, and applied research

Hong Xu[1]* and Edwin James Burns[2]

[1]Psychology, School of Social Sciences, Nanyang Technological University, Singapore, Singapore,
[2]Department of Psychology, Edge Hill University, Ormskirk, United Kingdom

KEYWORDS

perception, cognition, attention, memory, action, human computer interaction

Editorial on the Research Topic
Perception, cognition, and working memory: interactions, technology, and applied research

Research questions

The research questions addressed in this Research Topic span from our senses to perception to cognition and behavior, as well as theoretical and mathematical models to explain such processes. In particular, the studies have helped answer questions in visual perception (Bernardinis et al.; Orima and Motoyoshi; Wu et al.; Vargas and Moreno-Ríos), working memory (Liu et al.; Wang A. et al.; Zacharov et al.), attention (Kandana Arachchige et al.), detection (Branch et al.), emotion (Liang et al.), memory (Guo et al.; Chén et al.; Zhou et al.; Setti et al.), auditory processing (Fan et al.), language (Wang Y. et al.; Cui et al.; Zhu and Aryadoust), cognition (Krupitzer et al.; Ren et al.; Shi), actions (Bitu et al.; Zhu et al.; Lukashova-Sanz et al.; Ren et al.), and human-machine interactions (Loriette et al.; Zhang et al.).

Methods

To address the research questions in the above fields, the studies in our Research Topic used multiple research methods, such as psychophysical experiments (Vargas and Moreno-Ríos; Shi), behavioral tasks (Bitu et al.; Lukashova-Sanz et al.), survey questionnaires (Zhu et al.) and virtual reality simulations (Krupitzer et al.; Lukashova-Sanz et al.). They also used eye tracking to record the participants' eye movement patterns (Wang Y. et al.; Zacharov et al.), and revealed the underlying neural mechanisms through EEG (Liang et al.; Orima and Motoyoshi; Cui et al.; Shi) or its connectivity by transcranial direct current stimulation (tDCS) (Wu et al.).

Theories

Given the complexity of the tasks, it is a challenge to find an overarching unified theory to explain underlying mechanisms. By contrast, existing theories often explain a narrower collection of data derived from tasks. The current issue tries therefore to collate papers that help face down such challenges, bringing them to our attention, and thus provides an outlet for such research to be published. For example, cognitive load theory (Sweller, 1988; Mayer and Moreno, 2003; Sweller et al., 2011; Bitu et al.; Zhu and Aryadoust) explains how limited amount of cognitive resources are used for a task. Flexible resource (attention) theory (Sandry et al., 2014; Sandry and Ricker, 2020; Zhou et al.) argues that attention resources are allocated flexibly to items in working memory and that their allocation to one item is at the cost of the others. These theories suggest that we have limited capacity in attention, cognitive resources, and working memory (Miller, 1956). This Research Topic also brought up a mathematical modeling for priming (Chén et al.). However, there is still a lack of theories for brain-computer interfaces (Loriette et al.), especially for the connections among sensation, perception, cognition, memory and action.

Applications

In addition to neurotypical participants, these methods are beneficial for people at a disadvantage, such as the blind and visually impaired (Setti et al.) or those with Parkinson's disease (Bernardinis et al.). Similarly, they may also help provide practical support to those with autism spectrum disorder (Zacharov et al.). They cover studies that tested children (Zacharov et al.; Bitu et al.), soccer players (Ren et al.; Krupitzer et al.), college students (Ren et al.), adults (e.g., Branch et al.) and older adults (Zhu et al.). Setti et al. showed that the blind participants exhibited impaired performance in a spatial memory task of sound locations, adapted from the card game "Memory," thus confirming "the pivotal role of visual experience in the active manipulation of memorized spatial information." This suggests that auditory spatial memory benefited from visual integration. While auditory spatial memory of the location is affected by visual experience, it might also be interesting to test participants' responses to tasks that do not require any visual cues of the space, and to examine if the increased sensitivity of one sensory modality is at the cost of the other(s) or enhanced by the other(s).

Future research

Complex tasks and concepts can often be dissected into distinct steps that occur sequentially or simultaneously. Based on existing theories of limited cognitive resources, the former may be more reasonable, but it also depends on the time scale of this analysis. In a fine grained time-scale, such as milliseconds, it is more likely to be sequential than simultaneously. Different research methods may be able to address the tasks from various perspectives. Therefore, future research may call into the integration of these perspectives, and the underlying mechanisms and theories.

Author contributions

HX drafted the manuscript. EB provided critical revisions. Both authors approved the final version of the manuscript for submission.

Funding

This research was supported by the National Research Foundation, Singapore and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-RP-2020-019) and School of Social Sciences, Nanyang Technological University, Singapore.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Mayer, R. E., and Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educ. Psychol.* 38, 43–52. doi: 10.1207/S15326985EP3801_6

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97. doi: 10.1037/h0043158

Sandry, J., and Ricker, T. J. (2020). Prioritization within visual working memory reflects a flexible focus of attention. *Attent. Percept. Psychophys.* 82, 2985–3004. doi: 10.3758/s13414-020-02049-4

Sandry, J., Schwark, J. D., and Macdonald, J. (2014). Flexibility within working memory and the focus of attention for sequential verbal information does not depend on active maintenance. *Mem. Cogn.* 42, 1130–1142. doi: 10.3758/s13421-014-0422-1

Sweller, J (1988). Cognitive load during problem solving: effects on learning. *Cogn. Sci.* 12, 257–285. doi: 10.1207/s15516709cog1202_4

Sweller, J., Ayres, P., and Kalyuga, S. (2011). *Cognitive Load Theory*. New-York, NY: Springer. doi: 10.1007/978-1-4419-8126-4

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 17 June 2021
doi: 10.3389/fpsyg.2021.692785

Sensitivity to a Break in Interaural Correlation in Frequency-Gliding Noises

Langchen Fan[1,2†], Lingzhi Kong[3†], Liang Li[1,2] and Tianshu Qu[2]*

[1]Beijing Key Laboratory of Behavior and Mental Health, School of Psychological and Cognitive Sciences, Peking University, Beijing, China, [2]Key Laboratory on Machine Perception (Ministry of Education), Department of Machine Intelligence, Peking University, Beijing, China, [3]Language Pathology and Brain Science MEG Lab, School of Communication Sciences, Beijing Language and Culture University, Beijing, China

OPEN ACCESS

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Brian Gygi,
Veterans Affairs Northern California
Health Care System (VANCHCS),
United States
Pekcan Ungan,
The Science Academy Society of
Turkey, Turkey

***Correspondence:**
Tianshu Qu
qutianshu@pku.edu.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

Received: *09 April 2021*
Accepted: *25 May 2021*
Published: *17 June 2021*

Citation:
Fan L, Kong L, Li L and Qu T (2021)
Sensitivity to a Break in
Interaural Correlation in
Frequency-Gliding Noises.
Front. Psychol. 12:692785.
doi: 10.3389/fpsyg.2021.692785

This study was to investigate whether human listeners are able to detect a binaurally uncorrelated arbitrary-noise fragment embedded in binaurally identical arbitrary-noise markers [a break in correlation, break in interaural correlation (BIAC)] in either frequency-constant (frequency-steady) or frequency-varied (unidirectionally frequency gliding) noise. Ten participants with normal hearing were tested in Experiment 1 for up-gliding, down-gliding, and frequency-steady noises. Twenty-one participants with normal hearing were tested in Experiment 2a for both up-gliding and frequency-steady noises. Another nineteen participants with normal hearing were tested in Experiment 2b for both down-gliding and frequency-steady noises. Listeners were able to detect a BIAC in the frequency-steady noise (center frequency = 400 Hz) and two types of frequency-gliding noises (center frequency: between 100 and 1,600 Hz). The duration threshold for detecting the BIAC in frequency-gliding noises was significantly longer than that in the frequency-steady noise (Experiment 1), and the longest interaural delay at which a duration-fixed BIAC (200 ms) in frequency-gliding noises could be detected was significantly shorter than that in the frequency-steady noise (Experiment 2). Although human listeners can detect a BIAC in frequency-gliding noises, their sensitivity to a BIAC in frequency-gliding noises is much lower than that in frequency-steady noise.

Keywords: auditory system, binaural hearing, center frequency, interaural correlation, frequency gliding, interaural delay

INTRODUCTION

The auditory system usually implicates functions of two ears, integrating the sound information from both ears. The binaural hearing has been recognized as a critical function of the central auditory system, offering substantial advantages in localizing sounds, dealing with reflections, and improving speech recognition in adverse environments (Kohlrausch et al., 2013).

Interaural coherence (the degree of similarity of the sound waveforms at the two ears) can be physically measured as the maximum value of the cross correlation between the sound wave at the left ear and the sound wave at the right ear when one of the two sounds has been time shifted (within limits, e.g., ±1 or ±2 ms) to maximize the correlation (Grantham, 1995;

Aaronson and Hartmann, 2010), which is called "interaural correlation." If the sound wave at the left ear is an identical copy of the wave at the right ear, the interaural correlation is one. In contrast, if the sound waves at the left and right ears are independently generated, the interaural correlation is near to zero. The interaural correlation can be represented at both the neurophysiological level (Wang et al., 2018) and the perceptual level (Blauert and Lindemann, 1986). When sounds, i.e., arbitrary noises, arrive at the two ears simultaneously, identical sounds (interaural correlation = 1) at the two ears are perceptually fused into a single image at the center area of the head, while binaurally independent sounds are perceived as two separated sound images at each ear (Blauert and Lindemann, 1986).

Several previous studies have shown that human listeners are able to discriminate changes in the interaural correlation across two binaural noises. Particularly, the discrimination was extremely sensitive to a slight drop in the interaural correlation from binaurally identical noise (with an interaural correlation of one; Pollack and Trittipoe, 1959; Gabriel and Colburn, 1981; Akeroyd and Summerfield, 1999; Culling et al., 2001; Boehnke et al., 2002; Chait et al., 2005). Furthermore, the sensitivity to the dynamic change in interaural correlation has been investigated using a binaural analog of the gap-detection paradigm by placing a binaurally uncorrelated fragment, i.e., a break in interaural correlation (BIAC; a pair of binaurally independent noises), in the temporal center of two bursts of binaurally identical noise (markers: Akeroyd and Summerfield, 1999; Boehnke et al., 2002). Introducing a BIAC does not alter the energy or spectrum of the arbitrary noise but modifies the auditory images, including the perceptual compactness/diffuseness of the noise image (Blauert and Lindemann, 1986; Edmonds and Culling, 2009). The duration threshold (the minimum duration required to detect a BIAC) is measured to determine the sensitivity to a dynamic change in interaural correlation. Previous studies have proved that human listeners are sensitive to a BIAC in either broad-band or narrow-band noise whose spectral information does not vary with time monaurally (Akeroyd and Summerfield, 1999; Boehnke et al., 2002).

Moreover, understanding of the interaural correlation processing is incomplete without considering the impact of the interaural delay. As the interaural delay increases from zero, the perceptually fused single auditory image of binaurally identical noise initially moves toward the leading ear, then becoming increasingly diffuse and eventually indistinguishable from the sound image of the binaurally independent noise (Blodgett et al., 1956). Our previous studies have shown that the sensitivity to a BIAC decreased dramatically as the interaural delay increased from zero to several milliseconds, and the maximum interaural delay, at which a BIAC can be detected (the delay threshold), has been used to determine the impact of the time delay between the sounds at the two ears on the sensitivity to a change in interaural correlation (Huang et al., 2008, 2009a,b, 2019; Li et al., 2009, 2013; Kong et al., 2012, 2015; Qu et al., 2013).

Ecologically, communication sounds with time-varying spectra are common for humans and other species. For example, the frequency modulation is a fundamentally acoustic component in human speech, critical to the discrimination of vowels (Jenkins et al., 1983), the recognition of Mandarin tones (Kong and Zeng, 2004), and the speech recognition in noise (Zeng et al., 2005).

Moreover, it has been shown that the auditory system is sensitive to the binaural cues even in frequency-gliding tone (frequency range: 3–8 kHz; Hsieh and Saberi, 2009). To our knowledge, however, the issue of the sensitivity to changes in interaural correlation in frequency-gliding sound has not been reported.

Previous binaural models share a fundamental notion that binaural performance, e.g., interaural correlation processing, is based on frequency-band-by-frequency-band comparisons of bandpass-filtered signals at two ears (Stern and Trahiotis, 1995; Ungan et al., 2019). The frequency selectivity of binaural processing is not necessarily poorer than that for monaural processing (Verhey and van de Par, 2018), since it has been shown that the auditory system is capable of integrating binaural information across different frequency channels (Jain et al., 1991; Hsieh and Saberi, 2009). Thus, we hypothesized that human listeners can hear a dynamic change in interaural correlation in noises with center frequency varying unidirectionally when both the spectral and temporal integrations are involved.

MATERIALS AND METHODS

Participants

All participants were young adult university students at the Peking University. They had pure-tone thresholds no higher than 25 dB HL between 0.125 and 8 kHz, and the threshold difference between the two ears at each testing frequency was less than 15 dB HL. They gave written informed consent and were paid a modest stipend for their participation. All the experimental procedures were approved by the Committee for Protecting Human and Animal Subjects in the School of Psychological and Cognitive Sciences at Peking University.

Ten participants (eight females, 18–26 years old, mean age = 20.5 years) took part in Experiment 1. Twenty-one different participants (15 females, 17–24 years old, mean age = 19.1 years) were tested in Experiment 2a. Another group of the participants (13 females, 18–27 years old, mean age = 20.7 years), who did not participate in Experiment 1 and Experiment 2a, were tested in Experiment 2b.

Apparatus and Stimuli

The participant was seated in a chair at the center of a sound-attenuated chamber (EMI Shielded Audiometric Examination Acoustic Suite). Frequency-steady and frequency-gliding noises (sampling rate = 16 kHz; duration = 2,000 ms; rise/decay time = 50 ms) were synthesized using MATLAB (the MathWorks Inc., Natick, MA, United States).

To produce frequency-steady noise, Gaussian wideband noise (0–8 kHz) was generated and bandpass filtered (the 400-Hz geometric center frequency with a bandwidth of 1.585 octave).

To produce frequency-gliding noises, the wideband noises were cut into temporal frames using a Hanning window. The frame length was 62.5 ms (duration of on/off ramps = 31.25 ms) and the frameshift time was 15.6 ms (with an overlap between successive frames). The energy of each of the frames was set to a fixed value. These wide-band noise frames were filtered into narrow-band noise frames by a 1.585-octave wide bandpass filter in the frequency domain, and the frequency components outside the passband were set to zero. Each narrow-band frame had a center frequency in the range from 100 to 1,600 Hz (log spaced). The sequence of the center frequencies was from 100 to 1,600 Hz or from 1,600 to 100 Hz for the up-gliding and down-gliding noises, respectively. Note that all the frames were concatenated together by overlap and sum method, and played out. **Figure 1** shows spectrograms of up-gliding noise (left), frequency-steady noise (middle), and down-gliding (right) noise.

The BIAC was always in the temporal center of the noise. For example, to insert a BIAC with a 200-ms duration, the noise section between 900 and 1,100 ms (from the noise onset) in the left-ear channel was substituted by an interaurally independent segment (interaural correlation = 0) with the same parameters. For frequency-gliding noises, the center frequency of the BIAC in up-gliding noise changed from 348 to 459 Hz and the center frequency of the BIAC in down-gliding noise changed from 459 to 347 Hz during the 200-ms BIAC. The center frequency was always 400 Hz, 1,000 ms after the noise onset.

In Experiment 1, the duration of BIAC varied while the overall duration of the noise stimuli was kept at 2,000 ms. The minimum duration required to detect a BIAC (duration threshold) was examined using the frequency-steady noise, up-gliding noise, and down-gliding noise. In Experiment 2, the duration of the BIAC was fixed at 200 ms. When an interaural delay was introduced, a quiet segment with a duration equal to the interaural delay was added to the beginning of the stimulus for the right ear and the end of the stimulus for the left ear. The maximum interaural delay at which the 200-ms BIAC could be detected (delay threshold) was tested for both up-gliding noise and frequency-steady noise in Experiment 2a, and the delay threshold was determined for both down-gliding noise and frequency-steady noise in Experiment 2b.

Sound stimuli were generated using a Creative Sound Blaster PCI128 (Creative SB Audigy 2 ZS, Creative Technology Ltd., Singapore) and delivered by headphones (HD 265 linear, Sennheiser, Germany). The sound intensity was calibrated using a Larson Davis Audiometer Calibration and Electroacoustic Testing System (AUDit and System 824, Larson Davis, Depew, NY, United States). The overall sound level was 63 dB SPL.

Design and Procedure

The BIAC was perceived as a "central-to-diffuse" change in the noise. The percepts of the BIAC in frequency-gliding were similar to those embedded in frequency-steady noise, except the frequency-gliding noise has a continuous pitch gliding. Note that any auditory event coinciding in time with the BIAC could not be detected when only noise at one ear was delivered. A brief training session was used before Experiment 1 and Experiment 2a and 2b to ensure that each participant understood the instructions and was able to detect the BIAC in each of the three noise types, especially in frequency-gliding noises.

In Experiment 1, the duration threshold for detecting the BIAC was measured for each of the noise types using adaptive two-interval, two-alternative, and forced-choice procedures. In each trial, the BIAC was randomly assigned to one of the two intervals, which were separated by 1,000 ms. The participants' task was to detect an auditory change in the middle of the noises and identify which of the two intervals contained the change by pressing the left or right button on a response box. The BIAC duration was set to 65 ms at the beginning and manipulated using a three-down one-up procedure: The duration was decreased after three consecutive correct responses and increased after one incorrect response. The initial size of the change in the duration of the BIAC was 16 ms, and the step size was altered by a factor of 0.5 with each reversal in direction of duration change until the minimum value of 1 ms was reached. Feedback was given visually after each trial *via* a LCD monitor in front of the participant. Each adaptive procedure (i.e., a run) was terminated after 10 reversals, and the duration threshold for a run was defined as the arithmetic mean duration across the last six reversals. For each noise type, the arithmetic

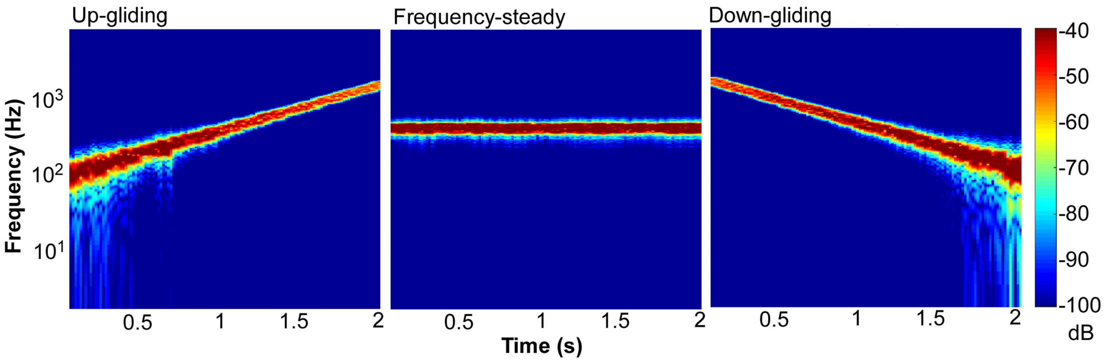


FIGURE 1 | Spectrograms of up-gliding, frequency-steady, and down-gliding noises.

mean of the duration thresholds for three runs was taken as the duration threshold.

In Experiment 2, the delay threshold for detecting the BIAC was measured using a similar procedure to that for Experiment 1, except that the BIAC duration was fixed at 200 ms. The interaural delay systematically varied in Experiment 2 and was at the beginning set to 0 ms, which is the easiest condition for a listener to detect the BIAC. The interaural delay was increased after three consecutive correct responses and decreased after one incorrect response. The step size started at 16 ms and decreased by half for each reversal until it reached 1 ms. Each adaptive procedure (i.e., a run) was terminated after 10 reversals, and the delay threshold for a run was defined as the arithmetic mean interaural delay across the last six reversals. For each noise type, the arithmetic mean of the delay thresholds for three runs was taken as the delay threshold. A brief training session was also provided before the experiment.

RESULTS

Experiment 1

The duration thresholds for detecting the BIAC were obtained from 10 participants for each of the noise types when the noise at one ear was delivered simultaneously with that delivered on the other ear. Our results clearly showed that listeners were able to detect a dynamic change in interaural correlation for binaural noises with the unidirectionally varied center frequency. **Figure 2** shows the group-mean duration thresholds and standard errors of the mean for each noise type. An important feature was that the binaurally uncorrelated fragments embedded in frequency-gliding noises were much harder to detect than that embedded in frequency-steady noise. The mean thresholds for detecting a BIAC for up-gliding noise, down-gliding noise, and frequency-steady noise were 48.4 ms, 34.4 ms, and 4.7 ms, respectively.

ANOVA across the three conditions of the noise type was performed to determine whether the duration threshold for frequency-gliding noises was much longer than that for frequency-steady noise. The ANOVA showed that the main effect of noise type on the duration threshold was significant [F (2,18) = 7.152, $p < 0.01$]. LSD *post-hoc* analyses showed that the duration threshold for detecting the BIAC in frequency-steady noise was significantly shorter than that in up-gliding noise ($p < 0.01$) and down-gliding noise ($p < 0.01$). Moreover, the duration threshold in up-gliding noise was not significantly different from that in down-gliding noise ($p = 0.354$).

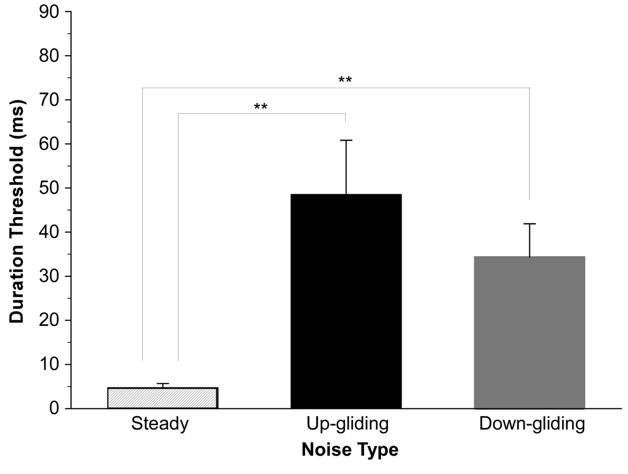


FIGURE 2 | Group-mean duration thresholds for detecting a break in interaural correlation (BIAC) in three types of noises: frequency-steady noise, up-gliding noise, and down-gliding noise (Experiment 1). The error bars represent the standard errors of the means (SEM). **$p < 0.01$.

Experiment 2

Among the duration thresholds obtained in Experiment 1, when the interaural delay was zero, the longest duration threshold for up-gliding noise was 119.4 ms and the longest duration threshold for down-gliding noise was 92.9 ms. Thus, it is reasonable to predict that most human listeners are able to detect a 200-ms BIAC in the frequency-gliding noises when the interaural delay is zero. In Experiment 2a, the longest interaural delays at which a 200-ms BIAC could be detected (delay thresholds) were obtained from 21 participants for up-gliding noise and frequency-steady noise. The delay thresholds for down-gliding noise and frequency-steady noise were obtained from another 19 participants in Experiment 2b.

In consistent with the results in Experiment 1 that the BIAC in frequency-gliding noises was much harder to detect than that in frequency-steady noise, the maximum interaural delay for detecting the BIAC in frequency-gliding noises was shorter than that for frequency-steady noise. The group-mean delay threshold for detecting the BIAC was 7.3 ms for up-gliding noise and 9.2 ms for down-gliding noise while that for frequency-steady noise was 12.0 ms in Experiment 2a and 12.1 ms in Experiment 2b. **Figure 3** shows the group-mean delay thresholds

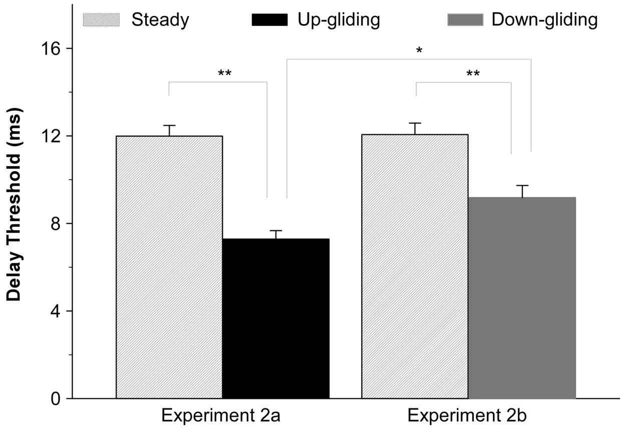


FIGURE 3 | Group-mean interaural delay thresholds for detecting a BIAC in for three types of noises: frequency-steady noise, up-gliding noise, and down-gliding noise (Experiment 2). The error bars represent the SEM. *$p < 0.05$, **$p < 0.01$.

for up-gliding noise and the frequency-steady noise in Experiment 2a, and those for down-gliding noise and frequency-steady noise in Experiment 2b.

A paired t-test was performed to determine if the delay threshold for up-gliding noise was significantly shorter than that for the frequency-steady noise. The paired t-test showed that the difference was significant ($t = -9.895$, $p < 0.001$, Experiment 2a). Similarly, a paired t-test showed that the delay threshold for the down-gliding noise was also significantly shorter than that for frequency-steady noise ($t = -5.846$, $p < 0.001$, Experiment 2b). In contrast to the comparison between the duration threshold for up-gliding noise and that for down-gliding noise, a non-matched samples *t*-test showed that the delay threshold for the up-gliding noise was significantly shorter than that for down-gliding noise ($t = -2.859$, $p < 0.01$).

DISCUSSION

The primary aim of our study was to determine the sensitivity to a change in interaural correlation when the center frequency of binaural noises varied over time (frequency-gliding noises). The results of this study showed that young adults with normal hearing were able to detect a BIAC embedded in frequency-gliding noises (center frequency: between 100 and 1,600 Hz). However, the duration threshold for frequency-gliding noises was significantly longer than that for frequency-steady noise (center frequency: 400 Hz).

As the detection of a BIAC is determined by the perceptual contrast in the interaural correlation between the uncorrelated segment and the marker (the noise sections flanking the BIAC), the detection difficulty in frequency-gliding noises might be based on the possible decrease in the perceptual contrast between the BIAC and marker in frequency-gliding noises. Given that most models of binaural processing are based on the assumption that sounds are filtered into narrow-band signals and the processing of the binaural information is based on comparisons of interaural differences in a band-by-band manner (Durlach et al., 1986; Stern and Trahiotis, 1995; Akeroyd and Summerfield, 1999), the detection of the BIAC for frequency-gliding noises should be based on both the processing of interaural correlation within the frequency band where the BIAC embedded and the across-band information from the other frequency bands which frequency-gliding noises passed through.

For the processing of interaural correlation within the frequency band where a BIAC embedded (center frequency: 400 Hz), the detection of the BIAC may have been influenced by forward and backward masking from the marker. The duration of the forward fringe for frequency-steady noise would be 997.65 ms, based on the duration of the whole noise, and the mean duration threshold for frequency-steady noise is 4.7 ms. For frequency-gliding noises, however, the effective duration of the marker noise within each frequency band would be affected by the speed of the sweep and the bandwidth of the auditory filters. According to the auditory filter bandwidth of Glasberg and Moore (1990), the bandwidth [equivalent rectangular bandwidth (ERB)], the frequency range of the frequency band (center frequency = 400 Hz), was calculated:

$$ERB(f) = 0.108f + 24.7 \tag{1}$$

where f is the center frequency. Based on the frequency range of the band (366 Hz–434 Hz) and frequency-gliding noises used in this study, the overall duration of frequency-gliding noises in the frequency band (center frequency = 400 Hz) was 125 ms. If the mean duration threshold for frequency-gliding noises is used to estimate the duration of the forward fringe in the frequency band centered on 400 Hz, the duration of the forward fringe is 38 ms for up-gliding noise and 45.3 ms for down-gliding noise. Although no prior study has assessed the effect of forward fringe duration on the detection of a BIAC, the discrimination between binaural noises with different interaural correlation was virtually impossible for durations of 10 and 32 ms (Pollack and Trittipoe, 1959). Although it is possible that the listener can detect the BIAC when only the frequency band where the BIAC embedded was monitored, the detection of BIAC is extremely hard based on the output of binaural processing from the frequency band with the center frequency of 400 Hz according to the findings of Pollack and Trittipoe (1959).

Comparisons of binaural information across frequency play an important role in the binaural lateralization of bandpass noises (Stern and Trahiotis, 1995; Ungan et al., 2019). For example, the interaural time difference of a bandpass noise which is consistent over frequency has been found to be the true interaural time difference of the stimuli (straightness; Stern and Trahiotis, 1995). The detection of the BIAC in frequency-gliding noises probably needs to integrate over a wider frequency range where more binaural information of the marker noises (interaural correlation = 1) is provided than the single frequency band centered on 400 Hz. Around the frequency band with a center frequency of 400 Hz, six frequency bands for the frequency from 82 to 366 Hz and eight frequency bands for the frequency from 434 to 1,724 Hz were included according to the auditory filter bandwidth of Glasberg and Moore (1990). The effective duration of noises in different frequency bands ranged from 78 to 250 ms. It has been shown that the percentage of correct discrimination between binaural noises with an interaural correlation of 0.998 and reference noises with an interaural correlation of 0.922 decreased from 100 to 60% as the noise duration decreased from 316 to 32 ms (Pollack and Trittipoe, 1959). It is speculated that the interaural correlation processing of the marker noises for frequency-gliding noises might be affected by the relatively short duration of marker noises in individual frequency bands.

Our results showed that participants were able to detect a BIAC in frequency-gliding noises even when an interaural delay of several milliseconds was introduced. The human auditory system is able to process binaural cues with interaural delays much longer than those experienced in free-field listening which is usually less than 600 microseconds (Blodgett et al., 1956). Our previous studies have also shown that human listeners can detect a BIAC at larger interaural delays than those experienced in free-field listening for broad-band or narrow-band noises (Huang et al., 2009a,b; Li et al., 2009; Kong et al., 2012, 2015). In the present study, the delay threshold for detecting a fixed-duration BIAC (200 ms) in frequency-gliding noises was significantly shorter than in frequency-steady noise. Fine-structure signals from the leading ear must be maintained (or delayed) for several milliseconds to allow interaural processing of binaural noises with large interaural delays and the maintained information progressively decays as the interaural delay is increased (Huang et al., 2009a; Li et al., 2013). Consistent with the difficulty in detecting a brief BIAC in frequency-gliding noises with no interaural delay, it appears that the maintenance of the fine-structure information for frequency-gliding noises is harder than that for frequency-steady noise.

It has been widely accepted that binaural responses are temporally sluggish when compared with monaural responses (Grantham, 1995; Akeroyd and Summerfield, 1999). The binaural sluggishness is supported by the temporal-window theory that the duration of the binaural temporal window is shown to be significantly longer than that of the monaural temporal window (Akeroyd and Summerfield, 1999). Given the sensitivity to a dynamic change in interaural correlation over time is affected by the binaural temporal window, one possible explanation for the difficulty in detecting the BIAC in frequency-gliding noises might be the increase in the duration of the binaural temporal window for frequency-gliding noises. It has been proved that the processing of interaural correlation makes it harder to perceive the temporal changes in the frequency (Krumbholz et al., 2009). However, whether the binaural temporal window for frequency-gliding noises is broader than that for frequency-steady noise cannot be determined until the just noticeable difference of the interaural correlation for frequency-gliding noises is tested in further studies.

Many sounds in natural environments have frequency modulations, e.g., speech and other communication sounds (Hsieh and Saberi, 2009). Investigation of the interaural correlation processing of sounds with frequency modulations should lead to a better understanding of the mechanism underlying their spatial coding and recognition against a noisy background. For example, considering the detection of a target sound, i.e., speech, against a noisy background when both the target and noise are delivered binaurally through headphones, the detection performance is significantly improved by inverting either the target signal wave or the masking noise wave in one ear (Licklider, 1948). This binaural unmasking effect is suggested to be closely related to the sensitivity to a change in interaural coherence (the degree of similarity of the sound waveforms at the two ears; Durlach et al., 1986). However, the interaural correlation of the stimulus has been found to be a poor predictor of this binaural unmasking effect (van der Heijden and Joris, 2010). Thus, further studies need to be performed to investigate the functional relationship between the interaural correlation of frequency-gliding noises and auditory target detection in noise.

In addition, interaural correlation processing is based on the neural processing of the temporal fine structures (Huang et al., 2009a; Li et al., 2013) which are vulnerable to diseases with auditory neural degeneration, e.g., Alzheimer's disease (Sinha et al., 1993). Several previous studies have found that the spatial coding and auditory scene analysis were impaired in patients with Alzheimer's disease (Goll et al., 2012; Golden et al., 2015). The interaural correlation processing of frequency-gliding noises in this study may have clinical significance as a manifestation of the prodromal stage of Alzheimer's disease.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Committee for Protecting Human and Animal Subjects in the School of Psychological and Cognitive Sciences at Peking University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LF conducted experiments. LK conducted data analysis and manuscript writing. LL contributed to the writing of the manuscript. TQ conducted sound generation and supervised this research project. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Program under grant number 2019YFC1408501; the National Natural Sciences Foundation of China under grant numbers 31771252 and 31800962; the Science Foundation of Beijing Language and Cultural University (the Fundamental Research Funds for the Central Universities) under grant number 17YBB20; and the State Key Laboratory of Robotics under grant number 2018-O09 and the High Performance Computing Platform of Peking University.

REFERENCES

Aaronson, N. L., and Hartmann, W. M. (2010). Interaural coherence for noise bands: waveforms and envelopes. *J. Acoust. Soc. Am.* 127, 1367–1372. doi: 10.1121/1.3290991

Akeroyd, M. A., and Summerfield, A. Q. (1999). A binaural analog of gap detection. *J. Acoust. Soc. Am.* 105, 2807–2820. doi: 10.1121/1.426897

Blauert, J., and Lindemann, W. (1986). Spatial mapping of intracranial auditory events for various degrees of interaural coherence. *J. Acoust. Soc. Am.* 79, 806–813. doi: 10.1121/1.393471

Blodgett, H. C., Wilbanks, W. A., and Jeffress, L. A. (1956). Effect of large interaural time differences upon the judgment of sidedness. *J. Acoust. Soc. Am.* 28, 639–643. doi: 10.1121/1.1908430

Boehnke, S. E., Hall, S. E., and Marquardt, T. (2002). Detection of static and dynamic changes in interaural correlation. *J. Acoust. Soc. Am.* 112, 1617–1626. doi: 10.1121/1.1504857

Chait, M., Poeppel, D., Cheveigné, A., and Simon, J. Z. (2005). Human auditory cortical processing of changes in interaural correlation. *J. Neurosci.* 25, 8518–8527. doi: 10.1523/JNEUROSCI.1266-05.2005

Culling, J. F., Colburn, H. S., and Spurchise, M. (2001). Interaural correlation sensitivity. *J. Acoust. Soc. Am.* 110, 1020–1029. doi: 10.1121/1.1383296

Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (1986). Interaural correlation discrimination: II. Relation to binaural unmasking. *J. Acoust. Soc. Am.* 79, 1548–1557. doi: 10.1121/1.393681

Edmonds, B., and Culling, J. F. (2009). Interaural correlation and the binaural summation of loudness. *J. Acoust. Soc. Am.* 125, 3865–3870. doi: 10.1121/1.3120412

Gabriel, K. J., and Colburn, H. S. (1981). Interaural correlation discrimination: I. Bandwidth and level dependence. *J. Acoust. Soc. Am.* 69, 1394–1401. doi: 10.1121/1.385821

Glasberg, B. R., and Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* 47, 103–138. doi: 10.1016/0378-5955(90)90170-t

Golden, H. L., Nicholas, J. M., Yong, K. X. X., Downey, L. E., Schott, J. M., Mummery, C. J., et al. (2015). Auditory spatial processing in Alzheimer's disease. *Brain* 138, 189–202. doi: 10.1093/brain/awu337

Goll, J. C., Kim, L. G., Ridgway, G. R., Hailstone, J. C., Lehmann, M., Buckley, A. H., et al. (2012). Impairments of auditory scene analysis in Alzheimer's disease. *Brain* 135, 190–200. doi: 10.1093/brain/awr260

Grantham, D. W. (1995). "Spatial hearing and related phenomena," in *Hearing*. ed. B. C. J. Moore (London: Academic), 297–345.

Hsieh, I. H., and Saberi, K. (2009). Detection of spatial cues in linear and logarithmic frequency-modulated sweeps. *Atten. Percept. Psychophysiol.* 71, 1876–1889. doi: 10.3758/APP.71.8.1876

Huang, Y., Huang, Q., Chen, X., Wu, X., and Li, L. (2009a). Transient auditory storage of acoustic details is associated with release of speech from informational masking in reverberant conditions. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1618–1628. doi: 10.1037/a0015791

Huang, Y., Kong, L., Fan, S., Wu, X., and Li, L. (2008). Both frequency and interaural delay affect event-related potential responses to binaural gap. *Neuro Report* 19, 1673–1678. doi: 10.1097/WNR.0b013e32831576c7

Huang, Y., Lu, H., and Li, L. (2019). Human scalp evoked potentials related to the fusion between a sound source and its simulated reflection. *PLoS One* 14:e0209173. doi: 10.1371/journal.pone.0209173.g008

Huang, Y., Wu, X., and Li, L. (2009b). Detection of the break in interaural correlation is affected by interaural delay, aging, and center frequency. *J. Acoust. Soc. Am.* 126, 300–309. doi: 10.1121/1.3147504

Jain, M., Gallagher, D. T., Koehnke, J., and Colburn, H. S. (1991). Fringed correlation discrimination and binaural detection. *J. Acoust. Soc. Am.* 90, 1918–1926. doi: 10.1121/1.401671

Jenkins, J. J., Strange, W., and Edman, T. R. (1983). Identification of vowels in "vowelless" syllables. *Percept. Psychophys.* 34, 441–450. doi: 10.3758/bf03203059

Kohlrausch, A., Braasch, J., Kolossa, D., and Blauert, J. (2013). "An introduction to binaural processing," in *The Technology of Binaural Listening*. ed. J. Blauert (Berlin, Heidelberg: Springer), 1–32.

Kong, L., Xie, Z., Lu, L., Qu, T., Wu, X., Yan, J., et al. (2015). Similar impacts of the interaural delay and interaural correlation on binaural gap detection. *PLoS One* 10:e0126342. doi: 10.1371/journal.pone.0126342

Kong, L., Xie, Z., Lu, L., Wu, X., and Li, L. (2012). Sensitivity to a break in interaural correlation is co-modulated by intensity level and interaural delay. *J. Acoust. Soc. Am.* 132, 114–118. doi: 10.1121/1.4734241

Kong, Y., and Zeng, F. (2004). Temporal and spectral cues in mandarin tone recognition. *J. Acoust. Soc. Am.* 120, 2830–2840. doi: 10.1121/1.2346009

Krumbholz, K., Magezi, D. A., Moore, R. C., and Patterson, R. D. (2009). Binaural sluggishness precludes temporal pitch processing based on envelope cues in conditions of binaural unmasking. *J. Acoust. Soc. Am.* 125, 1067–1074. doi: 10.1121/1.3056557

Li, L., Huang, J., Wu, X. H., Qi, J. G., and Schneider, B. A. (2009). The effects of aging and interaural delay on the detection of a break in the interaural correlation between two sounds. *Ear Hear.* 30, 273–286. doi: 10.1097/AUD.0b013e318198703d

Li, H., Kong, L., Wu, X., and Li, L. (2013). Primitive auditory memory is correlated with spatial unmasking that is based on direct-reflection integration. *PLoS One* 8:e63106. doi: 10.1371/journal.pone.0063106

Licklider, J. C. R. (1948). The influence of interaural phase relations upon the masking of speech by white noise. *J. Acoust. Soc. Am.* 20, 150–159. doi: 10.1121/1.1906358

Pollack, I., and Trittipoe, W. J. (1959). Interaural noise correlations: examination of variables. *J. Acoust. Soc. Am.* 31, 1250–1252. doi: 10.1121/1.1907669

Qu, T., Cao, S., Chen, X., Huang, Y., Li, L., Wu, X., et al. (2013). Aging effects on detection of spectral changes induced by a break in sound correlation. *Ear Hear.* 34, 280–287. doi: 10.1097/AUD.0b013e31826e4fe1

Sinha, U. K., Hollen, K. M., Rodriguez, R., and Miller, C. A. (1993). Auditory system degeneration in Alzheimer's disease. *Neurology* 43, 779–785. doi: 10.1212/wnl.43.4.779

Stern, R. M., and Trahiotis, C. (1995). "Models of binaural interaction," in *Handbook of Perception and Cognition—Hearing*. ed. B. C. J. Moore (San Diego: Academic), 347–386.

Ungan, P., Yagcioglu, S., and Ayik, E. (2019). Event-related potentials to single-cycle binaural beats of a pure tone, a click train, and a noise. *Exp. Brain Res.* 237, 2811–2828. doi: 10.1007/s00221-019-05638-4

van der Heijden, M., and Joris, P. X. (2010). Interaural correlation fails to account for detection in a classic binaural task: dynamic ITDs dominate N0Sπ detection. *JARO* 11, 113–131. doi: 10.1007/s10162-009-0185-8

Verhey, J. L., and van de Par, S. (2018). Binaural frequency selectivity in humans. *Eur. J. Neurosci.* 51, 1179–1190. doi: 10.1111/ejn.13837

Wang, Q., Lu, H., Wu, Z., and Li, L. (2018). Neural representation of interaural correlation in human auditory brainstem: comparisons between temporal-fine structure and envelope. *Hear. Res.* 365, 165–173. doi: 10.1016/j.heares.2018.05.015

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., et al. (2005). Speech recognition with amplitude and frequency modulations. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2293–2298. doi: 10.1073/pnas.0406460102

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 26 July 2021
doi: 10.3389/fnins.2021.676469

Abnormal Vision-Based Displacement Perception in Parkinson's Disease

Matthew Bernardinis[1,2,3]*, S. Farokh Atashzar[4]*, Rajni V. Patel[1,2,5†] and Mandar S. Jog[1,3,5†]

[1] School of Biomedical Engineering, Western University, London, ON, Canada, [2] Canadian Surgical Technologies and Advanced Robotics, London Health Sciences Centre (LHSC), London, ON, Canada, [3] London Movement Disorders Centre, LHSC, London, ON, Canada, [4] Electrical and Computer Engineering, Mechanical and Aerospace Engineering, New York University (NYU), New York, NY, United States, [5] Department of Clinical Neurological Sciences and Electrical and Computer Engineering, Western University, London, ON, Canada

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Abbas F. Sadikot,
McGill University, Canada
Jane Alty,
University of Tasmania, Australia

***Correspondence:**
Matthew Bernardinis
mberna26@uwo.ca
S. Farokh Atashzar
f.atashzar@nyu.edu

†These authors share senior authorship

Specialty section:
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

Received: 05 March 2021
Accepted: 02 July 2021
Published: 26 July 2021

Citation:
Bernardinis M, Atashzar SF,
Patel RV and Jog MS (2021)
Abnormal Vision-Based Displacement
Perception in Parkinson's Disease.
Front. Neurosci. 15:676469.
doi: 10.3389/fnins.2021.676469

In this work, we investigate the effect of Parkinson's disease (PD), and common corresponding therapies on vision-based perception of motion, a critical perceptual ability required for performing a wide range of activities of daily livings. While PD has been recognized as mainly a motor disorder, sensory manifestation of PD can also play a major role in the resulting disability. In this paper, for the first time, the effect of disease duration and common therapies on vision-based perception of displacement were investigated. The study is conducted in a movement-independent manner, to reject the shadowing effects and isolate the targeted perceptual disorder to the maximum possible extent. Data was collected using a computerized graphical tool on 37 PD patients [6 early-stage *de novo*, 25 mid-stage using levodopa therapy, six later-stage using deep brain stimulation (DBS)] and 15 control participants. Besides the absolute measurement of perception through a psychometric analysis on two tested position reference magnitudes, we also investigated the linearity in perception using Weber's fraction. The results showed that individuals with PD displayed significant perceptual impairments compared to controls, though early-stage patients were not impaired. Mid-stage patients displayed impairments at the greater of the two tested reference magnitudes, while late-stage patients were impaired at both reference magnitudes. Levodopa and DBS use did not cause statistically significant differences in absolute displacement perception. The findings suggest abnormal visual processing in PD increasing with disease development, perhaps contributing to sensory-based impairments of PD such as bradykinesia, visuospatial deficits, and abnormal object recognition.

Keywords: Parkinson's disease, perception, vision, displacement, levodopa, deep brain stimulation, *de novo*, non-motor

INTRODUCTION

Although movement abnormalities have clinically defined Parkinson's disease (PD) since its 19th century definition, the motor system is not necessarily the sole root of abnormalities (Jankovic, 2008). Accurate movements rely on the initial collection and processing of environmental information by the sensory nervous system, sensorimotor integration, and the production of motor output signals sent to muscles, with disruption of any system impairing movement (Singer, 1980;

Noback et al., 2005). As exemplified with force computing and reproduction that is accurate in simple tasks, but slowed and variable during complex tasks, aspects of the abnormal motor functionality of PD can be rooted in neural dysfunction (Stelmach et al., 1989; Lafargue et al., 2008). Accordingly, movement abnormalities in PD are not necessarily due to abnormal motor function alone. Rather, motor dysfunction may be influenced by improper neural processing of stimuli, which along with PD-induced sensorimotor integrative deficiencies could lead to the observed motor dysfunction. However, due to movement generation being heavily used as a measure of the response in most clinical studies on PD it is not possible to decipher if abnormalities arise through the perceptual, motor, or sensorimotor integration deficits.

Visual information involving object and spatial properties is crucial for navigation and the production of accurate active and reactive movements (Azulay et al., 1999; Davidsdottir et al., 2005). Modern visual experience theory suggests vision to be a means to gain knowledge used to explore and manipulate the space around the host through motor activities (O'Regan and Noë, 2001). Considering this, vision perception and motion production are linked processes, with visual information processing being imperative for movement. Anatomically, visual perceptions are linked to a dorsal parietal stream involved in mapping an object's location in space, and a ventral occipitotemporal stream involved in object identification and memory (Goodale and Milner, 1992; Horwitz et al., 1992).

Visuospatial abnormalities in memory and representation of three-dimensional space are common among individuals with PD (Davidsdottir et al., 2005). These impairments can contribute to balance and navigation deficits increasing fall risk and injury (Azulay et al., 1999; Wood et al., 2002; Davidsdottir et al., 2005). Although visually-related deficits are common characteristics of PD (Corin et al., 1972; Wright et al., 1990; Haug et al., 1994; Büttner et al., 1995; Adamovich et al., 2001; Barnes and David, 2001; Zhu et al., 2016; Joyce et al., 2018), there is no systematic investigation on processing abnormalities of visual information for tasks relevant to movement that are not confounded by motor impairment. Furthermore, past findings of object recognition and navigational impairments observed in PD (Wood et al., 2002; Laatu et al., 2004; Davidsdottir et al., 2005; Lawrence et al., 2007; Clark et al., 2008; Almeida and Lebold, 2010; Martens and Almeida, 2012; Nantel et al., 2012) indicate PD-induced impairments in ventral occipitotemporal visual processing may exist. Recently, we have designed a computerized movement independent task using a virtual reality environment to generate computational statistical models of visual perceptions of time, which showed deficits in the accurate discrimination of temporal durations for those with PD (Bernardinis et al., 2019). The current paper extends the use of the computerized module allowing for statistical understanding of vision-based displacement perception.

We propose that the perceptual "tuning" of individuals with PD may be distorted, leading to improper processing of perceptual stimuli, thus causing inappropriate motor output. Although this motor output may be what one might expect based on the perception (i.e., is congruent to the perception), it is still incorrect due to perceptual inaccuracy, exemplified in healthy individuals by the changing of one's stride length and speed when perceiving a surface to be icy even if it eventually determined to be dry. To investigate this phenomenon, we must first assess the pure perceptual ability of PD patients. In this work, we have studied visual processing in PD through a perceptual task resembling movement-related displacement perception tasks (Demirci et al., 1997; Konczak et al., 2007), while isolating visual processing from movements and sensorimotor integration. Thus, we observed the ability of individuals with PD in accurately perceiving movement-independent visual displacement information. The impact of disease duration and the effect of dopaminergic and surgical treatment were also assessed, providing deeper insight on how the disease affects visual processes and the treatment effect on these abnormalities.

MATERIALS AND METHODS

Experimental Design

To study visual displacement perception independent of movement, a two-alternative forced-choice experiment (displayed in **Figure 1A**) was conducted. In the task, the displacement distance between two circular displacements presented in series is compared. Each displacement began with a white circle presented near the top or bottom of the monitor, followed by a displaced green circle. The participant responded (without time constraint) which displacement they perceived to be the largest in distance. In each of the 160 trials, one of two "standard stimuli" (10 and 17.5 cm) were compared to one of 8 "comparison stimuli." The comparison stimuli magnitudes for the 10 cm standard stimulus were 7, 8.5, 9, 9.5, 10.5, 11, 11.5, and 13 cm; and the comparison magnitudes for the 17.5 cm standard stimulus were 12.25, 14.85,15.75, 16.62, 18.36, 19.25, 20.1, and 22.75 cm. Comparison values were chosen based on pilot testing of healthy adults, in which comparison stimuli differing in magnitude the most from the standards were always answered correctly, and those differing in magnitude the least were answered correctly 50% of the time.

Testing Apparatus

Experimental visual stimuli were solely displayed on an LG Flatron W2242PM 22-inch visual monitor (resolution: 1,680 × 1,050). Participants sat in a comfortable, upright position approximately 60 cm (∼2 feet) in front of the monitor (**Figure 1B**). Both the height of the chair and monitor were adjusted for optimum viewing. Each participant, along with the examiner, were in an isolated room, minimizing auditory and visual distractions. The visual perception test was run in a virtual-reality environment designed at the Canadian Surgical Technologies and Advanced Robotics (CSTAR) lab and was connected to a real-time Matlab-Simulink program controlled by the experimenter.

Participants

Thirty-seven patients with PD (30 males, seven females) and 15 healthy, age-matched controls (12 females, three males) with

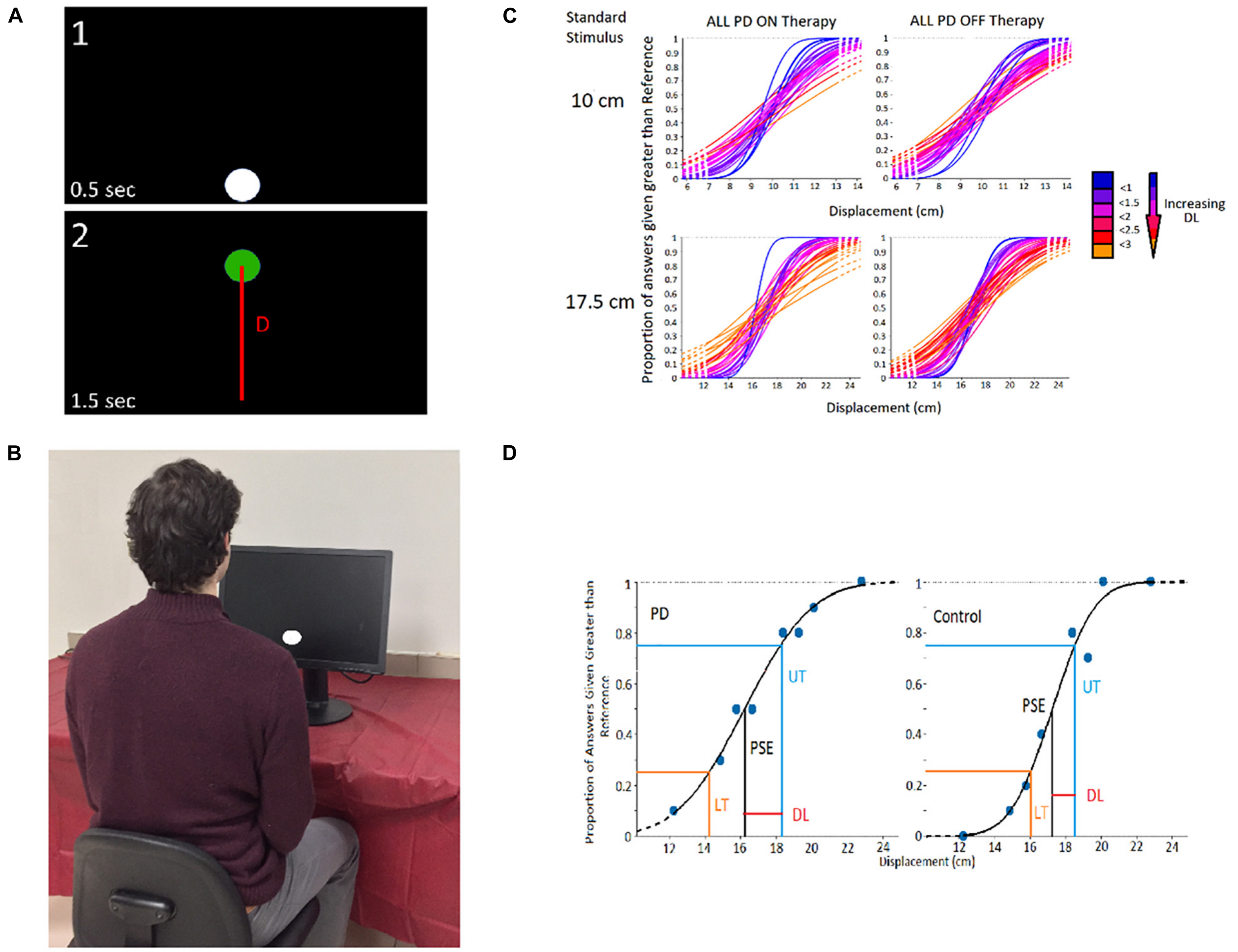


FIGURE 1 | Experimental Design. **(A)** Visual Displacement Perception Trial — Participant compares displacement distance between circles (D; line not visible in the experiment). **(B)** Experimental Setup. **(C)** Cumulative Gaussian Distributions heat map of PD participants and controls, in which distributions more blue in colour signify a greater slope (thus greater participant perceptual sensitivity), and those more red/orange in colour signifying a lesser slope (and worse participant perceptual ability). **(D)** Example analysis of Cumulative Gaussian Distribution to obtain DL.

no known neurological or psychiatric disorders were recruited from the Movement Disorders Program at London Health Sciences Centre, University Hospital in London, Ontario, Canada (**Table 1**). Recruitment involved assessing potential participant enrollment eligibility based on cognitive aptitude, participants fitting into the early-stage *de novo*, mid-stage levodopa using, or later-stage DBS using subgroups (based on the progression of PD and therapy usage), and the exclusion of candidates exhibiting PD symptoms that would impair experimental assessments (such as inabilities to focus, excessive fatigue, dystonia, etc.). These eligible candidates were met during clinic visits, where the experiment was described and their ability to conduct the experiment was further assessed based on cognitive fitness. Control participants generally had relation to PD participants, often being a spouse, family member, or friend, and were also contacted during clinic visits. The study protocol for this work was approved by the Research Ethics Board of the University of Western Ontario (REB 107253). All participants provided informed consent for participation in the study. All experiments were performed in accordance with the Tri-Council Policy Statement of Ethical Conduct for Research Involving Humans in Canada, as well as the Declaration of Helsinki. Of the 37 PD patients, 25 were treated using levodopa medication daily (half-life: 1–1.5 h) (Brooks, 2008). Though individual equivalent dose of levodopa differed from patient to patient, most consumed 200 mg of levodopa 3–4 times a day. Prior to experimentation, PD patients refrained from taking levodopa for at least 12 h to achieve a clinically defined OFF state. These patients initially conducted the experiment OFF levodopa, after which they were administered 300 mg of levodopa (unless their regular dose was 100 mg or lower, in which case they were administered 200 mg) and performed the task again (ON phase). No participants displayed dyskinesia with this acute dose. The ON and OFF experiments were conducted on the same day, involving a mandatory break of an hour after levodopa administration. Motor symptoms were assessed ON and OFF levodopa using section 3 (motor sub-scale) of the

Unified Parkinson's Disease Rating Scale (UPDRS). To further augment the findings and report additional observations, a pair of small-sample patient groups ($n = 6$ for both groups) relating to later-stage deep brain stimulation (DBS) using patients and early-stage *de novo* patients not yet using any PD therapy at the time of experimentation were analyzed as "case studies." The intention of these case studies was to observe potential trends regarding the visual displacement perception abilities of shorter and longer duration PD patients, and to observe the perceptual effects of DBS use. Due to the small sample sizes of these groups, the statistical tests act only as indicators of potential trends rather than a confirmation that differences are occurring between certain populations. If later-stage patients using DBS were also using levodopa, they refrained from using the medication 12 h prior to the experiment and throughout the day of testing. It should be noted that OFF DBS refers to OFF-OFF conditions (OFF stimulation and OFF dopaminergic medication), with ON DBS referring to ON-OFF conditions (ON stimulation, OFF dopaminergic medication). Prior to experimentation, DBS devices were turned OFF. After a 45-min waiting period, patients carried out the task in the same fashion as patients using levodopa. Experimentation occurred initially in the OFF-stimulation state, followed by the device being turned on and an hour break before experimentation in the ON state. Of note, the average age of later-stage patients using DBS was substantially lower than the levodopa patient group (**Table 1**). This is largely due to the presence of cognitive impairment in many elderly later-stage patients. Early-stage *de novo* PD patients ($n = 6$) only carried out the experiment once. Cognitive assessment of PD patients was conducted using the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005). Diagnostic assessments for visual acuity (using reading tasks and the Snellen eye chart), and smooth pursuit and saccadic eye movements were performed on all participants (PD and control) by an experienced clinician. PD patients were excluded from the study if they displayed visual, oculomotor, or substantial cognitive (MoCA < 25) impairments. Furthermore, PD patients experiencing visual hallucinations (PD-VH) or using PD medications other than levodopa were excluded from the study. It should be noted that patients included in the study did not exhibit severe impairments in color perception that would affect experimental performance.

Statistical Analysis

Initially, the correctness of patient responses was computed for each comparison value of a given standard stimulus. This data was then used to generate a probabilistic model (i.e., cumulative Gaussian distribution psychometric function) of the patient's perceptual ability for assessment where increased slope and shift signify perceptual impairment (**Figure 1C**; Fründ et al., 2011). For this, the Psignifit 4.0 third party Matlab toolbox was used. The participants' point of subjective equality (PSE) for both standard stimuli were calculated. The upper threshold (UT) and lower threshold (LT) were obtained through analysis of the psychometric function (**Figure 1D**), signifying the magnitude of displacement that was discerned from the standard stimulus 75% of the time (Gescheider, 2013). To assess an individual's absolute perceptual sensitivity for vision-based displacement perception, a two-forced alternative-choice assessment comparing two linear displacements was carried out as described by G. Gescheider (2013). The Difference Threshold (DL; DL = PSE–LT or DL = UT–PSE) was the unit used to measure perceptual sensitivity, signifying the difference in magnitude necessary to differentiate a stimulus from the standard stimulus. A participant's DL is inversely proportional to their perceptual sensitivity, with smaller DL's indicating greater perceptual ability. In each trial of the experiment one of the two standard stimuli magnitudes was compared to a smaller or larger comparison stimuli based on the standard stimulus present. Datum points were considered outliers and omitted from analysis if they were 1.5 × Interquartile Range (IQR) above the third quartile, or 1.5 × IQR below the first quartile.

The paired two-tailed t-test was utilized to statistically assess perceptual differences based on patient therapeutic state, and independent samples two-tailed t-tests were used for comparisons between PD and control groups. Furthermore, perceptual linearity rooted in "Weber's Law" (Baird and Noma, 1978) was analyzed to provide boosted sensitivity toward detecting potential abnormalities that may not have been observed through absolute assessment of the probabilistic models. Based on Weber's law, the ratio between an individual's DL and the amplitude of the standard stimulus is constant (Coren, 2003; Gescheider, 2013). The quantifiable value of Weber's Law, Weber's Fraction (WF), is defined as WF = DL/S, where S represents standard stimulus magnitude. Perceptions of healthy humans measured by WF have shown a strong linear relationship, following Weber's Law.

TABLE 1 | Summary of Demographic and Clinical Data for Tested PD Patients.

	Levodopa	DBS	*De novo*	Control
Demographic data				
Number (n)	25	6	6	15
Age (years)	70.04 ± 6.80	55.16 ± 8.89	74.17 ± 3.97	67.71 ± 8.82
Gender (m/f)	22/3	4/2	4/2	3/12
Total Years of Education	13.4 ± 4.36	13.33 ± 2.50	13.00 ± 1.67	13.76 ± 1.80
Years Since Diagnosis	6.88 ± 4.36	11.5 ± 4.04	3.12 ± 2.0	N/A
Clinical data				
MoCA (out of 30)	26.68 ± 2.17	26.67 ± 3.08	27.83 ± 2.14	27.23 ± 1.59
UPDRS motor sub-scale OFF Therapy	23.92 ± 6.69	34 ± 10.51	22.33 ± 7.91	N/A
UPDRS motor sub-scale ON Therapy	14.72 ± 6.07	22.33 ± 7.92	N/A	N/A
UPDRS motor subscale OFF vs. ON Difference	9.20 ± 5.09	21 ± 5.62	N/A	N/A

UPDRS, Unified Parkinson's Disease Rating Scale; MoCA, Montreal Cognitive Assessment.

RESULTS

PD vs. Control Displacement Perception Findings

Part A

When comparing all PD patients (those using levodopa, DBS, and *de novo* patients) with the control participants, there were no perceptual abnormalities seen at the smaller standard stimulus for those with Parkinson's disease OFF their respective primary therapies (*p*-value = 0.595; average DL for PD patients OFF therapy: 1.69 ± 0.48; average DL for control participants: 1.60 ± 0.58) and ON their respective therapies [*p*-value = 0.566; average DL for PD patients ON therapy: 1.50 ± 0.47) (**Figure 2A**)]. However, for the larger tested standard stimulus of 17.5 cm, PD patients displayed significant impairments in visual displacement perception OFF their respective therapies (*p*-value = 0.006; average DL for PD patients OFF therapy: 2.22 ± 0.75; average DL for control participants: 1.70 ± 0.44), as well as significant impairments while ON their PD therapies (*p*-value = 0.033; average DL for PD patients ON therapy: 2.10 ± 0.78) (**Figure 2B**).

Part B

Focusing only on mid-stage PD patients using levodopa, for the standard stimulus of 10 cm, the average DL for PD patients OFF levodopa did not differ (*p*-value = 0.954) from the DL of control participants (average DL for patients OFF levodopa: 1.61 ± 0.49; average DL for control participants: 1.60 ± 0.58). This group of PD patients also displayed insignificant differences (*p*-value = 0.372) in their DLs when ON levodopa compared to the tested controls (average DL for patients ON levodopa: 1.41 ± 0.36) for the standard stimulus of 10 cm (**Figure 2C**). For the larger tested stimuli (compared to the 17.5 cm standard stimulus), the DLs of PD patients OFF levodopa were significantly greater (*p*-value = 0.041) than control participant DLs (average DL for patients OFF levodopa: 2.09 ± 0.68; average DL for control participants: 1.70 ± 0.44). In addition, there was no significant difference (*p*-value = 0.120) regarding greater DLs for PD patients ON levodopa compared to control participants (average DL for patients ON levodopa: 2.03 ± 0.79) (**Figure 2D**). Levodopa administration did not directly elicit any significant effects on the absolute perceptual sensitivity of displacement for PD patients. Regarding the standard stimulus of 10 cm, an insignificant trend (*p*-value = 0.164) toward reduced DLs was observed after the patients received levodopa (average DL of patients OFF levodopa: 1.61 ± 0.51; average DL of patients ON levodopa: 1.44 ± 0.56). In addition, for the standard stimulus of 17.5 cm there were no changes to average DL (*p*-value = 0.655) after the participants received levodopa (average DL of patients OFF levodopa: 1.95 ± 0.59; average DL of patients ON levodopa: 2.03 ± 0.79) (**Figures 2C,D**).

Case Study A: Later-Stage Patients Using Deep Brain Stimulation

When looking at DLs between control participants and later-stage PD patients using DBS therapy at the standard stimulus of

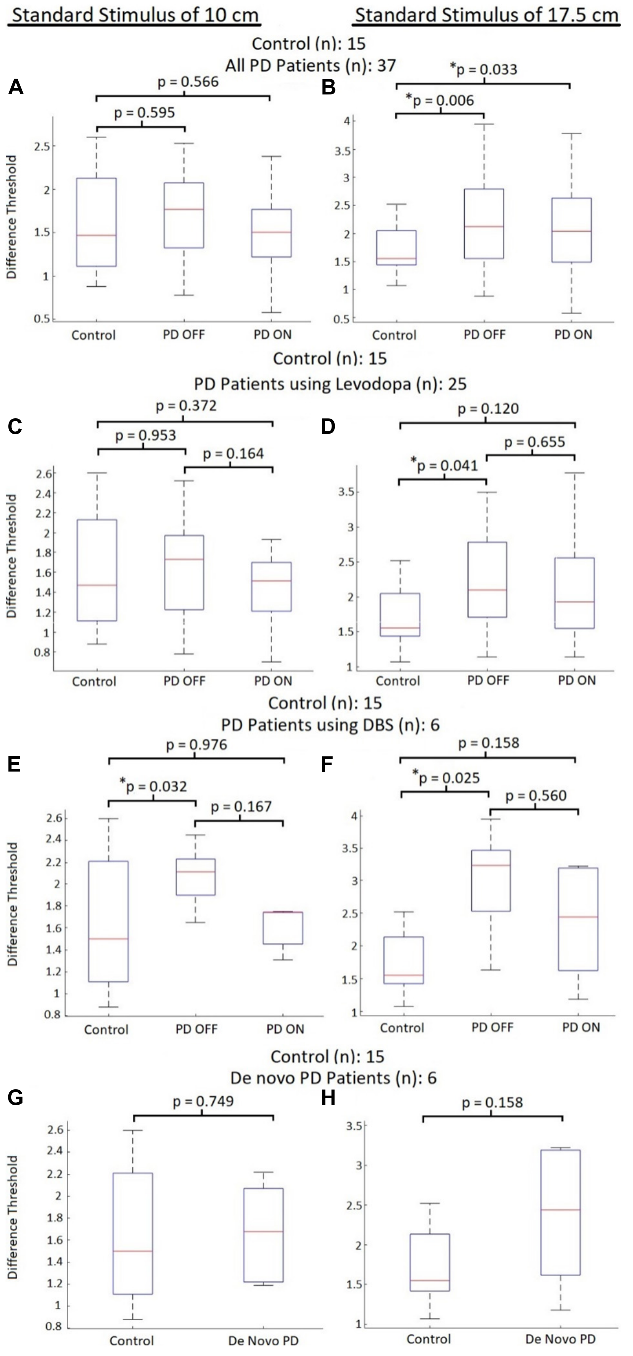


FIGURE 2 | Control vs. PD absolute perceptual ability comparisons. Comparison of visual displacement perceptual abilities (quantified using DL) between control participants and PD patients subdivided into groups based on therapeutic treatment. Red lines represent median DL for each group, with bars representing the data spectrum. **(A)** Comparisons to all PD participants with the 10 cm standard. **(B)** Comparisons to all PD participants with the 17.5 cm standard. **(C)** Comparisons to PD participants using levodopa with the 10 cm standard. **(D)** Comparisons to PD participants using levodopa with the 17.5 cm standard. **(E)** Comparisons to PD participants using DBS with the 10 cm standard. **(F)** Comparisons to PD participants using DBS with the 17.5 cm standard. **(G)** Comparisons to de novo PD participants with the 10 cm standard. **(H)** Comparisons to de novo PD participants with the 17.5 cm standard.

10 cm, PD participants displayed on average significantly greater DLs (p-value = 0.032) than controls when OFF DBS (average DL for patients OFF DBS: 2.07 ± 0.29). However, when ON DBS, there was no difference in DLs (p-value = 0.976) between PD patients and controls (average DL for patients ON DBS: 1.61 ± 0.20) (**Figure 2E**). Regarding the standard stimulus of 17.5 cm, patients OFF DBS displayed significantly greater DLs (p-value = 0.025) compared to control participants (average DL for patients OFF DBS: 2.99 ± 0.86). At the larger standard stimulus, no differences in DLs (p-value = 0.158) were seen for PD patients ON DBS compared to controls [average DL for patients ON DBS: 2.35 ± 0.94) (**Figure 2F**)].

For the PD patients using DBS, no significant differences in DLs were seen between ON and OFF states. For the smaller tested stimulus magnitudes (compared to the 10 cm standard stimulus), there was no significant DL difference (p-value = 0.167) when patients were ON DBS (average DL of patients OFF DBS: 2.06 ± 0.17; average DL for patients ON DBS: 1.57 ± 0.21) (**Figure 2E**). At the larger tested magnitudes (standard stimulus 17.5 cm) there was again no difference in DL (p-value = 0.560) when the patients were using DBS (average DL for patients OFF DBS: 2.99 ± 0.86; average DL for patients ON DBS: 2.58 ± 0.37) (**Figure 2F**). It should be noted that a relatively strong trend toward reduced DLs was observed when participants were ON DBS (**Figures 2E,F**). However, due to the small sample size of the DBS PD group (n = 6), it is possible that the statistical analysis is not representative of the therapy's impact on perceptual improvements. All but one of the patients using DBS displayed reduced DLs when ON DBS at both standard stimuli magnitudes.

Effect of Levodopa vs. DBS

Although no direct comparison on an individual's perceptual response to levodopa or DBS were made, comparisons of the therapies' efficacy can still be inferred from the data. First, when comparing the UPDRS motor subsection scores of patients using levodopa to DBS users OFF their respective therapies (in their base PD state), we see DBS users have significantly greater (p-value = 0.006) UPDRS scores than levodopa users (average UPDRS section III score for DBS PD patients: 34.00 ± 10.50; average UPDRS section III score for levodopa PD patients: 23.92 ± 6.69) (**Table 1**). As expected, later-stage DBS users had significantly greater motor impairment compared to mid-stage PD patients using levodopa therapy. Similarly, when comparing the DL for the standard stimulus of 17.5 cm, DBS patients OFF therapy (mean DL: 2.99 ± 0.86) displayed significantly greater (p-value = 0.015) DLs than patients using levodopa (mean DL: 2.09 ± 0.68). Furthermore, at the standard stimulus of 10 cm, very substantial trends (p-value = 0.059) toward greater DLs in DBS patients (mean DL: 2.07 ± 0.29) were observed compared to levodopa only patients (mean DL: 1.61 ± 0.49). This again is to be expected based on the earlier mentioned findings, as later-stage PD patients displayed more severe impairment in the tested vision-based perception compared to mid-stage PD patients. However, when these patients were ON their respective therapies no significant differences were observed between DLs at both the 10 cm standard (p-value = 0.478; mean DL for DBS PD patients: 1.61 ± 0.20; mean DL for levodopa PD patients: 1.43 ± 0.52) and the 17.5 cm standard (p-value = 0.412; mean DL for DBS PD patients: 2.35 ± 0.94; mean DL for levodopa PD patients: 2.03 ± 0.79). These findings may suggest that DBS therapy has a greater efficacy in treating the vision-based perception of displacement when movement is not involved compared to levodopa.

Case Study B: Early-Stage *de novo* Patients

Considering early stage *de novo* patients, for smaller stimuli magnitudes compared to the standard stimulus of 10 cm, there were insignificant differences between DLs (p-value = 0.749) of *de novo* PD patients compared to the control group (average DL for *de novo* patients was 1.68 ± 0.44) (**Figure 2G**). At the larger tested standard stimulus of 17.5 cm, *de novo* patients displayed an insignificant trend toward greater DLs (p-value = 0.158) compared to the control group (average DL for *de novo* patients: 2.35 ± 0.94) (**Figure 2H**). Thus, early-stage PD patients did not display significant differences in DL compared to controls.

Displacement Perception Linearity

Indeed, this study agreed with Weber's Law, showing a very strong correlation between the WF of the standard stimuli for healthy controls [Pearson correlation (R): 0.928, p-value < 0.001]. When comparing all PD patients OFF their respective therapies, they did not display significant correlations (R = 0.250, p-value = 0.135). However, when all PD patients were using their respective therapies, there were significant correlations seen between WFs (R = 0.762, p-value < 0.001). A similar pattern was observed when specifically looking at the levodopa group. For this group significant correlations were not observed when OFF levodopa (Pearson correlation: 0.235, p-value = 0.258). However, when these PD participants were administered levodopa strong correlations were observed between the WFs of different stimuli (Pearson Correlation: 0.821, p-value < 0.001) (**Figure 3**). Thus, administration of levodopa did appear to elicit some positive effects toward vision-based displacement perception.

DISCUSSION

This work shows that PD leads to visual, allocentric displacement perception impairments. These perceptual impairments arise without related movements, suggesting that the observed abnormality is intrinsic to the processing of visual information, and not dysfunctions occurring in sensorimotor integration or with the motor system. Although working memory and attentional deficit are well-noted symptoms of PD (Brown and Marsden, 1988; Calderon et al., 2001), it is improbable that these contributed to the observed perceptual abnormalities for the mid-stage PD patient group (using levodopa therapy) as no deficits were observed at the smaller tested magnitudes. Rather, the findings of the current study point toward impairment occurring in the ventral occipitotemporal and/or dorsal visual processing stream(s) in PD, showing a behavioral response to this known pathway. This provides a rationale regarding the potential basis for observed PD-induced deficits in activities utilizing ventral

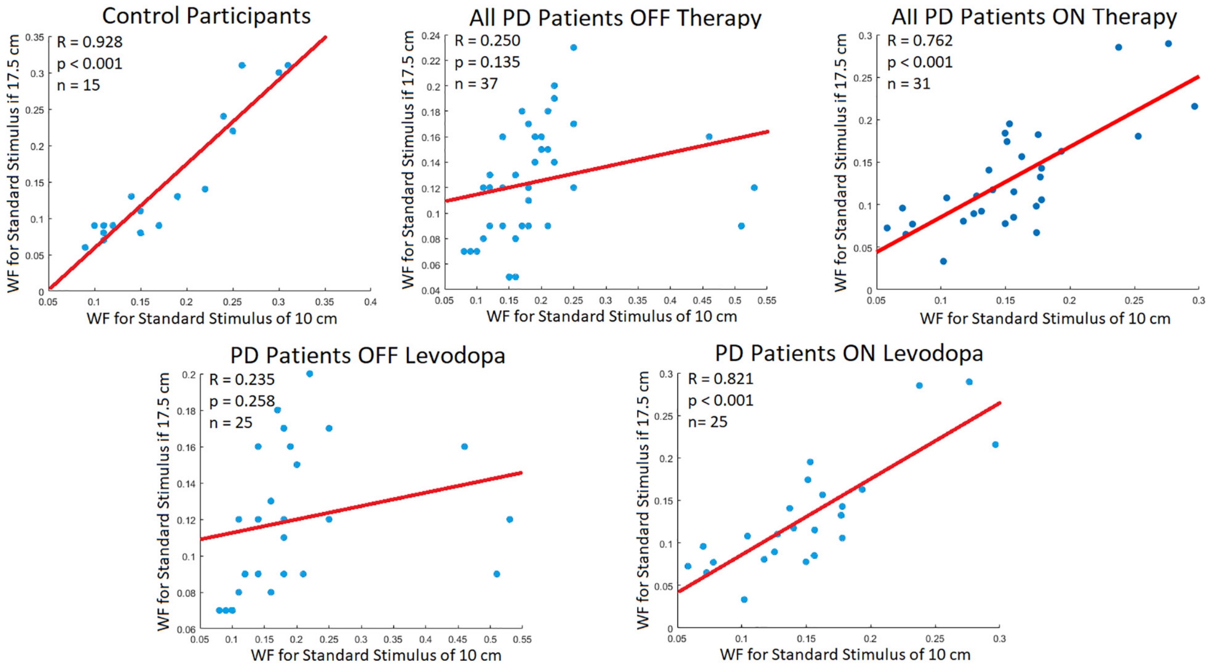


FIGURE 3 | Participant WF Correlation. Correlations (R; Pearson correlation coefficient) between participant WF at the standard stimuli of 10 and 17.5 cm. The red line signifies the line of best fit for the correlation of data points. According to Weber's Law, there should be strong correlations between WFs of different standard stimuli.

visual processing such as object and facial recognition (Laatu et al., 2004; Lawrence et al., 2007; Clark et al., 2008). This also is in similitude with the belief that freezing of gait (FoG) in PD is rooted in perceptual rather than motor deficiencies (Almeida and Lebold, 2010; Nantel et al., 2012). As proposed, the observed motor output may be an appropriate transformation of the instructions from the motor systems. However, the motor systems may be responding to errors in the processing and integration of visual information. Considering how this relates to FoG, errors in perception of distances processed through egocentric coordinates (using oneself as a reference) occurring in PD (Lee et al., 2001; Martens et al., 2013), as well as impairments in allocentric distances observed in the current study could lead to incorrect internal perceptions of one's own dimensions and the dimensions of objects and structures in their surrounding environment. Thus, movement outputs are produced in relation to the skewed visual processing, for example, leading to errors in which an individual overestimates their size while underestimating the width of the doorframe, causing gait freezing.

When considering the effect of PD therapies, levodopa and DBS were shown to not directly increase perceptual sensitivity on the tested visual displacement task for either tested magnitudes. Though this agrees with prior work questioning dopamine's relevance in movement-independent tasks impaired by PD and notions that common PD therapies are not beneficial for non-motor symptoms of the disease at their administered dosage (Deep-Brain Stimulation for Parkinson's Disease Study Group, et al., 2001; Ahlskog, 2005; Chaudhuri et al., 2006), it is interesting nonetheless due to the current studies task (and visual processing in general) being tightly linked to movement processes. However, linear relationships between perceptual sensitivity and magnitude (as per Weber's Law) were very weak when PD patients were OFF dopaminergic treatment, becoming strong when ON levodopa. We postulate this is due to dopaminergic treatment "tightening" the regulatory bounds of perception, in that the topology of the perceptual map re-orients after levodopa use, leading to greater overlap with controls in some domains. This would imply topographic perceptual maps are not normalized or "tightened" enough to allow all aspects of the perception to be improved. It is worth noting some aspects of non-motor disorders in PD do improve with levodopa use at the later-stages of the disease, specifically with the reduction of pain and anxiety (Fabbri et al., 2017).

Comparing patients using levodopa to those using DBS, we see that in OFF Parkinsonian states, the DBS-using participant group displayed significantly worse displacement perception abilities than levodopa-using patients. However, once these groups were administered their respective therapies, this gap in perceptual ability greatly shrank. This suggests that the improvements provided by subthalamic stimulation are greater than those brought on from dopaminergic treatment. However, it is possible that levodopa does indeed act in a beneficial manner toward correcting the visual displacement perceptual deficits in PD, and that it is other neurochemical imbalances (such as abnormal noradrenaline balances) or the widespread effect of levodopa targeting undesired neural regions that led to the observed perceptual abnormalities. It should be noted that a relatively

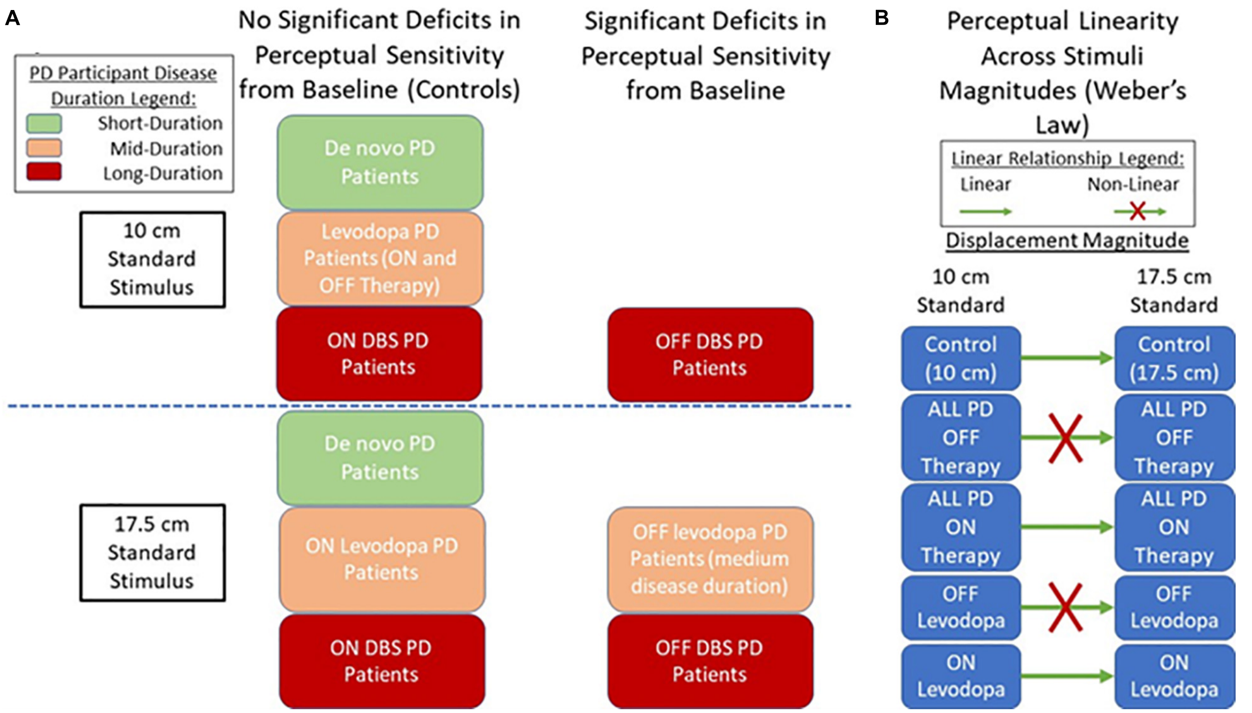


FIGURE 4 | Displacement Perception in PD Summary. **(A)** Findings on displacement perception sensitivity between PD patient groups compared to control (baseline) participants. **(B)** Findings on the state of perceptual linearity according to Weber's law.

strong trend regarding reductions in patient DL when ON DBS (compared to OFF DBS) was observed, along with impairments in subject DL (compared to controls) only being observed when OFF DBS (**Figure 2**). Thus, based on the results it appears that DBS might improve absolute visual displacement sensitivity for PD patients while levodopa did not cause improvement. However, statistical analysis was not fully representative of the population due to small power, necessitating future work for validation. If DBS does indeed display greater efficacy toward normalizing perceptual processing compared to dopaminergic therapies, it may support discussions and further investigations regarding the benefit of earlier surgical intervention.

The duration of PD appears to be related to performance in the tested allocentric visual perception task. Though the sample size is a limitation, *de novo* patients in the early stages of PD did not display any significant perceptual impairments in the current task compared to control participants. Patients in mid-stages of PD utilizing the dopaminergic medication as their primary PD therapy did display deficiencies in the displacement perception task. However, these were limited to the greater stimuli magnitudes compared to the standard stimulus of 17.5 cm, whereas late-stage patients (utilizing DBS therapy) displayed impairments at both tested standards. The results suggest that increased disease severity broadens the range of affected magnitudes. As tested individuals using DBS were at later stages of the disease, impaired memory and/or attentional performance might be involved (although subjects with observable deficits in these areas were rejected from study participation). Alternatively, increasingly severe PD symptoms may lead to a broader range of perceptual deficits through increasingly impaired occipitotemporal processing. This phenomenon should be further investigated as visual allocentric displacement perception may provide a valuable sensory modality that can be used to monitor PD progression without the use of motor function.

To conclude our findings (**Figure 4**), allocentric visual displacement perception deficits independent of associated movements were observed in PD, with longer disease duration appearing to lead to more widespread perceptual abnormalities. Levodopa therapy did not appear to directly improve base perceptual ability; however, it may have modulated the parameters of perception to be more like controls (seen through improved perceptual linearity). DBS appeared to be more effective toward improving the studied perception, warranting further work analyzing its effect on non-motor perceptions. Future work should further investigate the neurological basis for these abnormalities and investigate the use of visual displacement perception for disease monitoring. Furthermore, future work should expand on the major limitations of the study, namely, the small sample sizes (particularly for the later-stage DBS using and early-stage *de novo* patient subgroups), and the discrepancy in the sex-makeup of the PD group (largely male) and control group (largely female) that may have impact on sex-based perceptual abilities (Herrera-Guzmán et al., 2004). The age discrepancy between later-stage patients using DBS and mid-stage patients using levodopa should be noted, along with the possibility of slight residual effects from levodopa occurring in participants using the treatment, though they are in a clinically defined OFF state.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors upon reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Board of the University of Western Ontario (REB 107253). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MB was involved in the conception and organization of the project, recruiting and execution of the research, statistical analysis of the data, and wrote the first draft and carried out further editing in subsequent drafts. SA contributed to the development of the research problem's planning, execution, study design, and conduct, provided guidance for the experimental setup, and the statistical analysis for the project. RP proposed the research problem, contributed to the planning, organization, and development of the research project, and helped with the preparation and writing of the paper. RP and MJ supervised the thesis research of MB on which this manuscript is based. MJ assisted in the development of the clinical aspects of the research, and provided access to the participants included in the study, while also contributing to the analysis of the data and the preparing and writing of the manuscript. All authors have reviewed and edited the manuscript.

FUNDING

The work was supported by the AGE-WELL Network of Centres of Excellence under Grant Number AW CRP 2015 WP5.3 awarded to RP and MJ. SA was supported as a Postdoctoral Research Associate by grants awarded to RP.

ACKNOWLEDGMENTS

The authors would like to thank all the participants who took part in the experiments. The authors would also like to thank Olivia Samotus, Karen Siroen, Daphne Hui, Amit Srivastava, and Lisa Milligan for assisting with the experiments. The basic foundations of this work were published in the IEEE 2018 EMBC conference proceedings, entitled Visual Displacement Perception in Parkinson's Disease Analyzed Using a Computer-Generated Graphical Tool.

REFERENCES

Adamovich, S. V., Berkinblit, M. B., Hening, W., Sage, J., and Poizner, H. (2001). The interaction of visual and proprioceptive inputs in pointing to actual and remembered targets in Parkinson's disease. *Neuroscience* 104, 1027–1041. doi: 10.1016/s0306-4522(01)00099-9

Ahlskog, J. E. (2005). Challenging conventional wisdom: the etiologic role of dopamine oxidative stress in Parkinson's disease. *Move. Disord.* 20, 271–282. doi: 10.1002/mds.20362

Almeida, Q. J., and Lebold, C. A. (2010). Freezing of gait in Parkinson's disease: a perceptual cause for a motor impairment? *J. Neurol. Neurosurg. Psychiatry* 81, 513–518. doi: 10.1136/jnnp.2008.160580

Azulay, J. P., Mesure, S., Amblard, B., Blin, O., Sangla, I., and Pouget, J. (1999). Visual control of locomotion in Parkinson's disease. *Brain* 122, 111–120. doi: 10.1093/brain/122.1.111

Baird, J. C., and Noma, E. J. (1978). *Fundamentals of Scaling and Psychophysics*. New York, NJ: John Wiley & Sons.

Barnes, J., and David, A. S. (2001). Visual hallucinations in Parkinson's disease: a review and phenomenological survey. *J. Neurol. Neurosurg. Psychiatry* 70, 727–733. doi: 10.1136/jnnp.70.6.727

Bernardinis, M., Atashzar, S. F., Jog, M. S., and Patel, R. V. (2019). Differential temporal perception abilities in parkinson's disease patients based on timing magnitude. *Sci. Rep.* 9, 1–16.

Brooks, D. J. (2008). Optimizing levodopa therapy for Parkinson's disease with levodopa/carbidopa/entacapone:implications from a clinical and patient perspective. Neuropsychiatric disease and treatment. *Neuropsychiatry Disord. Treatment* 4, 39–47. doi: 10.2147/ndt.s1660

Brown, R. G., and Marsden, C. D. (1988). Internal versus external cues and the control of attention in Parkinson's disease. *Brain* 111, 323–345. doi: 10.1093/brain/111.2.323

Büttner, T., Kuhn, W., Müller, T., Patzold, T., Heidbrink, K., and Przuntek, H. (1995). Distorted color discrimination in 'de nova' Parkinsonian patients. *Neurology* 45, 386–387. doi: 10.1212/wnl.45.2.386

Calderon, J., Perry, R. J., Erzinclioglu, S. W., Berrios, G. E., Dening, T., and Hodges, J. R. (2001). Perception, attention, and working memory are disproportionately impaired in dementia with Lewy bodies compared with Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* 70, 157–164. doi: 10.1136/jnnp.70.2.157

Chaudhuri, K. R., Healy, D. G., and Schapira, A. H. (2006). Non-motor symptoms of Parkinson's disease: diagnosis and management. *Lancet Neurol.* 5, 235–245.

Clark, U. S., Neargarder, S., and Cronin-Golomb, A. (2008). Specific impairments in the recognition of emotional facial expressions in Parkinson's disease. *Neuropsychologia* 46, 2300–2309. doi: 10.1016/j.neuropsychologia.2008.03.014

Coren, S. (2003). *Sensation and Perception*. Hoboken, NJ: Wiley Online Library.

Corin, M. S., Elizan, T. S., and Bender, M. B. (1972). Oculomotor function in patients with Parkinson's disease. *J. Neurol. Sci.* 15, 251–265. doi: 10.1016/0022-510x(72)90068-8

Davidsdottir, S., Cronin-Golomb, A., and Lee, A. (2005). Visual and spatial symptoms in Parkinson's disease. *Vis. Res.* 45, 1285–1296. doi: 10.1016/j.visres.2004.11.006

Deep-Brain Stimulation for Parkinson's Disease Study Group, Obeso, J. A., Rodriguez-Oroz, M. C., Kumar, R., and Lang, A. E. (2001). Deep-brain stimulation of the subthalamic nucleus or the pars interna of the globus pallidus in Parkinson's disease. *N. Engl. J. Med.* 345, 956–963. doi: 10.1056/nejmoa000827

Demirci, M., Grill, S., McShane, L., and Hallett, M. (1997). A mismatch between kinesthetic and visual perception in Parkinson's disease. *Ann. Neurol.* 41, 781–788. doi: 10.1002/ana.410410614

Fabbri, M., Coelho, M., Guedes, L. C., Chendo, I., Sousa, C., Rosa, M. M., et al. (2017). Response of non-motor symptoms to levodopa in late-stage Parkinson's disease: Results of a levodopa challenge test. *Parkinsonism Relat. Disord.* 39, 37–43. doi: 10.1016/j.parkreldis.2017.02.007

Fründ, I., Haenel, N. V., and Wichmann, F. A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *J. Vis.* 11:16. doi: 10.1167/11.6.16

Gescheider, G. A. (2013). *Psychophysics: The Fundamentals*. Mahwah, NJ: Psychology Press.

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25. doi: 10.1016/0166-2236(92)90344-8

Haug, B. A., Trenkwalder, C., Arden, G. B., Oertel, W. H., and Paulus, W. (1994). Visual thresholds to low-contrast pattern displacement, color contrast, and luminance contrast stimuli in Parkinson's disease. *Move. Disord.* 9, 563–570. doi: 10.1002/mds.870090510

Herrera-Guzmán, I., Peña-Casanova, J., Lara, J. P., Gudayol-Ferré, E., and Böhm, P. (2004). Influence of age, sex, and education on the visual object and space perception battery (VOSP) in a healthy normal elderly population. *Clin. Neuropsychol.* 18, 385–394. doi: 10.1080/1385404049052421

Horwitz, B., Grady, C. L., Haxby, J. V., Schapiro, M. B., Rapoport, S. I., Ungerleider, L. G., et al. (1992). Functional associations among human posterior extrastriate brain regions during object and spatial vision. *J. Cogn. Neurosci.* 4, 311–322. doi: 10.1162/jocn.1992.4.4.311

Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. *J. Neurol. Neurosurg. Psychiatry* 79, 368–376.

Joyce, D. S., Feigl, B., Kerr, G., Roeder, L., and Zele, A. J. (2018). Melanopsin-mediated pupil function is impaired in Parkinson's disease. *Sci. Rep.* 8:7796.

Konczak, J., Krawczewski, K., Tuite, P., and Maschke, M. (2007). The perception of passive motion in Parkinson's disease. *J. Neurol.* 254, 655. doi: 10.1007/s00415-006-0426-2

Laatu, S., Revonsuo, A., Pihko, L., Portin, R., and Rinne, J. O. (2004). Visual object recognition deficits in early Parkinson's disease. *Parkinsonism Relat. Disord.* 10, 227–233. doi: 10.1016/j.parkreldis.2004.02.001

Lafargue, G., D'Amico, A., Thobois, S., Broussolle, E., and Sirigu, A. (2008). The ability to assess muscular force in asymmetrical Parkinson's disease. *Cortex* 44, 82–89. doi: 10.1016/j.cortex.2005.11.001

Lawrence, A. D., Goerendt, I. K., and Brooks, D. J. (2007). Impaired recognition of facial expressions of anger in Parkinson's disease patients acutely withdrawn from dopamine replacement therapy. *Neuropsychologia* 45, 65–74. doi: 10.1016/j.neuropsychologia.2006.04.016

Lee, A. C., Harris, J. P., Atkinson, E. A., and Fowler, M. S. (2001). Disruption of estimation of body-scaled aperture width in Hemiparkinson's disease. *Neuropsychologia* 39, 1097–1104. doi: 10.1016/s0028-3932(01)00032-x

Martens, K. A. E., and Almeida, Q. J. (2012). Dissociating between sensory and perceptual deficits in PD: more than simply a motor deficit. *Movement Disorders* 27, 387–392. doi: 10.1002/mds.24042

Martens, K. A. E., Ellard, C. G., and Almeida, Q. J. (2013). Dopaminergic contributions to distance estimation in Parkinson's disease: a sensory-perceptual deficit? *Neuropsychologia* 51, 1426–1434. doi: 10.1016/j.neuropsychologia.2013.04.015

Nantel, J., McDonald, J. C., Tan, S., and Bronte-Stewart, H. (2012). Deficits in visuospatial processing contribute to quantitative measures of freezing of gait in Parkinson's disease. *Neuroscience* 221, 151–156. doi: 10.1016/j.neuroscience.2012.07.007

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x

Noback, C. R., Ruggiero, D. A., Demarest, R. J., and Strominger, N. L. (2005). *The Human Nervous System: Structure and Function*. Totowa, NJ: Springer Science & Business Media.

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–973. doi: 10.1017/s0140525x01000115

Singer, R. N. (1980). *Motor Learning and Human Performance: An Application to Motor Skills and Movement Behaviors*. New York, NY: Macmillan.

Stelmach, G. E., Teasdale, N., Phillips, J., and Worringham, C. J. (1989). Force production characteristics in Parkinson's disease. *Exp. Brain Res.* 76, 165–172.

Wood, B. H., Bilclough, J. A., Bowron, A., and Walker, R. W. (2002). Incidence and prediction of falls in Parkinson's disease: a prospective multidisciplinary study. *J. Neurol. Neurosurg. Psychiatry* 72, 721–725. doi: 10.1136/jnnp.72.6.721

Wright, M. J., Burns, R. J., Geffen, G. M., and Geffen, L. B. (1990). Covert orientation of visual attention in Parkinson's disease: an impairment in the maintenance of attention. *Neuropsychologia* 28, 151–159. doi: 10.1016/0028-3932(90)90097-8

Zhu, M., Li, M., Ye, D., Jiang, W., Lei, T., and Shu, K. (2016). Sensory symptoms in Parkinson's disease: clinical features, pathophysiology, and treatment. *J. Neurosci. Res.* 94, 685–692. doi: 10.1002/jnr.23729

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 26 July 2021
doi: 10.3389/fnins.2021.698940

Analysis and Synthesis of Natural Texture Perception From Visual Evoked Potentials

Taiki Orima[1,2]* *and Isamu Motoyoshi*[1]

[1] *Department of Life Sciences, The University of Tokyo, Tokyo, Japan,* [2] *Japan Society for the Promotion of Science, Tokyo, Japan*

The primate visual system analyzes statistical information in natural images and uses it for the immediate perception of scenes, objects, and surface materials. To investigate the dynamical encoding of image statistics in the human brain, we measured visual evoked potentials (VEPs) for 166 natural textures and their synthetic versions, and performed a reverse-correlation analysis of the VEPs and representative texture statistics of the image. The analysis revealed occipital VEP components strongly correlated with particular texture statistics. VEPs correlated with low-level statistics, such as subband SDs, emerged rapidly from 100 to 250 ms in a spatial frequency dependent manner. VEPs correlated with higher-order statistics, such as subband kurtosis and cross-band correlations, were observed at slightly later times. Moreover, these robust correlations enabled us to inversely estimate texture statistics from VEP signals via linear regression and to reconstruct texture images that appear similar to those synthesized with the original statistics. Additionally, we found significant differences in VEPs at 200–300 ms between some natural textures and their Portilla–Simoncelli (PS) synthesized versions, even though they shared almost identical texture statistics. This differential VEP was related to the perceptual "unnaturalness" of PS-synthesized textures. These results suggest that the visual cortex rapidly encodes image statistics hidden in natural textures specifically enough to predict the visual appearance of a texture, while it also represents high-level information beyond image statistics, and that electroencephalography can be used to decode these cortical signals.

Keywords: image statistics, visual evoked potentials, texture perception, stimulus reconstruction, naturalness perception

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Rui Li,
Xi'an Jiaotong University, China
Vikram Shenoy Handiru,
Center for Mobility and Rehabilitation Engineering Research, Kessler Foundation, United States

***Correspondence:**
Taiki Orima
taikiorima@gmail.com

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *22 April 2021*
Accepted: *21 June 2021*
Published: *26 July 2021*

Citation:
Orima T and Motoyoshi I (2021) Analysis and Synthesis of Natural Texture Perception From Visual Evoked Potentials. Front. Neurosci. 15:698940. doi: 10.3389/fnins.2021.698940

INTRODUCTION

The visual field is full of complex image regions called "textures." Increasing evidence shows that textural information, or ensemble statistics, play a key role in the rapid perception and recognition of scenes, objects, and surface materials (Lowe, 1999; Oliva and Torralba, 2001; Motoyoshi et al., 2007; Whitney et al., 2014; De Cesarei et al., 2017; Fleming, 2017; Nishida, 2019).

It has widely been suggested that the perception of a texture is essentially based on the spatial distributions of low-level image features and their relationships (Julesz, 1965; Graham et al., 1992; Landy and Graham, 2004). Following extensive investigations into the neural computations underlying texture segregation (Bergen and Adelson, 1988; Zipser et al., 1996;

Baker and Mareschal, 2001), recent studies have re-formalized the theory in terms of image statistics (Portilla and Simoncelli, 2000; Freeman and Simoncelli, 2011; Freeman et al., 2013; Wallis et al., 2017). Specifically, the early visual cortex decomposes an image into multiple subbands of different orientation and spatial frequency, encodes moment statistics and correlations across subbands of different orientation and spatial frequency, and exploits these statistics to discriminate among various texture images. Compelling evidence for this framework is provided by texture-synthesis algorithms (Heeger and Bergen, 1995; Portilla and Simoncelli, 2000), which can synthesize a texture image that looks similar to a given texture by simply matching image statistics of white noise to those of the target texture.

Recent studies adopting functional magnetic resonant imaging and electrophysiology suggest that texture statistics are represented in the early visual cortex (Freeman and Simoncelli, 2011; Freeman et al., 2013; Okazawa et al., 2015, 2017). Yet, it is unclear how each class of statistic is encoded in the human brain, especially during the early processing of the image. To examine such a rapid cortical response in humans, electroencephalography (EEG) has widely been used as an easy and non-invasive measure. In visual neuroscience, classical studies have examined visual evoked potentials (VEPs) for a specific image feature, but with artificial patterns composed of lines and dots (Victor and Conte, 1991; Bach and Meigen, 1997, 1998; Peterzell and Norcia, 1997; Bach et al., 2000; Norcia et al., 2015; Kohler et al., 2018). More recently, several studies directly measured VEPs for natural images. Adopting reverse correlation analysis (DeAngelis et al., 1993), they successfully extracted VEP components correlated to particular image features, such as pixel statistics, phase statistics, the scene “gist,” and deep features (Rousselet et al., 2008; Scholte et al., 2009; Bieniek et al., 2012; Groen et al., 2012a,b, 2017; Hansen et al., 2012; Ghodrati et al., 2016; Greene and Hansen, 2020). However, these features are not powerful enough to fully describe the perception of individual images of scenes and objects they employed, and it is uncertain if the VEP components correlated with those features are truly relevant to the perception. In addition, those features are indifferent to texture perception.

In contrast to the perception of scenes and objects, the perception of textures is well described and even synthesized by a particular set of image statistics (Portilla and Simoncelli, 2000). Moreover, such image statistics are spatially global measurements, whose neural representations could be captured by EEG with a low spatial resolution. Taking advantage of these facts, the present study elucidates human cortical responses to texture statistics using a reverse correlation between VEPs for various natural textures and image statistics that are critical for the perceptual appearance of a texture. Our analysis revealed VEP components specifically correlated with low- and high-level texture statistics. On the basis of this robust correlation, we reconstructed image statistics from VEPs with linear regression and successfully synthesized perceptually mimicked textures simply from VEP signals. These results suggest that VEPs can capture neural responses to texture statistics specifically enough for the prediction of the perceptual appearance of individual images. We found different VEPs between natural textures and their synthetic versions, but those VEPs were limited to images in which texture statistics were not sufficient to synthesize the appearance of natural textures.

MATERIALS AND METHODS

Observers

Fifteen naïve, paid observers (22 years old on average) participated in the experiment. All participants had normal or corrected-to-normal vision. All experiments were conducted in accordance with the guidelines of the Ethics Committee for experiments on humans at the Graduate School of Arts and Sciences, The University of Tokyo. All participants provided written informed consent.

Apparatus

Visual stimuli were displayed on a gamma-corrected 24-inch liquid-crystal display (BENQ XL2420T) with a frame rate of 60 Hz. The pixel resolution was 1.34 min/pixel at a viewing distance of 100 cm, and the mean luminance of the uniform background was 33 cd/m^2.

Stimuli

The visual stimuli comprised 166 natural texture images, each subtending 5.7° × 5.7° (256 × 256 pixels; **Figure 1A**). Images were taken from our original natural-texture image database or from the Internet. All RGB images were converted to gray scale, and the mean luminance was normalized to 33 cd/m^2, which was equal to that of the gray background.

For comparison with the original natural textures, we additionally employed two types of synthesized image. One was an image synthesized by means of the Portilla–Simoncelli (PS) algorithm (**Figure 1B**; Portilla and Simoncelli, 2000), which can create a perceptually similar texture by matching low- and high-level image statistics of a white noise image, including moment statistics [i.e., standard deviation (SD), skew, and kurtosis] and cross-band correlations, to those of the original texture image. The synthesis was performed with a typical parameter setting as used in the original algorithm (except for the number of iterations) (Portilla and Simoncelli, 2000). The other synthetic textures were made by randomizing the spatial phase of the original natural textures (**Figure 1C**). These phase-randomized images were equivalent to the original image only in terms of the global spatial frequency spectrum.

Procedure

Electroencephalographys were measured in an electrically shielded, dark room. In each experimental session, each of 166 natural textures was presented once in random order, with a 500-ms duration followed by a 750-ms interval of the uniform gray background. Observers viewed the stimulus binocularly with steady fixation on a small black dot (10.8-min in diameter) that was shown at the center of the display throughout the session. For each observer, the sessions were repeated 24 times. The same measurements were also run as different blocks for the PS-synthesized textures and for the phase-randomized textures.

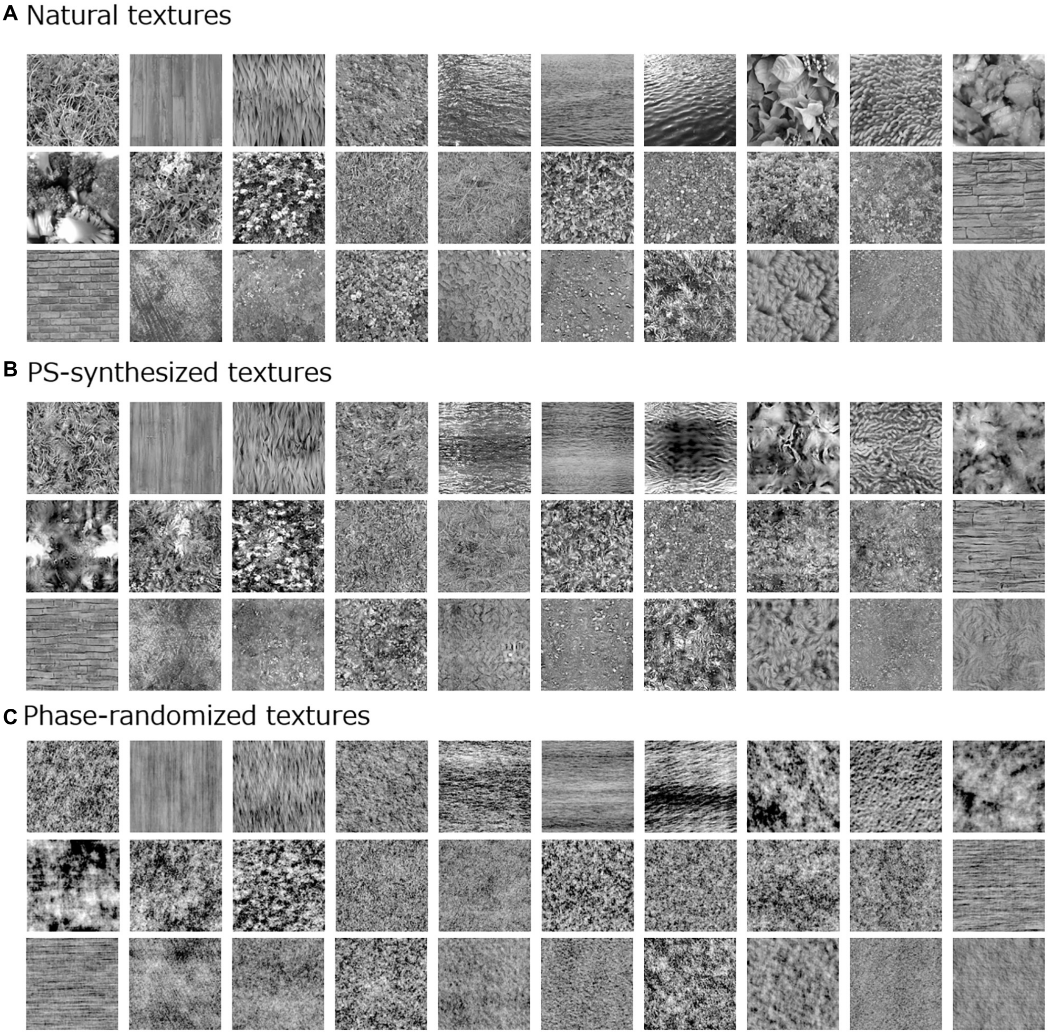


FIGURE 1 | Examples of visual stimuli used in the experiment: **(A)** natural textures; **(B)** Portilla–Simoncelli (PS)-synthesized versions; and **(C)** phase-randomized versions.

Each block was conducted in the same order for all participants on different days. Therefore, each observer spent 3 days in total participating in the EEG recordings (i.e., measurements were made for natural textures on the first day, PS-synthesized textures on the second day, and phase-randomized textures on the third day).

EEG Recordings and Preprocess

The EEG recordings were conducted using electrodes positioned at Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, Fz, Cz, and Pz, in accordance with the international 10–20 system, at a 1,000-Hz sampling rate, using Ag-AgCl electrodes and an electrode cap of appropriate size (BrainVision Recorder, BrainAmp amplifier, EasyCap; Brain Products GmbH). An additional electrode, which served as the common ground electrode, was placed midway between Fz and Fpz. All electrodes were referenced to another electrode positioned between Fz and Cz, and they were re-referenced off-line using the average amplitude of all electrodes. The EEG was resampled at 250 Hz, band-pass filtered at 0.1–100 Hz, and converted to epochs of −0.4 to 0.8 s from the stimulus onset. The power frequency component (50 Hz) was automatically rejected when the EEG was recorded. The baseline was from −0.1 to 0 s with respect to the stimulus onset, and the EEG was corrected relative to the baseline. Artifact components (i.e., eye movements) were removed by the heuristic examination of independent components. To remove epochs with eye blinks, epochs with an amplitude outside the range from −75 to 75 μV (i.e., 1.7% of all epochs) were rejected. VEPs for each image were defined as the average across the 24 repetitions. We compensated for machinery delay that was measured in each trial.

Analysis of Image Statistics

We analyzed image statistics for each texture image. In the analysis, the PS statistics space was not used directly because it was primarily designed for synthesis and consists of very complicated combinations of parameters, which are not suitable for visualizing the results. Instead, we chose several classes of statistics that are known to be particularly important in human texture models, including the PS model (Portilla and Simoncelli, 2000; Simoncelli and Olshausen, 2001; Landy and Graham, 2004). In any natural image, some of these statistics may be correlated with each other, but we defined them as independent classes in terms of their properties. Thus, we decomposed each image into different orientation and spatial frequency subbands and computed five representative image statistics: the SD, skew, kurtosis, correlation between different orientation subbands, and correlation between different spatial frequency subbands. In this space, we confirmed that natural textures and their PS-synthesized versions had almost identical, or very similar, image statistics ($r = 0.83$ on average).

For each texture, the luminance image was first decomposed to subbands of seven spatial frequencies (2–128 cycles/image, 1-octave steps: 0.35, 0.70, 1.40, 2.80, 5.61, 11.2, and 22.4 cycles/deg) (e.g., De Valois and De Valois, 1980) and eight orientation bands (0–157.5°, 22.5° steps) by using a linear Gaussian band-pass filter with a spatial frequency bandwidth (i.e., full width at half-maximum) of 1 octave and an orientation bandwidth of 30°. For each subband image, three moment statistics (i.e., log SD, skewness, and log kurtosis) were calculated. The central three panels in **Figure 2** show these three moment statistics obtained from a sample image (left-most image in **Figure 2**) and plotted as functions of orientation (x-axis) and spatial frequency (y-axis). We did not consider pixel statistics because visual cortical neurons have no direct access to pixel information.

In addition, correlations between subband "energy" images of different orientation and spatial frequency were calculated. These are known to be important high-level image statistics in texture synthesis (Portilla and Simoncelli, 2000). In detail, the cross-orientation energy correlations are related to how much local features in the image are oriented, and the cross-frequency energy correlations are related to how much the local luminance modulations are edgy or stepwise (Portilla and Simoncelli, 2000; Balas et al., 2009). Here, the energy image was given as a vector sum of the cosine and sine parts of the subband image. We calculated correlations in the energy image between different orientation bands along the same spatial frequency and between different spatial frequency bands along the same orientation. We then averaged the resulting correlations across orientation because the absolute orientation rarely matters in texture perception.

Specifically, we computed the "cross-orientation correlation" (XO) between subbands of variable orientation difference ($\Delta\theta$) at each spatial frequency (f) according to Eq. 1. The panel second from the right in **Figure 2** shows the resulting cross-orientation correlation plotted as a function of $\Delta\theta$ (x-axis) and f (y-axis).

$$XO_{\Delta\theta,f} = \sum_{\theta} \frac{corr\left(w_{\theta,f}, w_{\theta+\Delta\theta,f}\right)}{K} \tag{1}$$

In a similar manner, we also computed the "cross-frequency correlation" (XF) for the difference of a variable pair of spatial frequencies (f and f′) according to Eq. 2. The right-most panel in **Figure 2** shows the resulting cross-frequency correlation plotted as a function of f′ (x-axis) and f (y-axis).

$$XF_{f,f'} = \sum_{\theta} \frac{corr\left(w_{\theta,f}, w_{\theta,f'}\right)}{K} \tag{2}$$

Here, K is the number of orientations, corr stands for the correlation coefficient, and θ is the orientation of the subband.

We did not adopt correlation between "linear" subbands in our analysis because it had an extremely small variation across images (i.e., the variance was approximately 1/256 of that of energy subbands) owing to the narrow bandwidth of the spatial filters that we used, i.e., 30° in orientation and 1 octave in spatial frequency. While the linear cross-scale correlation is closely related to the cross-scale phase statistics and important in representing "edgy" structures in the image (Concetta Morrone and Burr, 1988; Kovesi, 2000; Portilla and Simoncelli, 2000), it plays a small role in texture perception unless one scrutinizes the image at the fovea (Balas, 2006; Balas et al., 2009).

Partial-Least-Squares Regression Analysis

To obtain the regression model for the VEPs and the image statistics of the visual stimulus, we conducted a partial-least-squares regression analysis between them. We assigned the VEPs to the predicator and the image statistics to response variables. We implemented the SIMPLS algorithm through the MATLAB function "plsregress". There were seven components, which minimized the prediction error of the response in a 10-fold cross validation in the training set (The mean squared error of the response was 80.0).

RESULTS

VEPs

Figure 3A shows the average VEPs for all images. Each row shows the results for one image type; i.e., natural textures, PS-synthesized textures, and phase-randomized textures. For all types, large-amplitude VEPs ($\sim$10 μV) were observed at the occipital electrodes (O1/O2). As we did not find any systematic and independent components in the other cortical regions, we here focus on VEPs from those two occipital electrodes.

Figure 3B shows the time course of VEP amplitudes at the occipital electrodes (i.e., the averaged responses from O1 and O2) for the different types of stimuli. The light-blue curves show the average VEPs for the individual images whereas the thick blue curves are the VEPs averaged across all images. The potentials at the occipital electrodes began to rise at 100 ms after the stimulus onset and reached a first small peak at around 120 ms followed by a second large peak at around 250 ms. The basic waveforms were also similar across images, but there were large variations across individual textures.

Correlation Between VEPs and Image Statistics

We conducted a reverse-correlation analysis of the VEPs and each image statistic. We conducted the reverse-correlation analysis

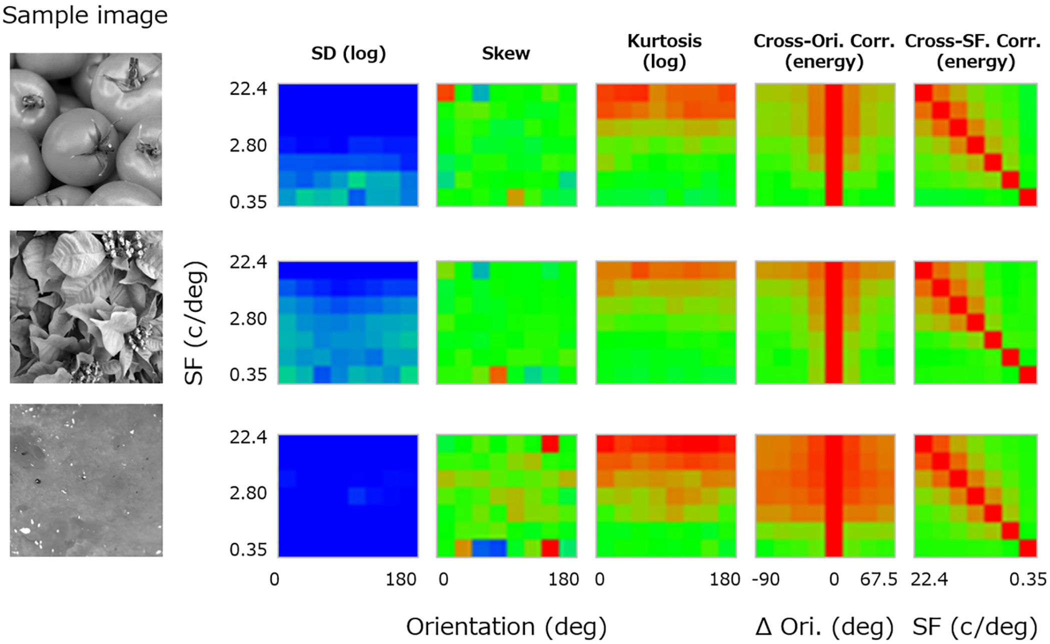


FIGURE 2 | Image statistics calculated for a sample texture image (far left). From the left, the central three panels show the log SD, skewness, and log kurtosis plotted as functions of the spatial frequency and orientation of the subband. The two right-most panels show the cross-orientation energy correlation plotted as a function of the spatial frequency and the orientation (Ori) difference between subbands, and the cross-frequency energy correlation plotted as a function of the spatial frequency (SF) and the paired SF. The color of each pixel represents the value of the statistics, separately scaled for each class of statistics.

for individual observers but the resulting data were noisy and lacking in robustness. This was thought to be because the number of repetitions for each image (24 repetitions) was small for the reverse-correlation analysis. To address this problem, in accordance with the method used in the previous studies (Scholte et al., 2009; Hansen et al., 2011), we computed z-scored VEPs at each time point for each observer and averaged them across observers. We then computed the coefficient of correlation between each image statistic and the z-scored VEP at each time point.

Figure 4 shows the dynamics of the correlations between image statistics and occipital VEPs (O1/O2). Each pixel in the heatmap indicates the coefficient of correlation between the VEPs at a particular timepoint (e.g., 100 ms) and a particular image statistic (e.g., log SD at 0-deg orientation and 2-c/image spatial frequency). Red indicates a positive correlation and blue indicates a negative correlation. Progressing downward, each row shows the results for a class of image statistics; i.e., log SD, skew, log kurtosis, cross-orientation correlation, and cross-frequency correlation. To address the multiple comparisons among time points and image statistic parameters, we adopted the Benjamini–Yekutieli false discovery rate (FDR)-correction method (Benjamini and Yekutieli, 2001). The significant correlations (FDR-corrected, $p < 0.05$) are indicated by vivid colors.

For all classes of image statistics, we found strong correlations with the VEPs that systematically develop over time. For instance, the VEPs had a strong positive correlation with the low-spatial-frequency SDs from ∼100 to ∼150 ms, a negative correlation with the mid-/high-spatial-frequency SDs from ∼150 to ∼180 ms, and a positive correlation with the mid-/high-spatial-frequency SDs from ∼190 to ∼260 ms. Such systematic rises and falls of correlations were found for the other classes of image statistics, with different timing. As we had obtained maps of the correlation dynamics for VEPs from other electrodes (F3, Fz, F4, P7, and P8), we confirmed that they were all similar to, or just sign-reversed from, the results obtained for the occipital electrodes (**Figure 4**).

Correlation Between VEPs and Summarized Image Statistics

The correlation maps shown in **Figure 4** appear somehow redundant. Regarding the moment statistics, for instance, the correlations with VEPs are nearly constant across all absolute orientations, as expressed by vertical "bands" in the maps. For the cross-band correlations, the absolute correlation with VEPs was always higher where the target subbands were close together in orientation (i.e., small $\Delta\theta$) and in spatial frequency (small $|f-f'|$), which is expressed as diagonal spreading on the maps. This is not surprising given that VEPs can hardly resolve a neural response across different absolute orientations. In addition, the absolute orientation plays a small role in the visual appearance of a texture. Accordingly, we calculated the correlations between the VEPs and further summarized measurements, so that we could interpret the temporal dynamics of VEPs correlated with each class of image statistics more easily. To that end, the summarized moment statistics (i.e., log SD, skew, and log kurtosis) were defined as the averages across the orientation for each spatial

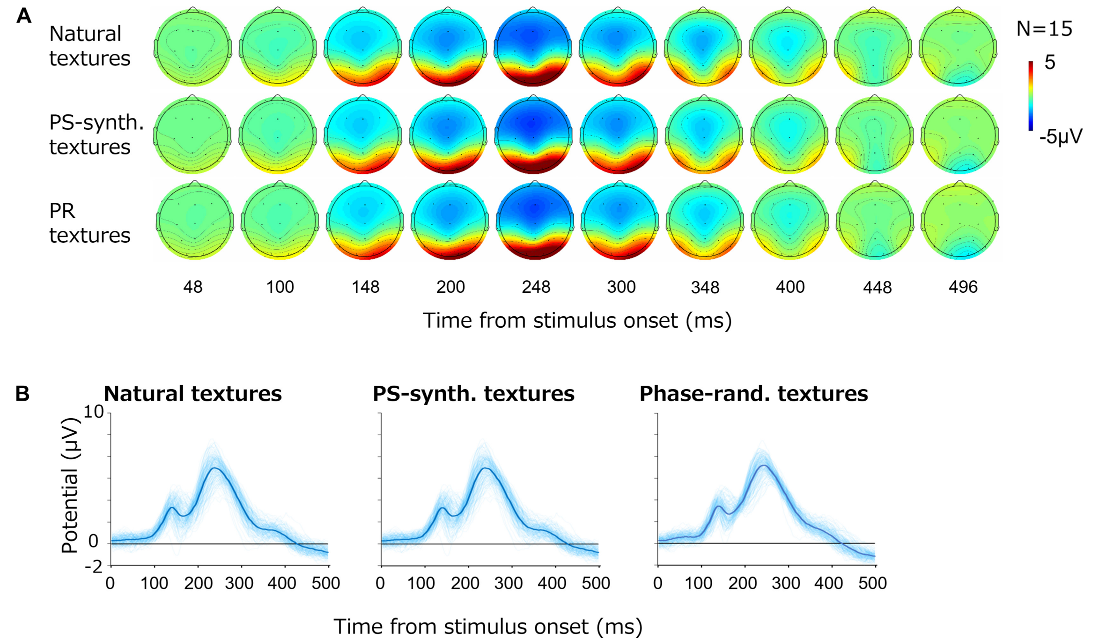


FIGURE 3 | Visual evoked potentials (VEPs) for texture images. **(A)** Topography of the grand average VEPs for the natural textures, PS-synthesized textures, and phase-randomized (PR) textures, in rows from top to bottom. **(B)** VEPs at the occipital electrodes (mean of O1 and O2). The light-blue traces show VEPs for individual images and the thick blue traces represent the averages across images.

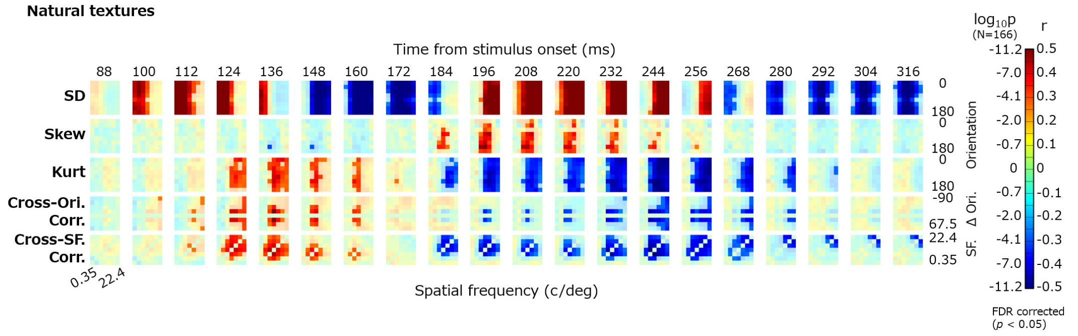


FIGURE 4 | Correlations between image statistics and VEPs. The rows from top to bottom show the correlation of VEPs with the log SD, skew, log kurtosis, cross-orientation correlation, and cross-frequency correlation. Red indicates positive correlations and blue indicates negative correlations. The colors are desaturated for values that are not statistically significant ($p \geq 0.05$, FDR corrected). The format of each panel follows that in **Figure 2**. The maps are arranged in columns for different time points, from 88 to 316 ms. SF, spatial frequency; Ori, orientation; and *r*, correlation coefficient.

frequency. The summarized cross-orientation correlation was given as the average across-orientation difference ($\Delta\theta$ except $\Delta\theta = 0$) for each spatial frequency. The summarized cross-frequency correlation was given as the average across-frequency difference (f-f′ except f = f′).

Figure 5 shows the dynamics of correlation between VEPs and the summary image statistics. The results are shown for the three types of texture stimulus: natural, PS-synthesized, and phase-randomized textures. The vividly colored regions indicate statistically significant correlations identified using the Benjamini–Yekutieli FDR-correction method ($p < 0.05$). Similar patterns of the results were obtained for the other electrodes. We also confirmed that nearly the same results are obtained if we use image statistics calculated within the central or peripheral region in the image.

The temporal development of VEPs correlated with the summary image statistics is now clearly visible. VEPs correlated with SDs were particularly strong ($r_{max} \approx 0.8$) and dynamically rose and fell in a spatial-frequency-dependent manner. They had a first peak at ~120 ms for low-spatial-frequency bands (2–16 c/image), a second negative peak at ~150 ms for middle spatial frequencies (4–64 c/image), and a third peak at ~200 ms for high spatial frequencies (8–128 c/image). VEPs correlated to skewness were observed at ~200 ms only for middle spatial frequencies (16–64 c/image). Even after 300 ms from the stimulus onset, we could observe significant correlations of VEPs to SDs and to some

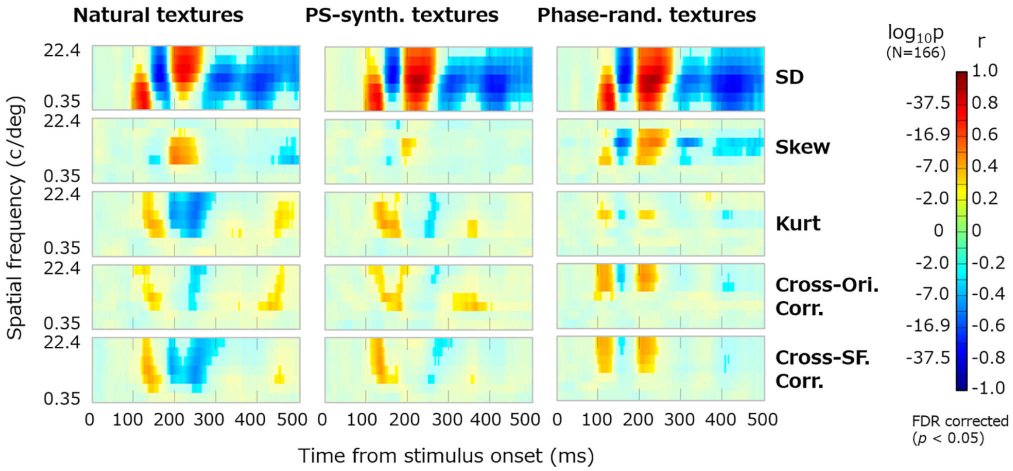


FIGURE 5 | Dynamics of VEPs correlated with summary image statistics. The horizontal axes represent the time from the stimulus onset (0–496 ms) and the vertical axes represent the spatial frequency (c/deg). Reddish pixels indicate positive correlation and blueish pixels indicate negative correlation, for which non-significant data ($p \geq 0.05$, FDR corrected) are desaturated. Panels in the successive rows show the correlations of VEPs with the SD, skew, kurtosis, and cross-orientation energy correlation and cross-frequency energy correlation. The results are shown for natural textures (left), PS-synthesized textures (middle), and phase-randomized textures (right).

other statistics. VEPs correlated to kurtosis, cross-orientation correlation, and cross-frequency correlation appeared to have similar dynamics. They commonly tended to have a first positive peak at ∼150 ms and a second negative peak at ∼200–250 ms, but only for middle and high spatial frequencies. This similarity may be partly due to mutual correlations among the three statistics, which we confirmed not only for our texture stimuli but also for a wide range of natural images. However, as many texture models assume, they have independent roles in the perceptual discrimination of textures, and we confirmed that merging these VEP components prevented us from reconstructing textures from VEP signals.

The temporal dynamics of correlation were qualitatively similar across different types of image, that is, original, PS-synthesized, and phase-randomized images (**Figure 5**). The correlation maps in **Figure 5** are highly correlated with each other; i.e., $r = 0.83$ ($p \approx 0$) for the original and PS-synthesized textures. However, we still found a small difference in the results between the original and PS-synthesized textures despite the equality of image statistics between the two types of texture. We will discuss this difference later in detail.

Reconstruction of Texture Image From the VEP

The series of analyses described above reveal a robust correlation structure between VEPs for natural textures and image statistics. This led us to the hypothesis that image statistics of a texture are predictable from VEP signals. In testing this possibility, we next sought to apply linear regression analysis, to inversely estimate the image statistics of texture stimuli from the VEP signals, and to determine if the estimated image statistics would enable us to synthesize images perceptually similar to the original texture. If such reconstruction was to be successful, it would further support the notion that the temporal pattern of VEPs for natural textures represents the neural processing of perceptually relevant image statistics.

For the purpose of texture synthesis from VEP signals, we adopted the texture statistics used in the PS texture-synthesis algorithm instead of the image statistics used in the above analyses (Note that most PS statistics are essentially equivalent or closely related to the image statistics used in the above reverse-correlation analysis). To construct a linear regression model of PS statistics and VEPs, we used partial-least-squares regression analysis. The number of statistics vectors in the PS texture space is too large to be used in such a regression model, and we therefore reduced PS statistics by applying a compression method inspired by a previous study (Okazawa et al., 2015): we set the number of orientation bands and number of scales each to 3, and the number of positions to 1; rejected the constant parameters; and utilized the symmetrical parameters in the cross-subband correlations. Thereafter, as mentioned in the section "Materials and Methods," we chose to utilize these reduced PS-synthesis (cPS) parameters instead of the original PS statistics. We took VEPs for a period of 0–496 ms (125 points) as the predicator, and the cPS statistics (110 points) as the response variables. The training data set consisted of 299 natural and PS-synthesized texture images used in the experiment (about 90% of all the data), and the test set consisted of the remaining 33 texture images (about 10%). The regression model from the VEPs to the cPS-synthesis parameters was trained on the training set. There were seven components, which minimized the prediction error of the response in a 10-fold cross validation on the training set. Finally, the cPS statistics for the test set were predicted using the trained regression model.

The results indicate that cPS statistics were well predicted by the temporal pattern of VEP signals, suggesting a robust relationship between image statistics and VEPs, as also demonstrated by the reverse-correlation analyses above. R^2

(train) was 0.35 and R^2 (test) was 0.20. The correlation between reconstructed cPS statistics and original cPS statistics was 0.88.

We synthesized textures using the estimated cPS statistics, to perceptually verify the quality of reconstruction. The synthesized textures are shown in **Figure 6**. The images in the upper row show the textures synthesized from the original cPS statistics, and the images in the lower row show textures synthesized with the cPS statistics as estimated from the VEPs. It is found that the VEP-based textures are very similar to, or almost indistinguishable from, the original cPS textures.

To obtain behavioral measures of this perceptual similarity between the original and VEP-based cPS textures, we had five observers (all of whom participated the EEG experiment) rate the quality of the VEP-based cPS textures in a separate experimental block after the EEG recordings. In the experiment, the original cPS textures and VEP-based cPS textures (7.8° × 7.8°) were displayed randomly on the left or right side on a uniform gray background of 40 cd/m^2. The observers inspected the two textures with free viewing and rated their dissimilarity on a five-point scale; that is, from 4 (not similar at all) to 3 (not similar), 2 (similar), 1 (very similar), and 0 (hard to see the difference). For each observer, the rating was done with three repetitions for each of 31 of the 33 textures from the test dataset (The PS-synthesis algorithm did not work for two images). The results showed that the average dissimilarity rating across images was 2.04 (s.e. of 0.22), with an average cross-observer correlation of 0.90. Defining a rating of less than 2.0 as a successful synthesis, 52% of the textures were successfully synthesized from VEPs.

Difference Between Natural and Synthetic Textures

While we observed that the average VEPs were similar among natural, PS-synthesized, and phase-randomized textures (**Figure 3**), we still found differences between the conditions with regard to individual images. **Figure 7** shows the differential VEPs between natural and PS-synthesized textures (**Figure 7A**) and those between PS-synthesized and phase-randomized textures (**Figure 7B**). By means of the statistical test introduced by VanRullen and Thorpe Vanrullen and Thorpe (2001)(i.e., significant if $p < 0.01$ for 15 consecutive periods), we found a significant mean difference between the natural textures and PS-synthesized textures at 148–384 ms and between the PS-synthesized textures and phase-randomized textures at 212–284 ms. Meanwhile, we found a large variation in the differential VEPs across individual images (light-blue traces); i.e., large differential VEPs were found for some images but little or no difference for other images.

What gave rise to these variations in the differential VEPs? Whereas the PS synthesis successfully equalized image statistics in the natural textures for all images, it did not always successfully replicate the appearance of the natural texture and occasionally produced texture images that appeared unnaturalistic. **Figure 7C** shows example textures that produced small (left) and large (right) differential VEPs, on average, from 148 to 248 ms. Especially for the difference between natural and PS-synthesized textures, these pairs of images illustrate that synthesized textures that produced large differential VEPs appeared to be unnatural and perceptually unlike the original natural texture. These observations led us to the notion that variations in the differential VEPs are related to variation in the "unnaturalness" of PS-synthesized textures.

In testing this possibility, we carried out a simple rating experiment to measure the unnaturalness of each PS-synthesized texture in a separate experimental block after the EEG recordings. In that experimental block, all observers who participated in the EEG experiment used a five-point scale to rate how closely each PS-synthesized texture appeared like a photograph of a natural texture (0, almost the same as a natural texture; 1, similar to a natural texture; 2, a little dissimilar to a natural texture; 3, a little unnatural; 4, obviously unnatural). We also asked the observers to rate the unnaturalness of phase-randomized textures, but we found extremely high ratings (unnatural) for almost all images, and we therefore did not use those data in the analysis. The other experimental settings were the same as in the rating experiment for EEG-based texture synthesis.

We then analyzed how the perceptual unnaturalness of a synthesized texture was related to the differential VEP between the natural and PS-synthesized textures. **Figure 7E** shows the dynamics of correlation between the PS-synthesized minus natural differential VEPs and the unnaturalness ratings. Significant correlations ($p < 0.05$, FDR-corrected) were observed at a temporal epoch (168–268 ms) similar to that for the differential VEPs shown in **Figure 7A**. This indicates that PS-synthesized textures that looked unnatural gave rise to VEPs

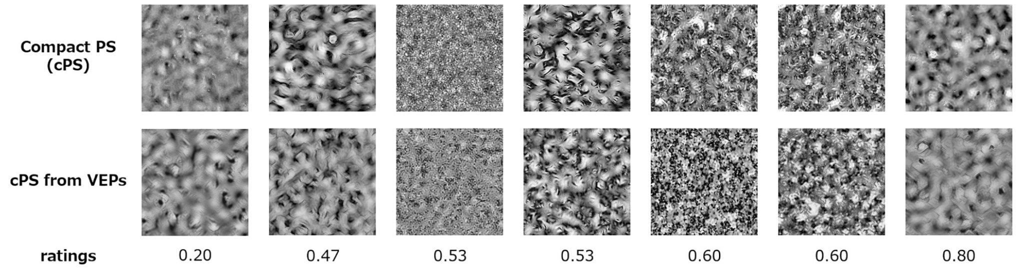


FIGURE 6 | Compact Portilla–Simoncelli (cPS) synthesized textures and compact-PS-synthesized textures with the image statistics as estimated from VEPs. The perceptual dissimilarity ratings (0–4) are given below the images.

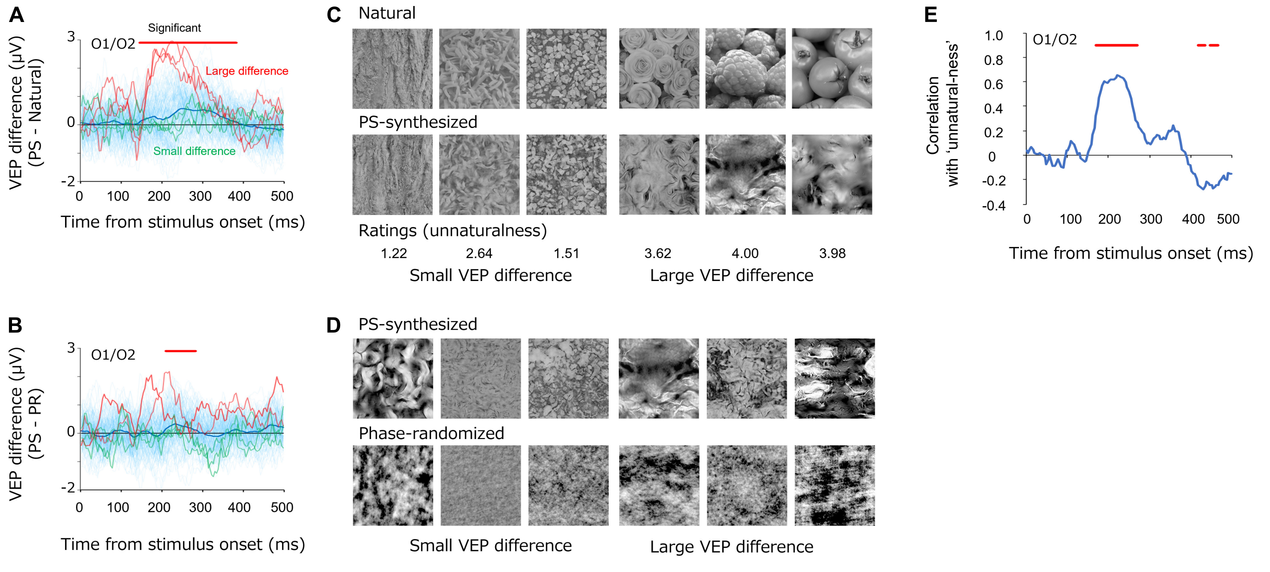


FIGURE 7 | (A,B) Differential VEPs at the occipital electrodes (O1/O2) between the PS-synthesized textures and natural textures **(A)** and between the PS-synthesized textures and phase-randomized textures **(B)**. The light-blue traces are the differential VEPs for each texture. The red traces show the three largest differential VEPs and the green traces show the three smallest differential VEPs. **(C,D)** Pairs of textures that elicited small (left three images) and large (right three images) differential VEPs. Numbers below the images in **(C)** represent the average "unnaturalness" rating of the PS-synthesized texture. **(E)** Correlations between the differential VEPs and the perceptual unnaturalness ratings. The red bars indicate the statistically significant periods ($p < 0.05$, FDR-corrected).

different from those of the original texture, even if they had nearly equal image statistics.

DISCUSSION

The present study investigated the temporal dynamics of cortical responses to biologically plausible image statistics of natural textures, by applying a reverse-correlation analysis between VEPs and image statistics. The analysis revealed that VEPs at the occipital electrodes are systematically correlated with image statistics that are known to be important for human texture perception. Moreover, on the basis of the robust relationship between the VEPs and image statistics, we successfully synthesized textures using image statistics as estimated from VEPs via a linear regression. These results support the notion that the human visual cortex rapidly encodes image statistics that play critical roles in the perception of natural textures. Although small differences were found for images that were not successfully synthesized, similar VEPs and correlation dynamics were observed for synthesized textures that had image statistics equivalent to those of the original natural textures.

Visual evoked potentials that correlated with the subband SD appeared in a spatial-frequency-dependent manner. They first peaked for low spatial frequencies at ∼100 ms after the stimulus onset, then peaked for middle spatial frequencies at ∼150 ms, and finally peaked for high spatial frequencies at ∼200 ms (**Figure 5**). This dynamic shift is consistent with "coarse-to-fine" processing, as suggested by a number of psychophysical studies on object/stereo processing (Schyns and Oliva, 1994; Hegdé, 2008). It is also consistent with physiological findings that magnocellular cells, which are tuned to low spatial frequencies, respond faster than parvocellular cells, which are tuned to high spatial frequencies (e.g., Nowak et al., 1995), and that the spatial frequency tuning of V1 cells shifts in a time-dependent manner from low to high spatial frequencies (Bredfeldt and Ringach, 2002; Mazer et al., 2002).

Visual evoked potentials also correlated with higher-order statistics, such as kurtosis and cross-subband energy correlations, with a similar temporal profile beginning as early as ∼120 ms after the stimulus onset. Considering the nature of each statistic, and past electrophysiological and psychophysical findings regarding texture processing, we speculate that these types of image statistic have a common functional and physiological basis. Kurtosis is primarily associated with spatial sparseness in the energy (complex-cell) outputs of a subband image (Kingdom et al., 2001; Olshausen and Field, 2004). As mentioned earlier, the cross-orientation energy correlations are related to the orientation of local features whereas the cross-frequency energy correlations are related to local luminance modulations (Portilla and Simoncelli, 2000; Balas et al., 2009). Neural computations for each of these three types of measurement are essentially based on inhibitory interactions among cortical neurons across space, orientation, and spatial frequency, respectively (Morrone et al., 1982; Ohzawa et al., 1982; Ferster, 1988; Zipser et al., 1996; Ferster and Miller, 2000; Nishimoto et al., 2006). These interactions are also functionally approximated as the second-order filters proposed in the human texture-vision model; i.e., filters that detect gradients of the energy output of a subband across space, orientation, and spatial frequency (Bergen and Adelson, 1988; Motoyoshi and Kingdom, 2003; Landy and Graham, 2004). It is likely that VEPs correlated with the three image statistics indicate the temporal dynamics of such interactive computations among neural channels in V1 and V2. It is not surprising that VEPs for such higher-order image statistics are observed at latencies as short as or only a little longer than those for SDs (except for very low spatial frequencies), given that the sharp orientation and spatial-frequency tuning of V1 cells emerges from the cross-channel interactions (Morrone et al., 1982; Ohzawa et al., 1982; Ferster and Miller, 2000).

The robust correlational structure between VEPs and image statistics allowed us to reconstruct texture images from image statistics that were inversely estimated from VEPs (**Figure 6**). In the present study, we deliberately applied a linear regression model even though it had lower prediction accuracy, in general, compared with prevailing non-linear "black box" models, including the deep neural network (DNN). Yet, the model we used still had an ability to reconstruct image statistics from occipital VEP signals accurately enough to synthesize textures that were perceptually similar to the target images. These results support the idea that the perceptual appearance of texture is ruled by such image statistics as encoded in the early visual cortex, and that the analysis of simple VEPs can extract these types of information.

While similar results were obtained for the natural and PS-synthesized textures, a small difference in VEP was found for some textures that were less successfully PS-synthesized and appeared "unnatural," even though they had virtually equivalent image statistics (**Figure 7**). When we reanalyzed the dynamic correlations without such mal-synthesized stimuli ("unnaturalness" rating exceeding 3.0), at 88–300 ms after the stimulus onset, the results of the natural images and the PS-synthesized image were closer (with a root-mean-square error of 0.12) than those for the whole visual stimuli (with a root-mean-square error of 0.17). This result further supports the notion that VEPs largely reflect cortical responses to image statistics. However, it is noted that differential VEPs of unnatural textures were clearly observed for the period of 180–250 ms from the stimulus onset. This VEP component indicates that there is a rapid neural processing of information beyond image statistics. We also found significant differences in VEPs between PS-synthesized and phase-randomized textures. According to previous imaging (Freeman et al., 2013) and electrophysiological (Ziemba et al., 2019) studies, these differences could be related to differential neural processing in V1 and V2 for naturalistic textures.

The present study was limited to achromatic natural textures, and the texture image reconstruction was restricted to the texture perception that can be described by image statistics. Despite these limitations, the results of the present study demonstrated that reverse-correlation analysis, which focuses on the holistic features within a relatively large space, enabled us to extract the characteristics of the response of the visual cortex to natural "textures," even with the low spatial resolution of EEG. In principle, the method proposed in the present study is general enough to be applicable to a wide variety of visual stimulus (e.g., natural scenes, materials, and objects) and image features [e.g., the spatial envelope, bags of features (such as the scale-invariant feature transform), and DNN features]. Future studies may extend the approach to better reconstruct the "'impression," using non-linear models such as the DNN. The present study revealed that the impression of natural images is, at least partially, processed in the early visual cortex as statistical features. Therefore, according to the findings of the present study, the impression of a visual stimulus may be summarized as compact features, which would be beneficial in forming the basis for the efficient communication and display of real-world, complex natural images.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee for experiments on humans at the Graduate School of Arts and Sciences, The University of Tokyo. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TO and IM designed the research and wrote the manuscript. TO conducted the experiment and analyzed the data. Both authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by Commissioned Research by NICT (1940101) and JSPS KAKENHI JP15H05916, JP18H04935, and JP20H01782.

REFERENCES

Bach, M., and Meigen, T. (1997). Similar electrophysiological correlates of texture segregation induced by luminance, orientation, motion and stereo. *Vis. Res.* 37, 1409–1414. doi: 10.1016/s0042-6989(96)00322-7

Bach, M., and Meigen, T. (1998). Electrophysiological correlates of human texture segregation, an overview. *Doc. Ophthalmol.* 95, 335–347.

Bach, M., Schmitt, C., Quenzer, T., Meigen, T., and Fahle, M. (2000). Summation of texture segregation across orientation and spatial frequency: electrophysiological and psychophysical findings. *Vis. Res.* 40, 3559–3566. doi: 10.1016/s0042-6989(00)00195-4

Baker, C. L., and Mareschal, I. (2001). Processing of second-order stimuli in the visual cortex. *Prog. Brain Res.* 134, 171–191. doi: 10.1016/s0079-6123(01)34013-x

Balas, B., Nakano, L., and Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *J. Vis.* 9:13. doi: 10.1167/9.12.13

Balas, B. J. (2006). Texture synthesis and perception: using computational models to study texture representations in the human visual system. *Vis. Res.* 46, 299–309. doi: 10.1016/j.visres.2005.04.013

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188.

Bergen, J. R., and Adelson, E. H. (1988). Early vision and texture perception. *Nature* 333, 363–364. doi: 10.1038/333363a0

Bieniek, M. M., Pernet, C. R., and Rousselet, G. A. (2012). Early ERPs to faces and objects are driven by phase, not amplitude spectrum information: evidence from parametric, test-retest, single-subject analyses. *J. Vis.* 12:12. doi: 10.1167/12.13.12

Bredfeldt, C. E., and Ringach, D. L. (2002). Dynamics of spatial frequency tuning in macaque V1. *J. Neurosci.* 22, 1976–1984. doi: 10.1523/jneurosci.22-05-01976.2002

Concetta Morrone, M., and Burr, D. C. (1988). Feature detection in human vision: A phase-dependent energy model. *Proc. R. Soc. Lond. Series B. Biol. Sci.* 235, 221–245. doi: 10.1098/rspb.1988.0073

De Cesarei, A., Loftus, G. R., Mastria, S., and Codispoti, M. (2017). Understanding natural scenes: contributions of image statistics. *Neurosci. Biobehav. Rev.* 74, 44–57. doi: 10.1016/j.neubiorev.2017.01.012

De Valois, R. L., and De Valois, K. K. (1980). Spatial vision. *Annu. Rev. Psychol.* 31, 309–341.

DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *J. Neurophysiol.* 69, 1118–1135. doi: 10.1152/jn.1993.69.4.1118

Ferster, D. (1988). Spatially opponent excitation and inhibition in simple cells of the cat visual cortex. *J. Neurosci.* 8, 1172–1180. doi: 10.1523/jneurosci.08-04-01172.1988

Ferster, D., and Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annu. Rev. Neurosci.* 23, 441–471. doi: 10.1146/annurev.neuro.23.1.441

Fleming, R. W. (2017). Material perception. *Annu. Rev. Vis. Sci.* 3, 365–388.

Freeman, J., and Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nat. Neurosci.* 14, 1195–1201. doi: 10.1038/nn.2889

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., and Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nat. Neurosci.* 16, 974–981. doi: 10.1038/nn.3402

Ghodrati, M., Ghodousi, M., and Yoonessi, A. (2016). Low-level contrast statistics of natural images can modulate the frequency of event-related potentials (ERP) in humans. *Front. Hum. Neurosci.* 10:630. doi: 10.3389/fnhum.2016.00630

Graham, N., Beck, J., and Sutter, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived texture segregation: effects of sign and amount of contrast. *Vis. Res.* 32, 719–743. doi: 10.1016/0042-6989(92)90188-o

Greene, M. R., and Hansen, B. C. (2020). Disentangling the independent contributions of visual and conceptual features to the spatiotemporal dynamics of scene categorization. *J. Neurosci.* 40, 5283–5299. doi: 10.1523/jneurosci.2088-19.2020

Groen, I. I., Ghebreab, S., Lamme, V. A., and Scholte, H. S. (2012a). Low-level contrast statistics are diagnostic of invariance of natural textures. *Front. Comput. Neurosci.* 6:34. doi: 10.3389/fncom.2012.00034

Groen, I. I., Ghebreab, S., Lamme, V. A., and Scholte, H. S. (2012b). Spatially pooled contrast responses predict neural and perceptual similarity of naturalistic image categories. *PLoS Comput. Biol.* 8:e1002726. doi: 10.1371/journal.pcbi.1002726

Groen, I. I., Silson, E. H., and Baker, C. I. (2017). Contributions of low-and high-level properties to neural processing of visual scenes in the human brain. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160102. doi: 10.1098/rstb.2016.0102

Hansen, B. C., Jacques, T., Johnson, A. P., and Ellemberg, D. (2011). From spatial frequency contrast to edge preponderance: the differential modulation of early visual evoked potentials by natural scene stimuli. *Vis. Neurosci.* 28:221. doi: 10.1017/s095252381100006x

Hansen, B. C., Johnson, A. P., and Ellemberg, D. (2012). Different spatial frequency bands selectively signal for natural image statistics in the early visual system. *J. Neurophysiol.* 108, 2160–2172. doi: 10.1152/jn.00288.2012

Heeger, D. J., and Bergen, J. R. (1995). "Pyramid-based texture analysis/synthesis," in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, (New York, NY), 229–238.

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Prog. Neurobiol.* 84, 405–439. doi: 10.1016/j.pneurobio.2007.09.001

Julesz, B. (1965). Texture and visual perception. *Sci. Am.* 212, 38–49.

Kingdom, F. A., Hayes, A., and Field, D. J. (2001). Sensitivity to contrast histogram differences in synthetic wavelet-textures. *Vis. Res.* 41, 585–598. doi: 10.1016/s0042-6989(00)00284-4

Kohler, P. J., Cottereau, B. R., and Norcia, A. M. (2018). Dynamics of perceptual decisions about symmetry in visual cortex. *NeuroImage* 167, 316–330. doi: 10.1016/j.neuroimage.2017.11.051

Kovesi, P. (2000). Phase congruency: a low-level image invariant. *Psychol. Res.* 64, 136–148. doi: 10.1007/s004260000024

Landy, M. S., and Graham, N. (2004). "73 Visual perception of texture," in *The Visual Neurosciences*, eds L. M. Chalupa and J. S. Werner (Cambridge, Mass: MIT Press), 1106–1118.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. *J. Comp. Vis.* 2, 1150–1157.

Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., and Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proc. Natl. Acad. Sci. U.S.A.* 99, 1645–1650. doi: 10.1073/pnas.022638499

Morrone, M. C., Burr, D. C., and Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proc. R. Soc. Lond. Ser B Biol. Sci.* 216, 335–354. doi: 10.1098/rspb.1982.0078

Motoyoshi, I., and Kingdom, F. A. (2003). Orientation opponency in human vision revealed by energy-frequency analysis. *Vis. Res.* 43, 2197–2205. doi: 10.1016/s0042-6989(03)00334-1

Motoyoshi, I., Nishida, S. Y., Sharan, L., and Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature* 447, 206–209. doi: 10.1038/nature05724

Nishida, S. Y. (2019). Image statistics for material perception. *Curr. Opin. Behav. Sci.* 30, 94–99. doi: 10.1016/j.cobeha.2019.07.003

Nishimoto, S., Ishida, T., and Ohzawa, I. (2006). Receptive field properties of neurons in the early visual cortex revealed by local spectral reverse correlation. *J. Neurosci.* 26, 3269–3280. doi: 10.1523/jneurosci.4558-05.2006

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., and Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *J. Vis.* 15:4.

Nowak, L. G., Munk, M. H. J., Girard, P., and Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Vis. Neurosci.* 12, 371–384. doi: 10.1017/s095252380000804x

Ohzawa, I., Sclar, G., and Freeman, R. D. (1982). Contrast gain control in the cat visual cortex. *Nature* 298, 266–268. doi: 10.1038/298266a0

Okazawa, G., Tajima, S., and Komatsu, H. (2015). Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proc. Natl. Acad. Sci. U.S.A.* 112, E351–E360.

Okazawa, G., Tajima, S., and Komatsu, H. (2017). Gradual development of visual texture-selective properties between macaque areas V2 and V4. *Cereb. Cortex* 27, 4867–4880.

Oliva, A., and Torralba, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comp. Vis.* 42, 145–175.

Olshausen, B. A., and Field, D. J. (2004). Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* 14, 481–487. doi: 10.1016/j.conb.2004.07.007

Peterzell, D. H., and Norcia, A. M. (1997). Spatial frequency masking with the sweep-VEP. *Vis. Res.* 37, 2349–2359. doi: 10.1016/s0042-6989(97)00046-1

Portilla, J., and Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comp. Vis.* 40, 49–70.

Rousselet, G. A., Pernet, C. R., Bennett, P. J., and Sekuler, A. B. (2008). Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neurosci.* 9:98. doi: 10.1186/1471-2202-9-98

Scholte, H. S., Ghebreab, S., Waldorp, L., Smeulders, A. W., and Lamme, V. A. (2009). Brain responses strongly correlate with Weibull image statistics when processing natural images. *J. Vis.* 9, 29–29. doi: 10.1167/9.4.29

Schyns, P. G., and Oliva, A. (1994). From blobs to boundary edges: evidence for time-and spatial-scale-dependent scene recognition. *Psychol. Sci.* 5, 195–200. doi: 10.1111/j.1467-9280.1994.tb00500.x

Simoncelli, E. P., and Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24, 1193–1216. doi: 10.1146/annurev.neuro.24.1.1193

Vanrullen, R., and Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *J. Cogn. Neurosci.* 13, 454–461. doi: 10.1162/08989290152001880

Victor, J. D., and Conte, M. M. (1991). Spatial organization of nonlinear interactions in form perception. *Vis. Res.* 31, 1457–1488. doi: 10.1016/0042-6989(91)90125-o

Wallis, T. S., Funke, C. M., Ecker, A. S., Gatys, L. A., Wichmann, F. A., and Bethge, M. (2017). A parametric texture model based on deep convolutional features closely matches texture appearance for humans. *J. Vis.* 17, 5–5. doi: 10.1167/17.12.5

Whitney, D., Haberman, J., and Sweeny, T. D. (2014). "49 From textures to crowds: multiple levels of summary statistical perception," in *The New Visual Neurosciences*, eds J. S. Werner and L. M. Chalupa (Cambridge, MA: MIT Press), 695–710.

Ziemba, C. M., Perez, R. K., Pai, J., Kelly, J. G., Hallum, L. E., Shooner, C., et al. (2019). Laminar differences in responses to naturalistic texture in macaque V1 and V2. *J. Neurosci.* 39, 9748–9756. doi: 10.1523/jneurosci.1743-19.2019

Zipser, K., Lamme, V. A., and Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *J. Neurosci.* 16, 7376–7389.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 29 September 2021
doi: 10.3389/fpsyg.2021.691577

Relationship Between Item and Source Memory: Explanation of Connection-Strength Model

*Junjun Guo[1], Keith Shubeck[2,3] and Xiangen Hu[1,2,3]**

[1] School of Psychology, Central China Normal University, Wuhan, China, [2] Department of Psychology, The University of Memphis, Memphis, TN, United States, [3] Institute for Intelligent Systems, The University of Memphis, Memphis, TN, United States

The controversy in the relationship between item memory and source memory is a focus of episodic memory. Some studies show the trade-off between item memory and source memory, some show the consistency between them, and others show the independence between them. This review attempts to point out the connection-strength model, implying the different types and strengths of the important role of the item–source connections in the relationship between item memory and source memory, which is based on the same essence in the unified framework. The logic of the model is that when item memory and source memory share the same or relevant connection between item and source, they positively connect, or they are independently or negatively connected. This review integrates empirical evidence from the domains of cognition, cognitive neuroscience, and mathematical modeling to validate our hypothesis.

Keywords: item, source, memory, connection-strength, model

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Vanessa R. Simmering,
University of Kansas, United States
Kate Ergo,
Ghent University, Belgium

****Correspondence:***
Xiangen Hu
xiangenhu@gmail.com

Specialty section:
This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: *06 April 2021*
Accepted: *11 August 2021*
Published: *29 September 2021*

Citation:
Guo J, Shubeck K and Hu X (2021) Relationship Between Item and Source Memory: Explanation of Connection-Strength Model. Front. Psychol. 12:691577. doi: 10.3389/fpsyg.2021.691577

INTRODUCTION

Effective retrieval cues play an important role in memory recovery. The effect of these cues is reflected in many research studies on memory: (1) directed forgetting paradigm (Sahakyan and Kelley, 2002); (2) mood-dependent paradigm (Lewis and Critchley, 2003); (3) emotional enhancement effects (Talmi et al., 2019); (4) false memory paradigm (Bookbinder and Brainerd, 2016), and (5) context maintenance and retrieval mathematical model to prove the importance of memory cues (Polyn et al., 2009). These phenomena reflect the implicit decision-making of memory based on effective cues that researchers call "sources" (Johnson et al., 1993), and these memory phenomena belong to declarative memory, which include episodic memory and semantic memory (Squire, 2004; Tulving, 2004).

Episodic memory and semantic memory are two different memory systems that were proposed by Tulving to cover human memory. Tulving believed that the differences between episodic and semantic memory are self-involvement, autonoetic awareness, and subjective sense of time (Tulving, 2004). Such features imply the specific attribute of episodic memory: connections. The differences between episodic and semantic memory are different types of connections. Researchers have confirmed the existence of a semantic network, which is called the spreading-activation theory (Collins and Loftus, 1975). Research shows that semantic memory also has connections, supported by connecting capacity in the hippocampus (Manns et al., 2003; Duff et al., 2020) and that patients with bilateral lesions have an impairment in semantic memory capacity relatively shortly before or after the damage has occurred but not in remote memory for factual knowledge. The relatively smaller damage to semantic memory, compared with episodic memory, in the hippocampus (Vargha-Khadem, 1997) is because semantic information is easier to connect to the semantic network, and this process is relatively automatic and unconscious. This idea will be discussed in section "Different Sources." As a result, episodic memory and semantic memory form a large network, including items and sources, and memory is divided into two categories, item memory and source memory, based on attention allocation (which will be explained in detail in the next section). The former emphasizes the retrieval of an item based on implicit sources, while the latter emphasizes the explicit retrieval of sources.

Item memory usually explains a lot of phenomena including false memory (Reyna, 2000), working memory (Raaijmakers and Schiffrin, 1981), emotional memory (Talmi et al., 2019), and other forms of memory, such as recognition and recall (Johnson, 2005). These memories are all item memories based on "source information." For example, time information is used as a source. Other research studies are focused on factors that affect the formation of source memory (for review, refer to Mather, 2007; Mitchell and Johnson, 2009). Relatively little attention has been paid to the relationship between item memory and source memory. The existing empirical evidence shows the dissociative (Glisky et al., 1995; Davachi et al., 2003; Slotnick et al., 2003), positive (Madan et al., 2017), or negative (Mather et al., 2006) relationship between item memory and source memory. Several research studies show that some factors have different effects on item memory and different source memory, such as attention and emotion (Mather, 2007).

However, there was a lack of a unified theoretical framework to explain the different relationships between item memory and source memory. We suggest that item memory and source memory share the same essence connections.

The connection-strength model emphasizes the important role of connections between item (semantic feature) and sources in the processes of encoding and retrieval of memory, and the relationship between the item and source memory, including positive, negative, or independent relationships.

Hence, we will form and introduce the connection-strength model to explain the relationships between item and source memory in the integrated framework, which is deeply based on item–source connections.

Commonness Among the Different Kinds of "Memories" – Connections

In the study of memory, there are many different forms of memory. From the view of time, the memory includes "working memory," "short-term memory," and "long-term memory" and from the view of retrieval, the memory includes "item memory," "source memory," "context memory," and "associative memory." For Tulving (2002), memory can be divided into "episodic memory" and "semantic memory." For Squire (2004), memory can be divided into "declarative memory" and "non-declarative memory." Without a proper framework, these different types of memories seem to represent different connotations.

However, further reflection shows that these memories share different connections: "working memory," "short-term memory," and "long-term memory" depend on different item-temporal connections; "item memory," "source memory," and "associative memory" depend on similar or different item–sources connections, which will be elucidated in the following sections. "Declarative memory" and "non-declarative memory" are retrieved relatively intentionally or automatically based on connection, respectively.

From the view of Chalfonte and Johnson (1996), there is no essential or inherent difference between "item" and "source," although they seem to be two different concepts. Such dichotomy comes from our attention focus: the focus of attention is called "item," and other information is called "source." For example, in psychological experiments, subjects always treat semantic features as the focus of our attention and treat other perceptual features as secondary information. Otherwise, only when the experimenters asked the subjects to focus on the secondary information, which researchers call source, the perceptual become "item" and "the semantic" become the new source. The "item" and "source" are two points of "connection" that are like a seesaw; sometimes one end is higher, and some other times, the other end is higher.

Consequently, memories are merely the associative network, and "episodic memory," "semantic memory," "source memory," "associative memory," and "context memory" are the subdivisions of such network. Different concepts underlie different endpoints in the connections. Source memory emphasizes retrieving source features, item memory emphasizes the retrieval of semantic features, and associative memory emphasizes the connection between different semantic features. Even emotions can be included in this network: mood-dependent memory (Lewis and Critchley, 2003). The view that emotion is merely one joint in our memory networks that tries to explain emotion-associated memory is called emotional priming.

However, there is an extensive concept or term confusion among these manifestations, such as "item" and "source": sometimes, researchers equate "item" to "semantic meaning"; sometimes, "item" means the episodic concept, the association. Therefore, in the remaining part of the manuscript, we will use the connection: "item"-"source," in which "item" means semantic features, and "source" means other information associated with "semantic." However, in memory systems, "item" in "item memory" means the connection. We use these depictions because most researchers utilize these terms in such a manner. In this logic, all kinds of memory can be divided into two types based on our attention focus. The relationship is shown in **Figure 1**.

Balance Between Item and Source Memory

From the perspective of the experiences of individuals, item memory and source memory are two parts of episodic memory, which are about the semantic and its association with subjective experiences, including when, where, and how an event happened. Source memories are results of questions that ask us to explicitly point out when, where, and how an event happened. The

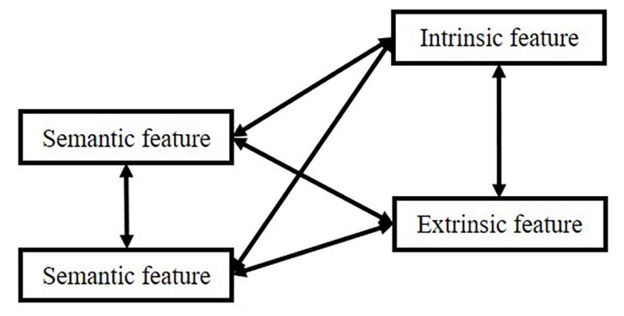


FIGURE 1 | Interactions between items and sources.

differences between item memory and source memory are as follows: the former tries to retrieve the item based on one kind or different kinds of sources, and the latter tries to retrieve one kind of source from the item. Hence, source memory and item memory are phenomena of connections.

From the perspective of the interaction between item memory and source memory, first, the memory of sources contributes to item memory; second, item memory also influences source memory. A wrong starting cue from the connection would impair true memory. This explains why the "feature-conjunction paradigm," which shares features with presented items, would facilitate a false memory (Nie, 2018).

Recognition shows the forms of "recollection," which is associated with the experience of vivid source attributes and "familiarity" without clear source memory (Yonelinas, 1999). Starns and Ksander (2016) used the zROC (the z-transformation of a operator characteristic curve which comes from the ratio between hit rate and false rate) slope to determine the relationship between item memory and source memory. They used three types of experimental conditions: (a) words associated with the source for once (no repetition); (b) words associated with two different kinds of sources (face or animal) (different source repetition); and (c) word and source connection repeats three times (same source repetition). Regardless of the condition, increased item memory confidence enhances the confidence of source memory, which may indicate the item as a cue to effectively retrieving the sources.

These examples reflect only a part of the scientific scenes we want to delineate. Johnson (2005) calls such phenomena "the different task shared the same processes." However, few research has tried to understand the different memories in the unified framework, which only contain two types of memory: item memory and source memory, which combined episodic memory and semantic memory that are based on "connections."

In various studies, there are always different relationships between item memory and source memory: (1) positive; (2) negative; and (3) irrelevant. However, there is a lack of integrated theories to explain such a controversy. Mather (2007) proposed an "object-based attention" framework to explain the better intrinsic source features memory and worse extrinsic source features memory for emotional stimuli. Nevertheless, this theory is insufficient to explain many inconsistent phenomena in memory. Sometimes, positive emotional events and positive context promote associative memory (Fredrickson and Branigan, 2005; Madan et al., 2019). However, sometimes, negative emotional context expands the scope of attention, and positive emotional context reduces the scope of attention (for review, refer to Huntsinger, 2013). In other studies, high or low motivational intensity related to emotion has different effects on the attention process (Harmon-Jones et al., 2011), and different emotions have different effects on attention and memory (Gable and Harmon-Jones, 2010; Harmon-Jones et al., 2011). These effects on attention will be reflected in the process of connection formation (which will be explained in section "Introduction of the Connection-Strength Model").

We suggested that the different types and different strengths of connections between item and source, item and item, and source and source (these can be called item-source connections) play an important role in the relationship between item and source memory. In the next section, we presented the premises of the strength-connection model followed by the introduction of the connection-strength model. In the section "Evidence From Cognition, Cognitive Neuroscience, and Mathematical Models," we collected evidence from domains of cognition, cognitive neuroscience, and mathematical cognitive psychology to validate our "strength-connection model" in interpreting the relationship between item and source memory.

PREMISES OF THE CONNECTION-STRENGTH MODEL

Different kinds of "connections" exist in a memory system, which is the reason for different relationships between item memory and source memory. Factors that affect the formation and extraction of "connections" are as follows: different sources of natural existence, single source or combined source, formation of connections, and extraction of connections.

Different Sources

There are always different source features that can be divided into different types. From the viewpoint of modality (Johnson et al., 1993), there are perceptual, contextual, semantic, and affective sources. From a relevant perspective, sources include the following: (1) external source monitoring, which demands us to discriminate different perceptions; (2) internal source monitoring, which involves two source discriminations in our thoughts; and (3) internal-external source monitoring, such as distinguishing source memory of thoughts from perception. The two classifications are hierarchical: the latter depends on the combination of the former.

From a spatial perspective, Mather (2007) tried to divide source features into two types: object-based source features, intrinsic features that share the same attention scope with the object, extrinsic source features that go beyond items such as context and associative objects. From the timeline, the source is presented before, after, and parallel with the item.

Bellezza and Elek (2018) illuminated that a bundle of two items and two sources in a unit affects different phenomena in memory. Bellezza and Elek presented word pairs with each word in 1 of 4 locations. In the test, subjects were shown one of the paired words and asked to recollect the other word and their locations. The results show that (1) the performance of source-location memory for the cue and the target is equal; (2) the source memory of unrecalled words is above chance; (3) the memory of the cue is associated with that of the target word; and (4) the location of the cue is always confused with that of the target. These results support the fact that item–source connections are always diverse and easy to change. A lot of different information will contribute to retrieval as an effective cue.

Other kinds of sources differ in the information process or extraction process: automatic or intentional encoding, automatic or deliberate retrieval, such as the temporal compared with the neutral environment, the semantic compared with the context. Based on our attention allocation, the process is automatic or strenuous. Divided attention deeply influences the deliberative process compared with the automatic process, either in encoding or retrieval.

Therefore, numerous connections are formed between and among "items" and "different sources." These connections would also influence item memory (for example, Talmi et al., 2019) and source memory (for review, refer to Mather, 2007). Consequently, there are many possibilities for a cue to be called memory.

Formations of the connections are dependent on the three periods as follows.

Three Periods of the Item–Source Formation

Although there are many different kinds of source features, three periods are involved in the formation of the item–source connections: "the perception of source features," "the appraisal of the importance of source features," and "connecting the item and source features." These processes are all indispensable. The structure of item–source formation is demonstrated by the following logic based on relative importance: only when participants try to form a connection that the strength will be higher. The appraisal is the core of these processes.

The effects of adaptive memory and emotional memory can demonstrate the importance of these three processes: when sources that connect to the item are important, the sources will be memorized better. However, when the item is relatively more important than the source, the source memory will be worse memorized. These two processes both impair the formation of connections, because the focus only influences its connections that reflect the importance of appraisal. Kroneisen and Bell (2018) used survival-based or moving-based appraisals for the item and found that the source memory associated with survival processing is better than the other conditions. Consistent effects are found in the memory of the location of foods or potential predators after appraisals of ease of collecting food or capturing wild animals (Nairne et al., 2012). According to the same logic, when the item is important, such as in a contaminated environment (Fernandes et al., 2017), the hunter information for males or gathering information for females (Nairne et al., 2009), the memory will be better.

From the perspective of emotional memory, there is also a trade-off: the emotional material decreases the memory of neutral context (Kensinger et al., 2007; Mather, 2007), and emotional context impairs neutral item memory (Zhang et al., 2015). This emotional enhancement memory trade-off is based on personal goals (for review, refer to Levine and Edelstein, 2009). Relative importance influences appraisal, perception, and connection in a conscious or unconscious form.

Deep empirical evidence will follow in section "Introduction of the Connection-Strength Model."

Goal and Different Processes in Encoding and Retrieval

The encoding and retrieval processes are two important stages in memory. However, there are contradictions between the mechanisms of encoding and retrieval: is the relationship symmetric or asymmetric? Tulving et al. (1994) found that encoding and retrieval are asymmetrical in the hemispheric cortex. However, encoding and retrieval also share the hippocampus (Fritch et al., 2020; Guo and Yang, 2020).

From the viewpoint of the item–source connections, symmetric and asymmetric relationships depend on whether the encoding and retrieval processes share the same item–source connection and the time of detection in different experiments. Such connections may influence the activity of the prefrontal cortex, parietal cortex, and hippocampus. There are types of goal-oriented spatial learning that influence spatial encoding and retrieval in the hippocampus (Turi et al., 2019). Research has found that goals deeply influence human memory in terms of items, sources, and connections (for review, refer to Levine and Edelstein, 2009; Kaplan et al., 2012).

The goals come from two sources: different appraisals of different individuals, attentional locations, and demands of the experimental design. Occasionally, the two kinds of goals compete with each other, and the winner plays an important role in the formation of connection. For example, the emotional context before or after attention always changes attention and memory by motivation or goal (Kaplan et al., 2012). The goals between encoding and retrieval facilitate the common or different processes in item–source connections and memory, which are reflected in the hippocampus (Levita and Muzzio, 2010).

As a result, when encoding and retrieval are based on different connections, it is more difficult to retrieve the item or source. This is mainly derived from Formula (2) and Formula (3).

Presentation of the Connection Models in Memory

The model originates from three existing popular theories: "spreading-activation theory" (Collins and Loftus, 1988), "searched for associative memory" (Raaijmakers and Schiffrin, 1981), and "hybrid model of source monitoring in paired-associates" (Bellezza and Elek, 2018). The description for these models is presented in **Table 1**.

The spreading-activation theory (Collins and Loftus, 1988) emphasizes the connections among different concepts based on experiences of individuals and different connected strengths

TABLE 1 | Old models that emphasize different connections in memory.

Theories	Authors and Cite	Emphasis	Example
Theory1 spreading-action theory	Collins and Loftus, 1975	Semantic features are connected networks.	Some semantic features are more closely related than others.
Theory2 hybrid model of source monitoring	Bellezza and Elek, 2018	The connections among items, sources exist.	One source of the item can be retrieved by the other item.
Theory3 search of associative memory	Raaijmakers and Schiffrin, 1981	Encoding in working memory; and importance of effective cues in retrieval.	In free recall, giving a cue-item always impairs the performance compared to no cued free recall, which means the importance of effective cue.

among them. When a concept is activated, the signal goes along the connection, and other concepts are activated. The stronger the connection, the easier it is to be activated.

The hybrid source monitoring model (Bellezza and Elek, 2018) attempts to explain the connections among items and their source features. Specific experiments are described in the first part. In the list, an item is not only associated with its source features but is also associated with other items and their sources. In the retrieval process, items and sources can also be used as cues to retrieve other items.

Raaijmakers and Schiffrin (1981) proposed a mathematical model to explain how a cue influences the performance of free recall. Such mathematical model is deeply embedded in the theory that retrieval and retrieval cues play an important role in episodic memory recovery, such as recognition and recall (Tulving and Thomson, 1973).

In encoding, connections are formed in working memory and come from the shared time in the capacity of working memory. Working memory cannot simultaneously maintain excessive information. Consequently, only the item and source share the same period in working memory, and connections can be formed.

Therefore, an inappropriate retrieval cue presentation impairs the free recall of a list. For example, when asking subjects to recall a list by a cued item, the recall performance would be worse than the condition of free recall with no cue (Raaijmakers and Schiffrin, 1981). The model emphasizes logic, as illustrated in **Figure 2**.

Raaijmakers and Schiffrin (1981) pointed out that the probability of retrieving the item in the list depends on the strength ratio of "context-item" to "the sum of context-all other items strengths," which can be simplified in the equation:

$$Probability_R(item_i) = \frac{strength_{(context_i - item_1)}}{\{strength_{(context_i - item_1)} + \ldots + strength_{(context_i - item_n)}\}} \tag{1}$$

However, there are limitations in the model: (1) only other items in the list have been considered in whole connections; (2) concepts are ambiguous: "other items" is part of "context," and the real context–temporal, semantic, and cognitive operations are not mentioned in the model; (3) such model is focused on the retrieval of item memory, not on source memory.

Three theories emphasize that information connection exists in our memory systems. In addition, effective retrieval cues are very important in memory recovery. However, these models cannot explain why there are differences between item and source memory, and why there are different relationships between the two memories.

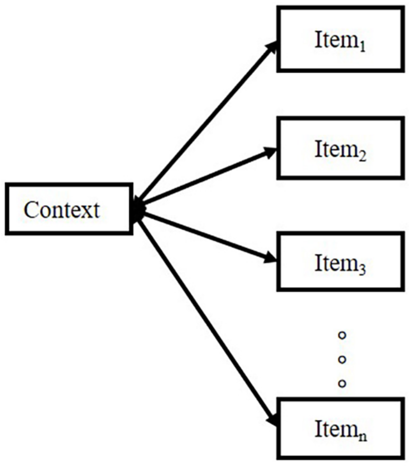


FIGURE 2 | The logic of the model studied by Raaijmakers and Schiffrin (1981).

INTRODUCTION OF THE CONNECTION-STRENGTH MODEL

The item always connects with different kinds of information, including semantic information, perceptual information, contextual information (spatial or temporal information), affective information, and operative track. This information can be integrated into one cluster or one bundle, which we call "object." The "object" involves a series of information that contributes to a whole representation. Although most of the time such integration cannot be realized by our consciousness, Kahneman et al. (1992) found that temporal proximity is very important in object-specific integration. Such capacity even happens in infants (Woodward, 1998). Attention plays an important role in object formation. Logan (1996) pointed out the two processes in the formation of an object: first, perceive the perceptual grouping from spatial proximity, and second, attention chooses several of them to form an object. Such theory is proved by a CODE theory–a mathematical model of visual attention. Attention always has a strong property of "selection" (for review, refer to

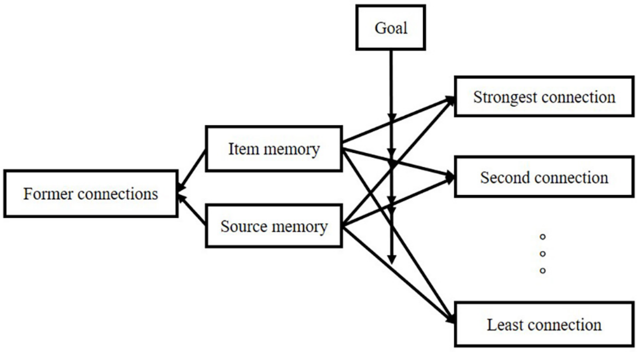


FIGURE 3 | A rough description of the "connection-strength model."

Heinke and Humphreys, 2005), which is very important in the formation of connections.

As a result, many connections are constructed with a difference in strengths. Among these connections, only parts of them are to be used in item memory and source memory. An effective connection is a connection with higher strength. The main differences between item memory and source memory can be seen in the connections they call. The relatively significant difference is that connections in item memory are sometimes combinations of item-sources connections. The memory retrieval probabilities for items and sources are presented in Equations (2) and (3). However, connection for source memory is a specific connection, or may be mediated by other items.

$$Probability_R(item_i) = \frac{\{strength_{(item_i - source_1)} + \ldots + strength_{(item_i - source_k)}\}}{\{strength_{(item_i - source_1)} + \ldots + strength_{(item_i - source_n)}\}} \quad (2)$$

$$Probability_R(source_i) = \frac{strength_{(source_i - item_i)}}{\{strength_{(source_i - source_1)} + \ldots + strength_{(source_i - source_n)}\}} \quad (3)$$

SI_i means the employed source-item strength in source memory decision making, and $\sum_{t=1}^{n} SS_t$ means the sum of all the strengths of connections related to the "source", whether the connection is formed in experiments or from past experiences. $P_R\,(source_i)$ means the probability of a specific source retrieval.

A rough description of the model is shown in **Figure 3**.

Item retrieval follows these principles: (1) performance of item retrieval depends on the connection between item and source; (2) there are two kinds of connection: item-one-source connection and item-sources connection; (3) the higher the connection strength, the more likely it is to be called; (4) the connection strength comes from encoding period; (5) appraisal and attention are very important in the formation of connection-strength; (6) in item-sources connection, our different goal would assign a different weight to a different connection in a different experimental design. The structure is shown in **Figure 4**.

Source retrieval follows these principles: (1) source retrieval depends on the ratio of a specific connection to the sum of all associated connections; (2) there are three types of connection: item-specific source connection, other item-the specific source connection, and other source-the specific source connection; (3) all the connections come from two stages: formation in the experiment and formation out of the experiment; (4) appraisal and attention are very important for the formation of connection-strength; (5) effective cues with higher strength are important for the performance of retrieval. The higher the connection strength, the more likely it is to be called. The structure is shown in **Figure 5**.

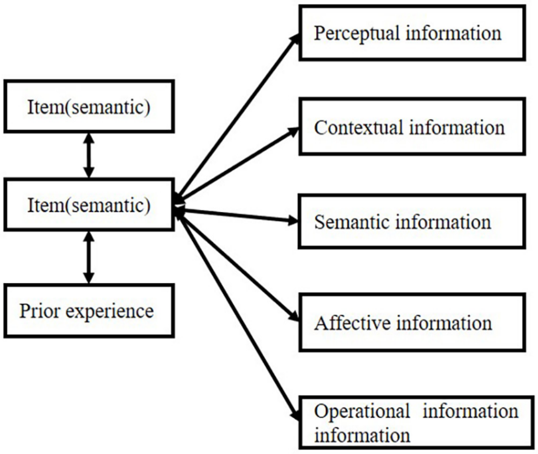


FIGURE 4 | Connection-strength model in item retrieval.

The essence of item memory and source memory is the same, which depends on the connection between item and source(s) and the ratio of the called connection(s) to the sum of all associated connections. The main difference is that focus shifts from item to source, and vice versa.

The relationship between item memory retrieval and source memory retrieval is: when the connection required by source memory and item memory is the same, there is a positive correlation between them; when item memory and source memory depend on different connections, for

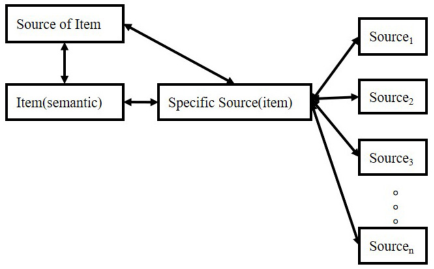


FIGURE 5 | Retrieval of source memory.

example, item memory depends on item-temporal connection, while source memory depends on item-color connection, item memory and source memory are independent of each other; when the connection required by item memory affects the formation of source memory connection, there is a negative correlation between item memory and source memory. These can be identified in the probability equation in equations (2) and (3): when the numerator of the ratio increases at the same time, the positive relationship appears, otherwise, the negative relationship will appear. When the numerator is irrelevant to each other, they are independent.

There is a lot of evidence to support our model: first, the performance of source memory is always worse than that of item memory, because the connection required by the source is a part of the connections required by item memory, such as affective state as the source; second, several source memories are better than others because they call stronger connections, such as object-based colors, compared with other associated objects; third, attention plays an important role in connection formation, the more attention paid to the connection, the higher the strength of the connection; fourth, experimental design affects which connection will be encoded and called by item memory and source memory; fifth, mathematical models show that both item memory and source memory are continuous processes, which indicates that the different connection-strength deeply affects the performance of item memory and source memory. A mathematical model also shows that the confidence of item memory and source memory influences each other (Starns et al., 2013).

The empirical evidence supporting these phenomena is as follows.

Different Experimental Designs

The two types of experimental designs are considerably different. First, studies presenting recognition and source judgments for the same item in immediate succession have revealed chance-level accuracy in source memory with no recognition. Second, studies presenting a block of recognition followed by a block of source judgments have revealed above-chance accuracy in source memory with no recognition (for review, refer to Fox and Osth, 2020).

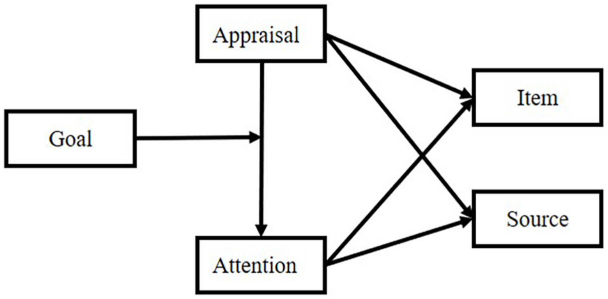


FIGURE 6 | Attention interaction between item and source.

The essence behind these phenomena is that subjects call different connections. When source judgments occur in the block after all the recognition has finished, other items associated with unrecognized items will also activate the source without recognition. Fox and Osth (2020) used a simultaneous, blocked, and reversed blocked design to demonstrate this idea.

Kim et al. (2012) found negative effects of item repetition on source memory. Experimenters show the items in two phases. In phase 1, line drawings present varying numbers of items; in phase 2, each item is associated with a critical new source. The results show that the more repetition in phase 1, the more difficult it is to memorize the critical new source in phase 2. This can be explained by the connection-strength model. The more repetition of the items in list 1, the more item–source (temporal source) in list 1, which induces a relatively weaker item-critical new sources strength ratio. Osth et al. (2018) found a list-strength effect, which means a proportion of items are strengthened to observe the effect on non-strengthened items in source memory but not in item memory. This effect can also be explained by our model.

EVIDENCE FROM COGNITION, COGNITIVE NEUROSCIENCE, AND MATHEMATICAL MODELS

Empirical evidence supports the connection-strength model, including evidence from cognition, cognitive neuroscience, and mathematical models.

Evidence in Cognition

Attention in Unitization

Attention plays the most important role in forming connections (for review, refer to Mather, 2007; Block and Gruber, 2014). The relationship is shown in **Figure 6**.

The object-based attention model (Mather, 2007) and space-based attention model (Belardinelli, 2016) support the importance of attention in the connection of different features in an object or space. However, not all connections are formed and stored in memory; for example, intrinsic source feature memory is always better than extrinsic source feature memory. This is because the intrinsic source features of the object share scope of attention with the object, which makes it easier to form a connection. These relationships are reflected in emotional enhancement memory (for review, refer to Talmi et al., 2019) and enhanced intrinsic source memory (for review, refer to Mather, 2007). Simultaneously, memories of extrinsic source features are impaired because they are beyond the core of attention.

Thus, the first evidence in cognition comes from the operation of attention to enhance the connection between the item and the source. The difficulty level of unitization deeply affects the connection between the item and the source, and several factors affect the difficulty: "pre-experimental associations" (e.g., Giovanello et al., 2006; Rhodes and Donaldson, 2008; Ford et al., 2010), "experimentally instructed encoding strategy" (Haskins et al., 2008; Bader et al., 2010; Parks and Yonelinas, 2015), and characteristics of the source. These factors are related to the same concept—unitization. The former and the latter reduce the difficulty of attention integration; the second is used as an effective strategy to enhance connection formation.

In research, the performance of aging declines in episodic memory (Friedman, 2013), semantic memory (Bertola et al., 2019; Venneri et al., 2019), source memory (Schacter et al., 1991), and associative memory (Greene and Naveh-Benjamin, 2020). Researchers suggest that these phenomena arise from a reduced binding capacity (Li et al., 2005). Unitization can reduce this tendency. Zheng et al. (2016) used two conditions: unitized condition (imaging the color as the internal parts of item) and non-unitized condition (imaging the color as the context), and then asked subjects to complete the source memory test. The results show that the difference in source memory between young and old people is smaller under the condition of unitization. Boywitt and Meiser (2012) found that under the condition of incidental attention, the source-source connection disappears in the extrinsic source features; and that this connection is preserved in intentional attention.

Further research (Kinjo, 2011) shows that the damage of extrinsic source memory is greater than that of intrinsic source memory because the former needs more attention.

Emotional Memory Enhancement Effects

The formation of connection plays an important role in emotional reinforcement memory, which is based on two different kinds of connection, what researchers call "organization" and "emotional context." Emotion enhances item–item connection and item–source connection. When emotional materials are highly clustering with each other, memory is always better than low-clustering materials (Talmi et al., 2007). Talmi and Moscovitch (2004) found that semantic relatedness was considerably important in list item memory: the memory of semantic-related neutral words is not worse than that of emotional words. Talmi et al. (2019) set up a retrieved-context model to explain emotional enhancement in memory. The emotional context maintenance and retrieval model points out that the emotional enhancement memory effect is based on the enhanced item-source connection that is associated with ever-changing temporal and emotional context. Consequently, the cues of organization and context play an important role in emotion-enhanced effects.

Interaction Between Item Memory and Source Memory

For intrinsic source features, item memory is positively correlated with source memory. This is because the intrinsic source features share an attention system with the item. Therefore, the enhancement of item memory is positively related to the enhancement of intrinsic source memory, which is deeply reflected in emotional items (Mather, 2007).

The important trade-off between the item and extrinsic source is very important for connection formation. Attention to the item enhances the connections associated with "this item," such as the temporal and color, rather than extrinsic sources (e.g., other items) that are beyond the item. Attention to "source" transfers the "source" to "item" and facilitates connections for the "new item." The new connections increase the denominator of the probability formula, thus reducing the ratio of the item to the extrinsic source. Therefore, there is a negative relationship between item memory and source memory. Such phenomena are more deeply reflected in the memory related to emotion.

Source memory can be divided into two types: intrinsic source memory, which is the features of the item itself, and extrinsic source memory, which is the associated features outside the item, including the context and objects that are associated with the item. Numerous studies support two opposite phenomena: compared with neutral items, emotional items always facilitate intrinsic source memory; however, emotional items always interfere with extrinsic source memory.

The trade-off between emotional items and emotional source memory validates our theory. The first evidence comes from the relative importance of the item. Compared with positive emotional stimuli, the extrinsic source memory of negative emotional stimuli is worse (Madan et al., 2019). Compared with low-arousal emotional stimuli, the extrinsic source memory of higher arousal stimuli is worse (Mather et al., 2006; Kensinger et al., 2007; Mather, 2007). The second evidence comes from the relative importance of the source. Compared with a neutral context, an emotional context is always remembered better at the cost of neutral items (Maratos et al., 2001; Maratos and Rugg, 2011; Chiu et al., 2013). We are always attracted by the emotional context even when asked to keep the attention on items, especially in an extremely important environment, such as the source memory of cheaters (Kroneisen and Bell, 2013) and goal-inconsistency phenomena (Bell et al., 2012). The third evidence comes from a relative comparison between items and sources. When an item is emotional, item memory is better at the cost of neutral context (Kensinger et al., 2007), and researchers ask participants to remember words in the emotional context and find the impaired word memory influenced by the emotional context (Zhang et al., 2015). The fourth evidence comes from the reappraisal that can change the trade-off memory between items and sources. Steinberger et al. (2011) found that different reappraisals of items and contexts would facilitate different memory trade-offs.

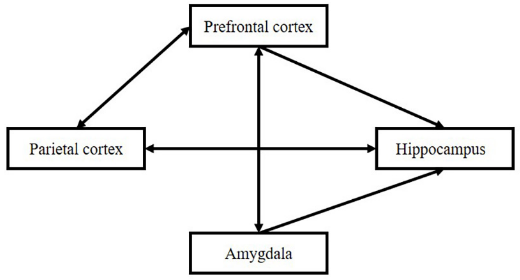


FIGURE 7 | Interactions among subregions of the brain.

Evidence From Cognitive Neuroscience

The formation of the item–source strength depends on three periods, "perception," "evaluation," and "association," which can be reflected in the function of the brain: the parietal cortex, prefrontal cortex, and hippocampus. The hippocampus is the core structure in Tulving's episodic memory. Its main function is to connect different types of information. At the same time, it is affected by different brain regions, such as the parietal cortex, frontal cortex, and amygdala. The relationship is shown in **Figure 7**.

Parietal Cortex in Perception

Only when a stimulus is perceived can a connection be formed. The posterior parietal cortex plays an important role in perception, such as speech, visual motion (Buchsbaum et al., 2010), and tactile perception (Ro et al., 2004). Transcranial magnetic stimulation of the right parietal lobe disrupts the perception of briefly presented stimuli (Howard et al., 2019). Damage to the intraparietal cortex impairs action and perception (Medina et al., 2020).

The parietal cortex is always a bridge between perception, action, and cognition (Gottlieb, 2007), which is associated with spatial attention and the "salience representation" of the external world that is relevant to us. The parietal cortex also combines with the amygdala to perceive biological motions (Bonda et al., 1996). Recently, researchers found that the parietal cortex plays an important role in transsaccadic memory and the integration of visual object features (Dunkley et al., 2016).

Extensive research has shown that cathodal stimulation on the left posterior parietal cortex decreases the retrieval performance in source memory (Chen et al., 2016. Guidotti et al. (2019) found that predictive activity in the parietal cortex predicts source-memory decisions: the greater the decision evidence, the greater the activation in the parietal cortex. The posterior eye field (Müri et al., 1996) and the ratio of posterior-anterior medial temporal lobe volumes can predict the performance of source memory (Snytte et al., 2020). Transcranial direct current stimulation of the parietal cortex decreases false recognition increases item and source memory accuracy compared with the situation with no stimulation (Pergollzzi and Chua, 2017), and improves associative memory (Vuli et al., 2021). The activation of the parietal cortex is associated with our confidence in memory: the activation of the old is larger than the new, and the perceived is larger than the imagined (King and Miller, 2017). Functional MRI (fMRI) shows parietal cortex plays an important role for information connections, especially in the object (von Stein et al., 1999), and lesions in parietal cortex will eliminate this processing (Decoteau and Kesner, 1998). Ben-Zvi et al. (2015) identified that parietal lesions impair associated learning, including word pairs, picture pairs, and picture–sound pairs.

Prefrontal Cortex in Evaluation

Classic research (Buschman and Miller, 2007) illustrates the role of the prefrontal cortex in the activation of posterior parietal cortices for top–down and bottom–up attention, which supports the idea that attention is the momentary enhanced reaction potential of the perceptual response (Berlyne, 1951). The low-frequency synchrony between the frontal and parietal cortices reflects top–down attention, and higher frequencies are associated with bottom–up attention. The prefrontal cortex plays an important role in the behavioral approach and inhibition processes (Sutton and Davidson, 1997), which is supported by observations of an activation of a "hot spot" at the cost of lateral inhibition through the call of norepinephrine (Mather et al., 2016). The interaction between the prefrontal cortext and parietal cortext is essential in the process of inhibition and activation (Buschman and Miller, 2007).

The prefrontal cortex has two main functions: "appraisal" and "attention allocation." The appraisal is reflected in the situation, which reflects the evaluation of importance among different kinds of information. Researchers call this effect the "appraisal-by-content model" (Dixon et al., 2017). The model points out that different areas of the prefrontal cortex are responsible for different kinds of input, including outside perceptions, episodic memories, future events, viscera sensory, action, and emotions. Based on the appraisal, the prefrontal cortex decides where attention should be located.

This effect is reflected in the processing of emotional stimuli. In the study, "A cognitive-motivational analysis of anxiety" (Mogg and Bradley, 1998) "appraisal" of the events plays an important role in negative attention bias in the population of "anxiety." Anxiety always shows a higher threat appraisal of negative stimuli than a healthy population. As a result, people with anxiety attempt to focus on negative stimuli because of a high threat appraisal of the negative stimuli; However, non-anxious groups think that negative stimuli do not pose a threat, and consequently they don't pay attention to them, but continue to do what they are doing (Huntsinger, 2013). Many studies support appraisal bias in attention and memory (Ma et al., 2017; Foley, 2018). An fMRI shows that the prefrontal cortex in anxiety controls attention to threat-related stimuli (Bishop et al., 2004).

Increasing research supports the function of appraisal in the prefrontal cortex. Kalisch et al. (2006) found that the medial prefrontal cortex plays an important role in the high-level appraisal of emotional materials. The decreased function of the medial prefrontal cortex in Alzheimer's patients will reduce their evaluation of their cognitive ability, especially memory (Ries et al., 2012). An explanation of the aversive would modulate the activation of appraisal in the medial prefrontal cortex (Mechias et al., 2009).

The prefrontal cortex plays an important role in the formation of connections. Research has found that normal aging and prefrontal cortex lesions are associated with poor performance in item memory and source memory (Swick et al., 2006). The activation of the medial prefrontal cortex contributes to the item and source memory of self-referent information compared with other referent information, which also reflects the role of appraisal (Leshikar and Duarte, 2012). Furthermore, prefrontal deficits impair episodic memory in patients with schizophrenia (Ragland et al., 2009). The subregions of the prefrontal cortex, left frontopolar cortex, left mid-ventrolateral region, left mid-dorsolateral region, and anterior cingulate cortex contribute commonly to working memory, semantic memory, and episodic memory (Nyberg et al., 2003).

Hippocampus in Association

The hippocampus is significant in connecting items and their sources. Dalton et al. (2018) used fMRI and found that the hippocampus cooperates with other areas to support the associative processes and scene constructions, which imply the binding ability of the hippocampus. Further research has found that the hippocampus is particularly important for the building of association across stimulus domains, such as combining visual features with auditory features (Borders et al., 2017). Implicit associative learning engages the hippocampus and interacts with explicit associative learning (Degonda et al., 2005).

Nordin et al. (2017) compared the performance of associative memory with the volume of the anterior hippocampus between middle-aged and older patients. The results show that the older population has poorer associative memory, which is accompanied by a smaller volume of the anterior hippocampus and less activation compared with the younger population. Iwasaki et al. (2021) found that beta oscillations in the hippocampus could forecast the performance of object-location associative memory. In mice, an increase in beta oscillations in the hippocampus during the encoding process would come along with better "source memory." The hippocampus of primates and humans contains spatial view neurons, which provide a representation of locations in the viewed space. Neuronal networks in the hippocampus activate together to form episodic memory, especially recent events that involve relations (Giovanello et al., 2010). Gradual changes in hippocampal activity are crucial in remembering the order of events (Manns et al., 2007).

Other studies have shown that the activation of the amygdala due to emotional stimuli processing impairs the connection between emotion items and source features by disrupting the function of the hippocampus (Roozendaal et al., 2009; Madan et al., 2017).

Association Among the Parietal Cortex, Prefrontal Cortex, and Hippocampus

The prefrontal hippocampus circuit is significant in associative memory. Different units are activated by different item presentations in the prefrontal cortex and hippocampus according to time for monkeys. The research demonstrates that both the prefrontal cortex and the hippocampus contribute to feature binding according to the timeline (Cruzado et al., 2020). Interactions between the prefrontal cortex and the hippocampus are particularly important for reactivating memories and their contexts to contribute to memory retrieval and assimilate the new memories-item-source connection to our schemas (Preston and Eichenbaum, 2013). The recovery of extinct fear memory in a special context requires both the prefrontal cortex and the hippocampus (Milad et al., 2007). Prefrontal hippocampal interactions are obvious during the encoding of new memories (Takehara-Nishiuchi, 2020). Two processes may exist in such interactions: first, the formation of new information into the old memory networks; and second, the formation of different types of information into unification. For the process of retrieval, prefrontal-hippocampal interaction is also found in rats when rats try to decide where they should go inside a maze (for humans, refer to Öztekin et al., 2009; for animals, refer to Cholvin et al., 2016). Attention from the prefrontal cortex influences the activation of the hippocampus (Córdova et al., 2019).

Prefrontal-parietal connections are also particularly important. Bor and Seth (2012) pointed out the significance of the prefrontal-parietal network in attention, working memory, and chunking. The prefrontal cortex influences the parietal cortex by enhancing attention to specific perceptions involved in source feature processing (Katsuki et al., 2015; Sofia and Gregoriou, 2017).

Amygdala's activation decreases or enhances the function of hippocampus (Madan et al., 2017) and interaction between parietal cortex and amygdala influences association forming (Kesner, 2000).

The research also found the combined contribution of the prefrontal cortex, parietal cortex, and hippocampus to working memory retrieval (Öztekin et al., 2009).

The different activation of areas that come along with source memory and episodic memory in neural imaging and neural physiology depends on the time course of detection. When detection time occurs simultaneously in one of the three proposed periods, perception, evaluation, and binding (association), the results would show that the activation of different areas in the brain supports episodic or source memory.

Evidence From Mathematical Models

Several mathematical models attempt to explain the relationship between item memory and source memory. To answer this question, two important questions must be clarified. First, how is the experiment designed? Second, what processes are included in item memory or source memory? The different experiment designs and "processes" will influence the relationship between the item and source memory, which is based on the same or different connections in the model that we propose.

These models include the "multinomial processing tree," "receiver operating characteristic analysis," "context maintenance and retrieval model," and "bivariate signal detection model."

However, a detailed description of the mathematical theory is beyond the scope of this article. We merely concentrate on how the mathematical models influence our knowledge of understanding the "connection-strength" model and validate our theories.

Contrary to the abovementioned evidence, mathematical models focus on retrieval processing in item and source memory. The essence that is for both encoding and retrieval is "item–source connection," where encoding is trying to form the connection and retrieval is trying to recover the "connection" with a cue or cues. The following shows the understanding of the relationship by deploying the connection in our retrieval.

Multinomial Processing Tree for Source Memory

Batchelder and Riefer (1990) used the multinomial processing model to explain the phenomena of source monitoring and presented item detection, source identification, and guessing bias parameters. The model is based on the hypothesis that the retrieval of source memory only occurs when an item is detected. Bell et al. (2016) supported this hypothesis: in retrieval, items and distractors are randomly present, and researchers first ask subjects to make the recognition followed by source decision. The results showed no source memory with no recognition. Additionally, source memory is always worse than item memory. Such phenomena come from the experimental design-simultaneous presentation described in section "Introduction," which cannot be treated as a theory but as the hypothesis rooted in the experimental design. The advanced multidimensional source model (Meiser and Bröder, 2002) shows that item memories are associated with source detection memories. Recollection is always associated with the joint memory of different source attributes, such as color and position; however, different sources are independent of each other in the condition of familiarity.

The multinomial processing tree model also supports different types of item–source connections with different strengths. Not all sources can be included in an item-source network (Dodson et al., 1998; Klauer and Wegener, 1998). Source memories of pictures are always better than visual words, and the source memory of the self-referent is better than other-referent information (Riefer et al., 1994). Meiser and Bröder (2002) showed different degrees of difficulty in source monitoring: when sources are similar, discrimination between sources is difficult.

Receiver Operating Characteristic (ROC) Analysis

Receiver operating characteristic curves are used as indices of whether the memory processes are based on the threshold criteria. The basic procedures behind the ROC are the first subjects to decide on the memory and then tell their confidence.

This topic focuses on three themes: whether two kinds of memory, "recollection" and "familiarity," share the same process; whether the item memory and source memory share the same process; and how item memory and source memory influence each other.

Some studies state that "recollection" and "familiarity" share different processes: "recollection" is a threshold process, and "familiarity" is a graded process. Jacoby (1991) identified that "recollection" was deeply influenced by attention compared with "familiarity," and had the same effect on older adults (Jacoby et al., 2005) and individuals with amnesia (Kensinger and Corkin, 2008). However, the main evidence comes only from neuroscience. This is because connections in "recollection" are more difficult to form and more easily influenced by other factors. From the viewpoint of source monitoring, researchers believe that the two processes are both based on the graded experiences of the subjects in asssociation with combined source information.

An excellent model, which is computational and based on neurobiology, was deployed by Elfman et al. (2008). The research found that the features of sources play a significant role: recollection will fit a threshold model when sources are considerably distinct and a continuous model when there is similarity (feature overlap) in sources. Such research finds that different activation of the hippocampus is extremely crucial in the encoding of different kinds of memory: distinct sources are associated with higher activity in the hippocampus, and lower distinct sources are associated with lower activity in the hippocampus. This classic research may solve the conflict in theory by the different activation in the hippocampus, which implies continuous item memory regardless of recollection or familiarity. The use of different source materials (In these types of experiments, variables such as pictures, words, varying colors, and auditory input are all significant predictors of forms in operator characteristic curves) in the experiments is significant in ROC analysis. In the experiment, Slotnick (2010) presented objects on the left or right position of the screen and asked subjects to remember. Recollection-based ROCs are formed by source memory confidence ratings connecting to judge "remember" or the highest item confidence rating response. The results of the ROCs show that recollection-based ROCs identify the hypothesis of continuous models. This evidence shows that the recollection and familiarity of recognition both follow continuous processing. Onyper et al. (2010) used the new model, "some-or-none," to emphasize the importance of the continuous model. This is because the researchers cannot integrate the data from "words" and "travel scenes" by the dual-process signal detection theory or unequal-variance signal detection model (the threshold or continuous model). Researchers combine the dual-process model and continuous model into the "variable-recollection dual-process model," which suggests that familiarity and recollection are based on the continuous process. However, the difference between familiarity and recollection is a variable criterion. The above studies mainly focus on how source memory affects item memory. Starns et al. (2013) used zROC slopes to explore the relationship between item memory and source memory. The experimenters presented the item with different sources, female or male voice (strong voice-presenting four times; weak voice-presenting one time), for different times and then asked the subjects how many times the item was present and what confidence they believed for the source. The results indicate that the confidence of item memory strengthens the confidence of source memory, which is called "the converging criteria account." The essence is underlain by the experimental design: the repetition to present item, the source would simultaneously increase the strength of item–source connections.

For the question of whether the retrieval of item memory and source memory shares the same process, researchers answer this question. Slotnick and Dodson (2005) found that recognition and source memory are both continuous by removing non-diagnostic source information in the analysis. In the experiment, 160 words were presented by female or male voices and then were combined with 80 new words to ask subjects to evaluate the confidence of the item and source memory. This supports the fact that model-item memory and source memory share the same mechanism.

Context Maintenance and Retrieval Model: How Context Influences Item Memory

Howard and Kahana (2002) focused on how context influences item memory. Howard and Kahana (2002) first pointed out the temporal context model (TCM) to explain the well-known phenomena in human memory: the recency effect and contiguity effect. TCM considers the temporal context as the cue to retrieve the item; for the recency effect, the slightly changed item-retrieval temporal context, compared with the encoding context, contributes to recency, and the similar temporal context between continuous items supports the contiguity effect. Second, Sederberg et al. (2008) further evidenced the context-based theory of recency and contiguity by simulating the internal contextual state as an effective cue to retrieval that goes beyond the information of the time, which comes from a combination of different contextual information. Polyn et al. (2009) pointed out in context maintenance and retrieval model: the semantic source which comes from longstanding semantic association among words, the temporal source which reflects the presentation of sequences and the modality source which is the presentation form, all contribute to item retrieval. The interaction among the three parts jointly contributes to retrieval as an effective cue in free recall. This model can also explain the enhanced emotional memory by increasing the connection between emotional items and contexts (Talmi et al., 2019).

Bivariate Signal Detection Model

The "bivariate signal detection model" is an effective mathematical model for identifying the connection-strength model, which underlines existing different item–source connections. The "bivariate signal detection model," as the special case of "multidimensional signal detection theory," is the extension of the signal detection theory. The model considers that the mental state is large and noisy, and every action needs a judgment, and judgments can be modeled as a random sample from a multivariate probability distribution that reflects individual perceptual space. Decision-making always depends on the different axis mapping from the space in distribution.

Banks (2010) proposed that item memory and source memory share a single analytic model. The multidimensional signal detection theory states that recognition memory and source memory depend on the projection of the multidimensional configuration onto an appropriate unidimensional axis, which is deployed as evidence to make memory decisions. This explanation is consistent with our strength model, which implies that source memory and recognition share the same or different item–source connections; however, they are based on the same mechanism. Researchers have used mathematical model bivariate signal detection to demonstrate this hypothesis. The experiment presents two kinds of words (words and first names) along with two kinds of sources (visual and auditory). Subsequently, the displayed items and new items are presented on the screen, and subjects should give the confidence of the item and source memory. The results, which are analyzed in bivariate signal detection theory, show that different tasks (recognition, source memory) can be performed based on different decision axes from the projection of multidimensional configuration. In addition, the results show the orthogonal relationship between item recognition memory and source memory, which deploys different connections. However, recognition and source memory use the same memory database that employs different information connections.

CONCLUSION

According to Johnson, memory can be divided into two types: information in the focus of our attention and information out of the focus of our attention. This dichotomy integrates a memory system into a new perspective, dividing it into item memory and source memory.

Different item memory and source memory depend on the same or different item–source connections. As a result, the relationship between item and source memory is positive, negative, or irrelevant.

Different item-source calls depend on the strength of the item-source(s) connection, which means that when the connection between item and source is stronger, the probability of retrieval of item memory or source memory is greater. However, when goals are proposed, different weights are added to different item-source connections; the memory then changes.

In different environments, there are different goals that affect the choice of connections. This explains why item memory and source memory are either consistent or inconsistent.

Cognitive processes and brain mechanisms affect the formation and intensity of connections. From the perspective of the cognitive process, attention allocation and the appraisal of the importance of items and sources would influence the formation of the item–source connection. From the perspective of brain mechanisms, the prefrontal cortex, parietal cortex, and hippocampus are associated with perception, appraisal, and connection formation, respectively. From the perspective of experimental design, the single source memory decision-making after item-random presents or block-items present determines the relationship between the item and source memory. The mathematical models support the hypothesis of the connection-strength model.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

JG was responsible for the title selection, literature collection, and writing. XH proposed the suggestions for title selection and manuscript writing. KS provided input and feedback and helped with writing. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Science Foundation of China (NSFC, #61937001).

REFERENCES

Bader, R., Mecklinger, A., Hoppstadter, M., and Meyer, P. (2010). Recognition memory for one-trial-unitized word pairs: evidence from event-related potentials. *NeuroImage* 50, 772–781. doi: 10.1016/j.neuroimage.2009.12.100

Banks, W. P. (2010). Recognition and source memory as multivariate decision processes. *Psychol. Sci.* 11, 267–273. doi: 10.1111/1467-9280.00254

Batchelder, W. H., and Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychol. Rev.* 97, 548–564. doi: 10.1037/0033-295X.97.4.548

Belardinelli, A. (2016). *Object-Based Attention: Cognitive and Computational Perspectives. From Human Attention to Computational Attention*. New York, NY: Springer New York.

Bell, R., Buchner, A., Kroneisen, M., and Giang, T. (2012). On the flexibility of social source memory: a test of the emotional incongruity hypothesis. *J. Exp. Psychol. Learn. Memory Cogn.* 38, 1512–1529. doi: 10.1037/a0028219

Bell, R., Mieth, L., and Buchner, A. (2016). Emotional memory: no source memory without old-new recognition. *Emotion* 17, 120–130. doi: 10.1037/emo0000211

Bellezza, F. S., and Elek, J. K. (2018). A hybrid model of source monitoring in paired-associates learning. *J. Exp. Psychol. Learn. Memory Cogn.* 45, 1042–1065. doi: 10.1037/xlm0000639

Ben-Zvi, S., Soroker, N., and Levy, D. A. (2015). Parietal lesion effects on cued recall following pair associate learning. *Neuropsychologia* 73, 176–194. doi: 10.1016/j.neuropsychologia.2015.05.009

Berlyne, D. E. (1951). Attention, perception and behavior theory. *Psychol. Rev.* 58, 137–146. doi: 10.1037/h0058364

Bertola, L., Ávila, Rafaela, T., Bicalho, M., and Malloy-Diniz, L. F. (2019). Semantic memory, but not education or intelligence, moderates cognitive aging: a cross-sectional study. *Braz. J. Psychiatry* 41, 535–539. doi: 10.1590/1516-4446-2018-0290

Bishop, S., Duncan, J., Brett, M., and Lawrence, A. D. (2004). Prefrontal cortical function and anxiety: controlling attention to threat-related stimuli. *Nat. Neurosci.* 7, 184–188. doi: 10.1038/nn1173

Block, R. A., and Gruber, R. P. (2014). Time perception, attention, and memory: a selective review. *Acta Psychol.* 149, 129–133. doi: 10.1016/j.actpsy.2013.11.003

Bonda, E., Petrides, M., Ostry, D., and Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J. Neurosci.* 16, 3737–3744. doi: 10.1523/jneurosci.16-11-03737.1996

Bookbinder, S. H., and Brainerd, C. J. (2016). Emotion and false memory: the context-content paradox. *Psychol. Bull.* 142, 1315–1351. doi: 10.1037/bul0000077

Bor, D., and Seth, A. K. (2012). Consciousness and the prefrontal parietal network: insights from attention, working memory, and chunking. *Front. Psychol.* 3:63. doi: 10.3389/fpsyg.2012.00063

Borders, A. A., Aly, M., Parks, C. M., and Yonelinas, A. P. (2017). The hippocampus is particularly important for building associations across stimulus domains. *Neuropsychologia* 99, 335–342. doi: 10.1016/j.neuropsychologia.2017.03.032

Boywitt, C. D., and Meiser, T. (2012). The role of attention for context–context binding of intrinsic and extrinsic features. *J. Exp. Psychol. Learn. Memory Cogn.* 38, 1099–1107. doi: 10.1037/a0026988

Buchsbaum, B. R., Hickok, G., and Humphries, C. (2010). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn. Sci.* 25, 663–678. doi: 10.1207/s15516709cog2505_2

Buschman, T. J., and Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315, 1860–1862. doi: 10.1126/science.1138071

Chalfonte, B. L., and Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory Cogn.* 24, 403–416. doi: 10.3758/BF03200930

Chen, N. F., Lo, C. M., Liu, T. L., Juan, C. H., Muggleton, N. G., and Cheng, S. K. (2016). Source memory performance is modulated by transcranial direct current stimulation over the left posterior parietal cortex. *NeuroImage* 139, 462–469. doi: 10.1016/j.neuroimage.2016.06.032

Chiu, Y., Dolcos, F., Gonsalves, B. D., and Cohen, N. J. (2013). On opposing effects of emotion on contextual or relational memory. *Front. Psychol.* 4:103. doi: 10.3389/fpsyg.2013.00103

Cholvin, T., Loureiro, M., Cassel, R., Cosquer, B., Herbeaux, K., Vasconcelos, A. D., et al. (2016). Dorsal hippocampus and medial prefrontal cortex each contribute to the retrieval of a recent spatial memory in rats. *Brain Structure Funct.* 221, 91–102. doi: 10.1007/s00429-014-0894-6

Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407–428. doi: 1037/0033-295X.82.6.407

Collins, A. M., and Loftus, E. F. (1988). A spreading-activation theory of semantic processing. *Read. Cogn. Sci.* 82, 126–136. doi: 10.1037/0033-295X.82.6.407

Córdova, N. I., Turk-Browne, N. B., and Aly, M. (2019). Focusing on what matters: modulation of the human hippocampus by relational attention. *Hippocampus* 29, 1025–1037. doi: 10.1002/hipo.23082

Cruzado, N. A., Tiganj, Z., Brincat, S. L., Miller, E. K., and Howard, M. W. (2020). Conjunctive representation of what and when in monkey hippocampus and lateral prefrontal cortex during an associative memory task. *Hippocampus* 30, 1332–1346. doi: 10.1002/hipo.23282

Dalton, M. A., Zeidman, P., McCormick, C., and Maguire, E. A. (2018). Differentiable processing of objects, associations and scenes within the hippocampus. *J. Neurosci.* 38, 8146–8159. doi: 10.1523/jneurosci.0263-18.2018

Davachi, L., Mitchell, J. P., and Wagner, A. D. (2003). Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proc. Natl. Acad. Sci. U S A.* 100, 2157–2162. doi: 10.1073/pnas.0337195100

Decoteau, W. E., and Kesner, R. P. (1998). Effects of hippocampal and parietal cortex lesions on the processing of multiple-object scenes. *Behav. Neurosci.* 112, 68–82. doi: 10.1037/0735-7044.112.1.68

Degonda, N., Mondadori, C. R., Bosshardt, S., Schmidt, C. F., Boesiger, P., Nitsch, R. M., et al. (2005). Implicit associative learning engages the hippocampus and interacts with explicit associative learning. *Neuron* 46, 505–520. doi: 10.1016/j.neuron.2005.02.030

Dixon, M. L., Thiruchselvam, R., Todd, R., and Christoff, K. (2017). Emotion and the prefrontal cortex: an integrative review. *Psychol. Bull.* 143, 1033–1081. doi: 10.1037/bul0000096

Dodson, C. S., Holland, P. W., and Shimamura, A. P. (1998). Onthe recollection of specific- and partial-source information. *J. Exp. Psychol. Learn. Memory Cogn.* 24, 1121–1136. doi: 10.1037/0278-7393.24.5.1121

Duff, M. C., Covington, N. V., Hilverman, C., and Cohen, N. J. (2020). Semantic memory and the hippocampus: revisiting, reaffirming, and extending the reach of their critical relationship. *Front. Hum. Neurosci.* 13:471. doi: 10.3389/fnhum.2019.00471

Dunkley, B. T., Baltaretu, B., and Crawford, J. D. (2016). Trans-saccadic interactions in human parietal and occipital cortex during the retention and comparison of object orientation. *Cortex* 82, 263–276. doi: 10.1016/j.cortex.2016.06.012

Elfman, K. W., Parks, C. M., and Yonelinas, A. P. (2008). Testing a neurocomputational model of recollection, familiarity, and source recognition. *J. Exp. Psychol. Learn. Memory Cogn.* 34, 752–768. doi: 10.1037/0278-7393.34.4.752

Fernandes, N. L., Pandeirada, J. N. S., Soares, S. C., and Nairne, J. S. (2017). Adaptive memory: the mnemonic value of contamination. *Evol. Hum. Behav.* 38, 451–460. doi: 10.1016/j.evolhumbehav.2017.04.003

Foley, M. A. (2018). Reflecting on how we remember the personal past: missing components in the study of memory appraisal and theoretical implications. *Memory* 26, 634–652. doi: 10.1080/09658211.2017.1387667

Ford, J. H., Verfaellie, M., and Giovanello, K. S. (2010). Neural correlates of familiarity-based associative retrieval. *Neuropsychologia* 48, 3019–3025. doi: 10.1016/j.neuropsychologia.2010.06.010

Fox, J., and Osth, A. F. (2020). Does source memory exist for unrecognized items? *PsyArXiv* [Preprint]. doi: 10.31234/osf.io/brpwu

Fredrickson, B. L., and Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cogn. Emot.* 19, 313–332. doi: 10.1080/02699930441000238

Friedman, D. (2013). The cognitive aging of episodic memory: a view based on the event-related brain potential. *Front. Behav. Neuroence* 7:111. doi: 10.3389/fnbeh.2013.00111

Fritch, H. A., Macevoy, S. P., Thakral, P. P., Jeye, B. M., and Slotnick, S. D. (2020). The anterior hippocampus is associated with spatial memory encoding. *Brain Res.* 1732, 146696. doi: 10.1016/j.brainres.2020.146696

Gable, P. A., and Harmon-Jones, E. (2010). The effect of low versus high approach-motivated positive affect on memory for peripherally versus centrally presented information. *Emotion* 10, 599–603. doi: 10.1037/a0018426

Giovanello, K. S., Keane, M. M., and Verfaellie, M. (2006). The contribution of familiarity to associative memory in amnesia. *Neuropsychologia* 44, 1859–1865. doi: 10.1016/j.neuropsychologia.2006.03.004

Giovanello, K. S., Schnyer, D. M., and Verfaellie, M. (2010). A critical role for the anterior hippocampus in relational memory: evidence from an fMRI study comparing associative and item recognition. *Hippocampus* 14, 5–8. doi: 10.1002/hipo.10182

Glisky, E. L., Polster, M. R., and Routhieaux, B. C. (1995). Double dissociation between item and source memory. *Neuropsychology* 9, 229–235. doi: 10.1037/0894-4105.9.2.229

Gottlieb, J. (2007). From thought to action: the parietal cortex as a bridge between perception, action, and cognition. *Neuron* 53, 9–16. doi: 10.1016/j.neuron.2006.12.009

Greene, N. R., and Naveh-Benjamin, M. (2020). A specificity principle of memory: evidence from aging and associative memory. *Psychol. Sci.* 31, 316–331. doi: 10.1177/0956797620901760

Guidotti, R., Tosoni, A., Perrucci, M. G., and Sestieri, C. (2019). Choice-predictive activity in parietal cortex during source memory decisions. *NeuroImage* 189, 589–600. doi: 10.1016/j.neuroimage.2019.01.071

Guo, D., and Yang, J. (2020). Interplay of the long axis of the hippocampus and ventromedial prefrontal cortex in schema-related memory retrieval. *Hippocampus* 30, 263–277. doi: 10.1002/hipo.23154

Harmon-Jones, E., Harmon-Jones, C., Amodio, D. M., and Gable, P. A. (2011). Attitudes toward emotions. *J. Personal. Soc. Psychol.* 101, 1332–1350. doi: 10.1037/a0024951

Haskins, A. L., Yonelinas, A. P., Quamme, J. R., and Ranganath, C. (2008). Perirhinal cortex supports encoding and familiarity-based recognition of novel associations. *Neuron* 59, 554–560. doi: 10.1016/j.neuron.2008.07.035

Heinke, D., and Humphreys, G. W. (2005). "Computational models of visual selective attention: a review," in *Connectionist Models in Psychology*, ed. G. W. Houghton (Hove: Psychological Press), 273–312.

Howard, C. J., Boulton, H., Bedwell, S. A., Boatman, C. A., Roberts, K. L., and Mitra, S. (2019). Low-frequency repetitive transcranial magnetic stimulation to right parietal cortex disrupts perception of briefly presented stimuli. *Perception* 48, 346–355. doi: 10.1177/0301006619834251

Howard, M. W., and Kahana, M. J. (2002). A distributed representation of temporal context. *J. Mathemat. Psychol.* 46, 269–299. doi: 10.1006/jmps.2001.1388

Huntsinger, J. R. (2013). Does emotion directly tune the scope of attention? *Curr. Direct. Psychol. Sci.* 22, 265–270. doi: 10.1177/0963721413480364

Iwasaki, S., Sasaki, T., and Ikegaya, Y. (2021). Hippocampal beta oscillations predict mouse object-location associative memory performance. *Hippocampus* 31, 503–511. doi: 10.1002/hipo.23311

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *J. Memory Lang.* 30, 513–541. doi: 10.1016/0749-596X(91)90025-F

Jacoby, L. L., Bishara, A. J., Hessels, S., and Toth, J. P. (2005). Aging, subjective experience, and cognitive control:dramatic false remembering by older adults. *J. Exp. Psychol. General* 134, 131–148. doi: 10.1037/0096-3445.134.2.131

Johnson, M. K. (2005). The relation between source memory and episodic memory: comment on Siedlecki et al. (2005). *Psychol. Aging* 20, 529–531. doi: 10.1037/0882-7974.20.3.529

Johnson, M. K., Hashtroudi, S., and Lindsay, D. S. (1993). Source monitoring. *Psychol. Bull.* 114, 3–28. doi: 10.1037/0033-2909.114.1.3

Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* 24, 175–219. doi: 10.1016/0010-0285(92)90007-O

Kalisch, R., Wiech, K., Critchley, H. D., and Dolan, R. J. (2006). Levels of appraisal: a medial prefrontal role in high-level appraisal of emotional material. *NeuroImage* 30, 1458–1466. doi: 10.1016/j.neuroimage.2005.11.011

Kaplan, R. L., Ilse, V. D., and Levine, L. J. (2012). Motivation matters: differing effects of pre-goal and post-goal emotions on attention and memory. *Front. Psychol.* 3:404. doi: 10.3389/fpsyg.2012.00404

Katsuki, F., Saito, M., and Constantinidis, C. (2015). Influence of monkey dorsolateral prefrontal and posterior parietal activity on behavioral choice during attention tasks. *Eur. J. Neurosci.* 40, 2910–2921. doi: 10.1111/ejn.12662

Kensinger, E. A., and Corkin, S. (2008). "Amnesia: point and counterpoint," in *Learning Theory and Behavior. Learning and Memory: A Comprehensive Reference*, Vol. 1, eds J. Byrne and R. Menzel (Amsterdam: Elsevier), 259–286. doi: 10.1007/BF00157784

Kensinger, E. A., Garoff-Eaton, R. J., and Schacter, D. L. (2007). Effects of emotion on memory specificity: memory trade-offs elicited by negative visually arousing stimuli. *J. Memory Lang.* 56, 575–591. doi: 10.1016/j.jml.2006.05.004

Kesner, R. P. (2000). Behavioral analysis of the contribution of the hippocampus and parietal cortex to the processing of information: interactions and dissociations. *Hippocampus* 10, 483–490. doi: 10.1002/1098-1063(2000)10:43.0.CO;2-Z

Kim, K., Yi, D., and Johnson, M. K. (2012). Negative effects of item repetition on source memory. *Memory Cogn.* 40, 889–901. doi: 10.3758/s13421-012-0196-2

King, D. R., and Miller, M. B. (2017). Influence of response bias and internal/external source on lateral posterior parietal successful retrieval activity. *Cortex* 91, 126–141. doi: 10.1016/j.cortex.2017.04.002

Kinjo, H. (2011). Effects of aging and divided attention on recognition memory processes for single and associative information. *Psychol. Rep.* 108, 405–419. doi: 10.2466/04.10.22.PR0.108.2.405-419

Klauer, K. C., and Wegener, I. (1998). Unraveling socialcategorization in the "Who said What?" paradigm. *J. Personal. Soc. Psychol.* 75, 1155–1178. doi: 10.1037/0022-3514.75.5.1155

Kroneisen, M., and Bell, R. (2013). Sex, cheating, and disgust: enhanced source memory for trait information that violates gender stereotypes. *Memory* 21, 167–181. doi: 10.1080/09658211.2012.713971

Kroneisen, M., and Bell, R. (2018). Remembering the place with the tiger: survival processing can enhance source memory. *Psychonomic Bull. Rev.* 25, 667–673. doi: 10.3758/s13423-018-1431-z

Leshikar, E. D., and Duarte, A. (2012). Medial prefrontal cortex supports source memory accuracy for self-referenced items. *Soc. Neurosci.* 7, 126–145.

Levine, L. J., and Edelstein, R. S. (2009). Emotion and memory narrowing: a review and goal-relevance approach. *Cogn. Emot.* 23, 833–875. doi: 10.1080/02699930902738863

Levita, L., and Muzzio, I. A. (2010). Role of the hippocampus in goal-oriented tasks requiring retrieval of spatial versus non-spatial information. *Neurobiol. Learn. Mem.* 93, 581–588. doi: 10.1016/j.nlm.2010.02.006

Lewis, P. A., and Critchley, H. D. (2003). Mood-dependent memory. *Trends Cogn. Sci.* 7, 431–433. doi: 10.1016/j.tics.2003.08.005

Li, S. C., Naveh-Benjamin, M., and Lindenberger, U. (2005). Aging neuromodulation impairs associative binding neurocomputational account. *Psychol. Sci.* 16, 445–450. doi: 10.1111/j.0956-7976.2005.01555.x

Logan, G. D. (1996). The code theory of visual attention: an integration of space-based and object-based attention. *Psychol. Rev.* 103, 603–649. doi: 10.1037/0033-295X.103.4.603

Ma, S. T., Abelson, J. L., Okada, G., Taylor, S. F., and Liberzon, I. (2017). Neural circuitry of emotion regulation: effects of appraisal, attention, and cortisol administration. *Cogn. Affect. Behav. Neurosci.* 17, 437–451. doi: 10.3758/s13415-016-0489-1

Madan, C. R., Fujiwara, E., Caplan, J. B., and Sommer, T. (2017). Emotional arousal impairs association-memory: roles of amygdala and hippocampus. *NeuroImage* 156, 14–28. doi: 10.1016/j.neuroimage.2017.04.065

Madan, C. R., Scott, S. M. E., and Kensinger, E. A. (2019). Positive emotion enhances association-memory. *Emotion* 19, 733–740. doi: 10.1037/emo0000465

Manns, J. R., Hopkins, R. O., and Squire, L. R. (2003). Semantic memory and the human hippocampus. *Neuron* 38, 127–133. doi: 10.1016/S0896-6273(03)00146-6

Manns, J. R., Howard, M. W., and Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. *Neuron* 56, 530–540. doi: 10.1016/j.neuron.2007.08.017

Maratos, E. J., Dolan, R. J., Morris, J. S., Henson, R. N., and Rugg, M. D. (2001). Neural activity associated with episodic memory for emotional context. *Neuropsychologia* 39, 910–920. doi: 10.1016/s0028-3932(01)00025-2

Maratos, E. J., and Rugg, M. D. (2011). Electrophysiological correlates of the retrieval of emotional and non-emotional context. *J. Cogn. Neurosci.* 13, 877–891. doi: 10.1162/08989290l753165809

Mather, M. (2007). Emotional arousal and memory binding: an object-based framework. *Perspect. Psychol. Sci.* 2, 33–52. doi: 10.1111/j.1745-6916.2007.00028.x

Mather, M., Clewett, d., Sakaki, M., and Harley, C. W. (2016). Norepinephrine ignites local hotspots of neuronal excitation: how arousal amplifies selectivity in perception and memory. *Behav. Brain Sci.* 39. doi: 10.1017/S0140525X15000667

Mather, M., Mitchell, K. J., Raye, C. L., Novak, D., Greene, E. J., and Johnson, M. K. (2006). Emotional arousal can impair feature binding in working memory. *J. Cogn. Neurosci.* 18, 614–625. doi: 10.1162/jocn.2006.18.4.614

Mechias, M., Etkin, A., and Kalisch, R. (2009). A meta-analysis of instructed fear studies: Implications for conscious appraisal of threat. *NeuroImage* 49, 1760–1768. doi: 10.1016/j.neuroimage.2009.09.040

Medina, J., Jax, S. A., and Coslett, H. B. (2020). Impairments in action and perception after right intraparietal damage. *Cortex* 122, 288–299. doi: 10.1016/j.cortex.2019.02.004

Meiser, T., and Bröder, A. (2002). Memory for multidimensional source information. *J. Exp. Psychol. Learn. Memory Cogn.* 28, 116–137. doi: 10.1037//0278-7393.28.1.116

Milad, M. R., Wright, C., Orr, S. P., Pitman, R. K., Quirk, G. J., and Rauch, S. L. (2007). Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biol. Psychiatry* 62, 446–454. doi: 10.1016/j.biopsych.2006.10.011

Mitchell, K. J., and Johnson, M. K. (2009). Source monitoring 15 years later: what have we learned from fMRI about the neural mechanisms of source memory? *Psychol. Bull.* 135, 638–677. doi: 10.1037/a0015849

Mogg, K., and Bradley, B. P. (1998). A cognitive-motivational analysis of anxiety. *Behav. Therapy* 36, 809–848. doi: 10.1016/S0005-7967(98)00063-1

Müri, R., Iba-Zizen, M. T., Derosier, C., Cabanis, E. A., and Pierrot-Deseilligny, C. (1996). Location of the human posterior eye field with functional magnetic resonance imaging. *J. Neurol. Neurosurg. Psychiatry* 60, 445–448. doi: 10.1136/jnnp.60.4.445

Nairne, J. S., Pandeirada, J. N. S., Gregory, K. J., Van, and Arsdall, J. E. (2009). Adaptive memory: fitness relevance and the hunter-gatherer mind. *Psychol. Sci.* 20, 470–476. doi: 10.1111/j.1467-9280.2009.02356.x

Nairne, J. S., Vanarsdall, J. E., Pandeirada, J., and Blunt, J. R. (2012). Adaptive memory: enhanced location memory after survival processing. *J. Exp. Psychol. Learn. Memory Cogn.* 38, 495–501. doi: 10.1037/a0025728

Nie, A. (2018). Facial recall: feature–conjunction effects in source retrieval versus item recognition. *Percept. Mot. Skills* 125, 369–386.

Nordin, K., Herlitz, A., Larsson, E. M., and Söderlund, H. (2017). Overlapping effects of age on associative memory and the anterior hippocampus from middle to older age. *Behav. Brain Res.* 317, 350–359. doi: 10.1016/j.bbr.2016.10.002

Nyberg, L., Marklund, P., Persson, J., Cabeza, R., and Ingvar, M. (2003). Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologia* 41, 371–377. doi: 10.1016/S0028-3932(02)00168-9

Onyper, S. V., Zhang, Y. X., and Howard, M. W. (2010). Some-or-none recollection: evidence from item and source memory. *J. Exp. Psychol. General* 139, 341–364. doi: 10.1037/a0018926

Osth, A. F., Fox, J., McKague, M., Heathcote, A., and Dennis, S. (2018). The list strength effect in source memory: data and a global matching model. *J. Memory Lang.* 103, 91–113. doi: 10.31219/osf.io/ky9ax

Öztekin, I., McElree, B., Staresina, B. P., and Davachi, L. (2009). Working memory retrieval: contributions of the left prefrontal cortex, the left posterior parietal cortex, and the hippocampus. *J. Cogn. Neurosci.* 21, 581–593. doi: 10.1162/jocn.2008.21016

Parks, C. M., and Yonelinas, A. P. (2015). The importance of unitization for familiarity-based learning. *J. Exp. Psychol. Learn. Memory Cogn.* 41, 881–903. doi: 10.1037/xlm0000068

Pergollzzi, D., and Chua, E. F. (2017). Transcranial direct current stimulation over the parietal cortex alters bias in item and source memory tasks. *Brain Cogn.* 108, 56–65. doi: 10.1016/j.bandc.2016.06.009

Polyn, S. M., Norman, K. A., and Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychol. Rev.* 116, 129–156. doi: 10.1037/a0014420

Preston, A., and Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Curr. Biol.* 23, 764–773. doi: 10.1016/j.cub.2013.05.041

Raaijmakers, J. G. W., and Schiffrin, R. (1981). Search of associative memory. *Psychol. Rev.* 8, 98–134. doi: 10.1037/0033-295X.88.2.93

Ragland, J. D., Laird, A. R., Ranganath, C., Blumenfeld, R. S., and Glahn, D. C. (2009). Prefrontal activation deficits during episodic memory in schizophrenia. *Am. J. Psychiatry* 166, 863–874. doi: 10.1176/appi.ajp.2009.08091307

Reyna, V. F. (2000). Fuzzy-trace theory and source monitoring: an evaluation of theory and false-memory data. *Learn. Individual Diff.* 12, 163–175. doi: 10.1016/S1041-6080(01)00034-6

Rhodes, S. M., and Donaldson, D. I. (2008). Electrophysiological evidence for the effect of interactive imagery on episodic memory: encouraging familiarity for non-unitized stimuli during associative recognition. *NeuroImage* 39, 873–884. doi: 10.1016/j.neuroimage.2007.08.041

Riefer, D. M., Hu, X., and Batchelder, W. H. (1994). Response strategies in source monitoring. *J. Exp. Psychol. Learn. Memory Cogn.* 20, 680–693. doi: 10.1037/0278-7393.20.3.680

Ries, M. L., Mclaren, D. G., Bendlin, B. B., Xu, G., Rowley, H. A., Birn, R., et al. (2012). Medial prefrontal functional connectivity—relation to memory self-appraisal accuracy in older adults with and without memory disorders. *Neuropsychologia* 50, 603–611. doi: 10.1016/j.neuropsychologia.2011.12.014

Ro, T., Wallace, R., Hagedorn, J., Farnè, A., and Pienkos, E. (2004). Visual enhancing of tactile perception in the posterior parietal cortex. *J. Cogn. Neurosci.* 16, 24–30. doi: 10.1162/089892904322755520

Roozendaal, B., McEwen, B. S., and Chattarji, S. (2009). Stress, memory and the amygdala. *Nat. Rev. Neurosci.* 10, 423–433. doi: 10.1038/nrn2651

Sahakyan, L., and Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *J. Exp. Psychol. Learn. Memory Cogn.* 28, 1064–1072. doi: 10.1037/0278-7393.28.6.1064

Schacter, D. L., Kaszniak, A. W., Kihlstrom, J. F., and Valdiserri, M. (1991). The relation between source memory and aging. *Psychol. Aging* 6, 559–568. doi: 10.1037/0882-7974.6.4.559

Sederberg, P. B., Howard, M. W., and Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychol. Rev.* 115, 893–912. doi: 10.1037/a0013396

Slotnick, S. D. (2010). 'Remember' source memory rocs indicate recollection is a continuous process. *Memory* 18, 27–39. doi: 10.1080/09658210903390061

Slotnick, S. D., and Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory Cogn.* 33, 151–170. doi: 10.3758/BF03195305

Slotnick, S. D., Moo, L. R., Segal, J. B., and Hart, J. Jr. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cogn. Brain Res.* 17, 75–82. doi: 10.1016/S0926-6410(03)00082-X

Snytte, J., Elshiekh, A., Subramaniapillai, S., Manning, L., and Rajah, M. N. (2020). The ratio of posterior–anterior medial temporal lobe volumes predicts source memory performance in healthy young adults. *Hippocampus* 30, 1209–1227. doi: 10.1002/hipo.23251

Sofia, P., and Gregoriou, G. G. (2017). Top-down control of visual attention by the prefrontal cortex: functional specialization and long-range interactions. *Front. Neurosci.* 11:545. doi: 10.3389/fnins.2017.00545

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiol. Learn. Mem.* 82, 171–177. doi: 10.1016/j.nlm.2004.06.005

Starns, J. J., and Ksander, J. C. (2016). Item strength influences source confidence and alters source memory zROC slopes. *J. Exp. Psychol. Learn. Memory Cogn.* 42, 351–365. doi: 10.1037/xlm0000177

Starns, J. J., Pazzaglia, A. M., Rotello, C. M., Hautus, M. J., and Macmillan, N. A. (2013). Unequal-strength source zROC slopes reflect criteria placement and not (necessarily) memory processes. *J. Exp. Psychol. Learn. Memory Cogn.* 39, 1377–1392. doi: 10.1037/a0032328

Steinberger, A., Payne, J. D., and Kensinger, E. A. (2011). The effect of cognitive reappraisal on the emotional memory trade-off. *Cogn. Emot.* 25, 1237–1245. doi: 10.1080/02699931.2010.538373

Sutton, S. K., and Davidson, R. J. (1997). Resting prefrontal asymmetry: a biological substrate of the behavioral approach and behavioral inhibition system. *Psychol. Sci.* 8, 204–210. doi: 10.1111/j.1467-9280.1997.tb00413.x

Swick, D., Senkfor, A., and Petten, C. V. (2006). Source memory retrieval is affected by aging and prefrontal lesions: behavioral and ERP evidence. *Brain Res.* 1107, 161–176. doi: 10.1016/j.brainres.2006.06.013

Takehara-Nishiuchi, K. (2020). Prefrontal–hippocampal interaction during the encoding of new memories. *Brain Neurosci. Adv.* 4:2398212820925580. doi: 10.1177/2398212820925580

Talmi, D., Lohnas, L. J., and Daw, N. D. (2019). A retrieved context model of the emotional modulation of memory. *Psychol. Rev.* 126, 455–485. doi: 10.1101/175653

Talmi, D., Luk, B. T. C., McGarry, L. M., and Moscovitch, M. (2007). The contribution of relatedness and distinctiveness to emotionally-enhanced memory. *J. Memory Lang.* 56, 555–574. doi: 10.1016/j.jml.2007.01.002

Talmi, D., and Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory Cogn.* 32, 742–751. doi: 10.3758/BF03195864

Tulving, E. (2002). Episodic memory: from mind to brain. *Annu. Rev. Psychol.* 53, 1–25. doi: 10.1146/annurev.psych.53.100901.135114

Tulving, E. (2004). Episodic memory: from mind to brain. *Annu. Rev. Psychol.* 53, 1–25.

Tulving, E., Kapur, S., Craik, F., and Houle, M. S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: positron emission tomography findings. *Proc. Natl. Acad. Sci. U. S. A.* 91, 2016–2020. doi: 10.1073/pnas.91.6.2016

Tulving, E., and Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev.* 80, 352–373. doi: 10.1037/h0020071

Turi, G. F., Li, W., Chalis, S., Pandi, I., O'Hare, J., Priestley, J. B., et al. (2019). Vasoactive intestinal polypeptide-expressing interneurons in the hippocampus support goal-oriented spatial learning. *Neuron* 101, 1150–1165. doi: 10.1016/j.neuron.2019.01.009

Vargha-Khadem, F. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science* 277, 376–380. doi: 10.1126/science.277.5324.376

Venneri, A., Mitolo, M., Beltrachini, L., Varma, S., Pietà, C. D., Jahn-Carta, C., et al. (2019). Beyond episodic memory: semantic processing as independent predictor of hippocampal/perirhinal volume in aging and mild cognitive impairment due to Alzheimer's disease. *Neuropsychology* 33, 523–533. doi: 10.1037/neu0000534

von Stein, A., Rappelsberger, P., Sarnthein, J., and Petsche, H. (1999). Synchronization between temporal and parietal cortex during multimodal object processing in man. *Cerebral Cortex* 9, 137–150. doi: 10.1093/cercor/9.2.137

Vuli, K., Bjeki, J., Paunovi, D., Jovanovi, M., and Filipovi, S. R. (2021). Theta-modulated oscillatory transcranial direct current stimulation over posterior parietal cortex improves associative memory. *Sci. Rep.* 11:3013. doi: 10.1038/s41598-021-82577-7

Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition* 69, 1–34. doi: 10.1016/S0010-0277(98)00058-4

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: a formal dual-process model and an analysis of receiver operating characteristics. *J. Exp. Psychol. Learn. Memory Cogn.* 25, 1415–1434. doi: 10.1037/0278-7393.25.6.1415

Zhang, Q., Liu, X., An, W., Yang, Y., and Wang, Y. (2015). Recognition memory of neutral words can be impaired by task-irrelevant emotional encoding contexts: behavioral and electrophysiological evidence. *Front. Hum. Neurosci.* 9:73. doi: 10.3389/fnhum.2015.00073

Zheng, Z., Li, J., Xiao, F., Ren, W., and He, R. (2016). Unitization improves source memory in older adults: an event-related potential study. *Neuropsychologia* 89, 232–244. doi: 10.1016/j.neuropsychologia.2016.06.025

Author Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of NSFC.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

BRIEF RESEARCH REPORT
published: 06 October 2021
doi: 10.3389/fnins.2021.698877

The Effect of Task-Irrelevant Emotional Valence on Limited Attentional Resources During Deception: An ERPs Study

Jing Liang[1], *Qian-Nan Ruan*[2], *Ke-Ke Fu*[3], *Yu-Hsin Chen*[4]* *and Wen-Jing Yan*[5]*

[1] School of Educational Science, Ludong University, Yantai, China, [2] Wenzhou Seventh People's Hospital, Wenzhou, China, [3] School of Education and Psychology, University of Jinan, Jinan, China, [4] College of Liberal Arts, Wenzhou-Kean University, Wenzhou, China, [5] School of Mental Health, Wenzhou Medical University, Wenzhou, China

Deception is a complex and cognitively draining dyadic process that simultaneously involves cognitive and emotional processes, both of which demand/capture attentional resources. However, few studies have investigated the allocation of attentional resources between cognitive and emotional processes during deception. The current study presented facial expressions of different valences to 36 participants. While an electroencephalogram was recording, they were asked to make either truthful or deceptive gender judgments according to preceding cues. The results showed that deceptive responses induced smaller P300 amplitudes than did truthful responses. Task-irrelevant negative emotional information (TiN) elicited larger P300 amplitudes than did task-irrelevant positive emotional information (TiP). Furthermore, the results showed that TiN elicited larger LPP amplitudes than did TiP in deceptive responses, but not in truthful ones. The results suggested that attentional resources were directed away to deception-related cognitive processes and TiN, but not TiP, was consistently able to compete for and obtain attentional resources during deception. The results indicated that TiN could disrupt with deception and may facilitate deception detection.

Keywords: deception, emotional valence, task relevance, attentional resources, ERP

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Casey S. Gilmore,
Minneapolis VA Health Care System,
United States
Antigona Martinez,
Nathan Kline Institute for Psychiatric
Research, United States

****Correspondence:***
Wen-Jing Yan
yanwj@wzu.edu.cn
Yu-Hsin Chen
yhandrewc@gmail.com

Specialty section:
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

Received: *22 April 2021*
Accepted: *30 August 2021*
Published: *06 October 2021*

Citation:
Liang J, Ruan Q-N, Fu K-K, Chen Y-H and Yan W-J (2021) The Effect of Task-Irrelevant Emotional Valence on Limited Attentional Resources During Deception: An ERPs Study.
Front. Neurosci. 15:698877.
doi: 10.3389/fnins.2021.698877

INTRODUCTION

Deception is typically defined as a psychological process by which an individual deliberately attempts to convince another to accept as true what the first individual knows to be false, typically to gain benefits or avoid losses for the liar, but sometimes for others (Abe, 2009). Previous studies have shown that deception is a highly complex and cognitively draining dyadic process that simultaneously involves both cognitive and emotional mechanisms (Buller and Burgoon, 1996; Burgoon and Buller, 2015). To successfully deceive another individual, deception simultaneously calls upon numerous cognitive processes to execute, such as response-conflict monitoring, inhibitory control, and/or task switching (Debey et al., 2015; Suchotzki et al., 2015). Furthermore, the deceiver has to constantly monitor his or her target's demeanor and emotions and control their own throughout the deception process, in order to appear credible and convincing (Vrij et al., 2019). As such, recent deception studies have begun to investigate possible interactions among the cognitive and emotional processes present during deception (Lee et al., 2010; Frank and Svetieva, 2012;

Liang et al., 2018). For example, Liang et al. (2018) asked participants to make truthful or deceptive gender judgments when viewing positively and negatively valenced facial expressions. This approach enabled the researchers to investigate the cognitive processes underlying individuals' truthful and deceptive responses when task-irrelevant emotional information was present. Their results indicated an influence of emotional intensity upon the difference in truthful and deceptive response times, suggesting that the automatic attention-orienting mechanisms of task-irrelevant emotional information were influenced on the cognitive cost of deception. However, it remains unclear how the mind resolves the allocation of finite attentional resources during deception when cognitive and emotional processes are present simultaneously.

A prominent event-related potentials (ERP) in deception tasks is the parietal P3 (also referred to as early P3/late P3/P3b/LPC/LPP according to the time window used in each study) which is a positive deflection typically occurring between 300 and 1,000 ms post-stimulus (Li et al., 2008). The parietal P3 amplitude is selectively sensitive to resources of a perceptual/cognitive nature. Recognition of salient/known stimuli induces larger P3 amplitude than unknown stimuli especially when the knowledge of the stimuli is concealed, as reported in Concealed Information Test studies (Meijer et al., 2014; Leue and Beauducel, 2015). Whereas other deception studies speculate that smaller P3 amplitude reflects attentional resources directed away from the primary task to cognitive processes related to deception (Johnson et al., 2003, 2005; Wu et al., 2009). Specifically, deception researchers generally agree that deception is a cognitively complex process that simultaneously enlists additional mechanisms (i.e., maintenance of truthful information in working memory, inhibition of truthful information, and execution of deceptive responses) beyond the primary task/act of lying itself. As a result, it has been reported through both behavioral (Debey et al., 2012; Suchotzki et al., 2017) and neuroimaging researches (Abe, 2009; Christ et al., 2009) that attentional resources are strained and simultaneously directed across multiple processes during deception. Recent findings on parietal P3 amplitude in tasks requiring deception (Leue et al., 2012; Leue and Beauducel, 2015, 2019) reported individual differences in a deception task for both early P3 (between 300 and 400 ms or 280 and 350 ms post-stimulus) and late P3 (between 400 and 700 ms or 440 and 610 ms post-stimulus). Leue et al. (2012) observed an effect of personality for early P3 amplitude but not for the late P3 amplitude, indicating that early P3 and late P3 amplitudes represent different processes (Leue and Beauducel, 2015).

According to findings of Kayser et al. (2000) who observed larger early and late P3 amplitudes with parietal topographies following emotionally salient stimuli compared to neutral stimuli. It was suggested that early parietal P3 amplitude serve as an indicator of initial affective stimulus salience and should probably not be regarded as a P3a or novelty P3, which is known to have a frontal topography, whereas late parietal P3 amplitude serves as a somatic marker to signal stimulus significance and to guide behavior (Damasio et al., 1991; Kayser et al., 2000; Leue and Beauducel, 2015). Many research found increased LPP amplitude (usually refers to the late P3 amplitude between 400 and 800 ms post-stimulus) following the presentation of emotional rather than neutral stimuli (Hajcak and Olvet, 2008; Hajcak et al., 2010). Furthermore, the LPP amplitudes elicited by negatively valenced pictures were observed to be significantly greater than those elicited by positively valenced pictures, even though both were equally probable, evaluatively extreme, and arousing (Vaish et al., 2008). One study also reported how LPP amplitudes were larger for negatively rather than positively valenced pictures, even when subjects were not asked to explicitly evaluate the valence of the stimulus (Ito and Cacioppo, 2000), demonstrating a negativity bias. In addition, Ferrari et al. (2008) investigated whether LPP amplitudes were simultaneously influenced by task relevance and emotionality, observing the largest LPP amplitudes from task-relevant emotional pictures and the smallest from task-irrelevant neutral pictures. When taken together, these studies suggest that the LPP amplitudes can reflect attentional resources diverted away from the processing of task-relevant stimulus information when both cognitive and emotional processes are simultaneously present in a task that involves a multifaceted stimulus.

However, researchers have yet to explore the division and allocation of finite attentional resources during deception when task-relevant cognitive processes and task-irrelevant emotional information are simultaneously present in a multifaceted stimulus. The present study aims to investigate the above question. Participants were asked to make truthful and deceptive gender judgments about positive and negative facial expressions while behavioral and electroencephalogramic data were recorded. Accordingly, the experiment stimuli included task-relevant information (i.e., gender relevance data such as hair, facial physique, facial features), task-irrelevant positively valenced emotional information (happy facial expression images; from here on referred to as TiP), and task-irrelevant negatively valenced emotional information (angry facial expression images; from here on referred to as TiN).

According to previous studies (Johnson et al., 2003, 2005), we expected significantly reduced parietal P3 amplitudes (for both early and late P3) for deceptive rather than truthful responses indicating that attentional and/or processing resources were being diverted away from processing the stimulus information related to the gender judgment task in order to attend to deception related processes such as maintaining truthful information in working memory, inhibition of truthful information, and execution of deceptive responses which cost more mental effort (Leue and Beauducel, 2019, p. 2). According to negativity bias (Ito and Cacioppo, 2000; Vaish et al., 2008) which indicated negatively valenced emotional information tends to be more salient, we postulated larger parietal P3 amplitudes (both early and late P3) for TiN rather than TiP. Since late P3 amplitude was simultaneously influenced by task relevance and emotionality (Ferrari et al., 2008), we expected the interaction between response (truthful vs. deceptive) and emotion to emerge on late P3 which is required to guide the individual in response selection and decision making by signaling stimulus significance.

METHOD

Participants

Thirty-six undergraduate and graduate students (18 males, average age = 21 years) were recruited. Participants were given a monetary reward for their participation. All were right-handed and had normal or corrected-to-normal vision. All signed informed consents and were approved by the Ethical Committee of Ludong University, in accordance with the Declaration of Helsinki.

Stimuli and Procedure

Based on the stimulus evaluation results in the study by Liang et al. (2018), 18 happy and 18 angry faces (half female) with intensities above 4.5 from the NimStim and MUG databases were selected for the formal experiment. The average intensity scores were 5.58 (*SD* = 0.57) for the happy faces and 5.76 (*SD* = 0.63) for the angry faces on a 9-point Likert Scale. Independent *t*-tests on the average intensity scores revealed no significant differences between happy and angry faces, $t(34) = 0.93$, $p = 0.36$.

For each trial, following a fixation cross (+) that was visible for 500–750 ms, the cue "T" or "L" appeared for 300 ms. Next, a blank gray screen was presented for 1,000 ms. Then, an image of a facial expression was presented for 300 ms, followed by a blank gray screen for 1,200 ms. The participants judged the gender of the model presented as accurately and quickly as possible, according to the preceding cue. The "F" key was labeled "Male" and the "J" key was labeled "Female." Participants were instructed to make truthful responses if the cue presented was "T" and deceptive responses if the cue presented was "L." The "Male" and "Female" keys were counterbalanced among the participants. There were equal numbers of "T" and "L" cues for each block, and all were presented in random order (see **Figure 1**).

The experiment consisted of three blocks. Each included 72 trials, with each picture repeated twice within a single block. Participants were given as much time as needed to rest before proceeding to the next block.

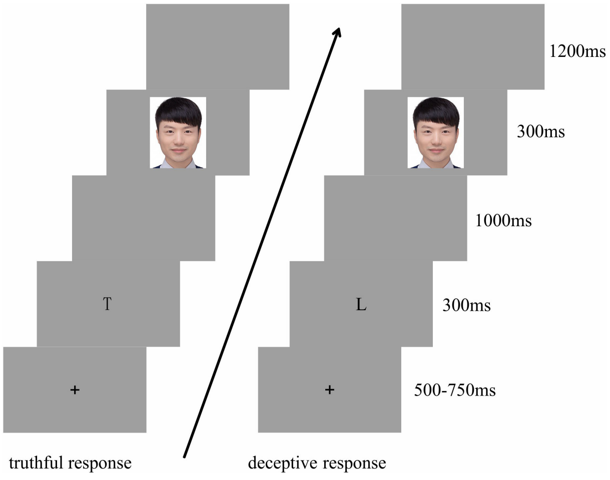


FIGURE 1 | Experiment procedure.

Electrophysiological Recording and Analysis

Continuous EEG readings were recorded using 64 electrodes mounted in an elastic cap (Electro-Cap international, Inc.) that was connected to the left mastoid. The data were removed offline and then re-referenced to the average of the left and right mastoids (M1 and M2). The vertical and horizontal electrooculograms were recorded with bipolar channels from sites above and below the midpoint of the left eye and next to the outer canthi of each eye. Mild skin abrasion was performed to reduce electrode impedance below 5 kΩ. The EEG was band-pass filtered from 0.05 to 100 Hz, amplified with a gain of 500, and stored on a computer hard drive at a sample rate of 1,000 Hz (Syn-Amps 4.5, Neuroscan, Inc.).

EOG artifacts from eye blinking and horizontal movement were automatically corrected in all trials using the Scan 4.5 software package. EEGs contaminated with artifacts due to amplifier clipping, bursts of electromyographic activity, or peak-to-peak deflection exceeding ± 75 μV were excluded from the trials. In the present study, the ERP waveforms were time-locked to the time of appearance of the facial expression images. The averaged epoch for the ERPs was 800 ms, including a 100 ms pre-stimulus baseline. Based on previous studies (Leue et al., 2012; Leue and Beauducel, 2015) and inspection of the topographical distribution of the grand-averaged ERP activities, there are two positive peaks in the time interval relevant for the P3 (300–800 ms), one between 300 and 400 ms and another one between 450 and 750 ms after stimulus onset. This indicates that an average across the relevant time window could represent two different components so that both time intervals were used for EEG component quantification. According to Luck (2005), we performed a combined amplitude and latency measurement: In each individual ERP we searched for the most positive segment average of 50 ms within the time window of 300–400 ms after stimulus onset for the early P3 due to its narrow window. The mean amplitude of the late P3 was measured between 450 and 750 ms. The following 15 electrode sites [Fz, F3, F4 (three frontal sites), FCz, FC3, FC4 (three frontal-central sites), Cz, C3, C4 (three central sites), CPz, CP3, CP4 (three central-parietal sites), and Pz, P3, P4 (three parietal sites)] were chosen for the statistical analysis. The amplitude of each ERP component was then submitted in an ANOVA with response type (truth vs. deception), valence (positive vs. negative), and electrode zones (frontal vs. frontal-central vs. central vs. central-parietal vs. parietal) as within-subject variables. ERP amplitudes were averaged over electrodes within a zone.

RESULTS

Behavioral Data

A series of 2 (valence: positive vs. negative) × 2 (response type: truthful vs. deceptive) repeated ANOVAs were conducted on both accuracy and reaction time. The main effect of response type was found for accuracy, [$F(1, 35) = 15.33$, $p < 0.001$, $\eta^2_p = 0.31$], indicating a higher accuracy for truthful ($M = 0.93$, $SE = 0.01$) rather than deceptive responses ($M = 0.90$, $SE = 0.01$). The main effect of response type was found for reaction time, [$F(1, 35) = 118.10$, $p < 0.001$, $\eta^2_p = 0.77$], indicating a shorter reaction time for truthful ($M = 678.12$, $SE = 25.90$) rather than deceptive responses ($M = 740.77$, $SE = 25.74$). The main effect of valence was found for reaction time, [$F(1, 35) = 4.42$, $p = 0.04$, $\eta^2_p = 0.11$], indicating a shorter reaction time for TiP ($M = 705.33$, $SE = 25.99$) rather than TiN ($M = 713.56$, $SE = 25.47$). The other main and interaction effects were not significant (**Figure 2**).

Event-Related Potentials Data

Figure 3 shows the overall ERP waveforms for the 36 subjects in the four conditions, which were detected using three electrodes on the midline of the scalp. After artifact rejection, the mean numbers of valid trials were 41, 41, 42, and 41 for deceptive responses to negative facial expressions, deceptive responses to positive facial expressions, truthful responses to negative facial expressions, and truthful responses to positive facial expressions, respectively.

Early P3 (300–400 ms)

A significant main effect of response type was revealed for early P3, [$F(1, 35) = 9.54$, $p = 0.004$, $\eta^2_p = 0.21$]; the mean amplitude for truthful responses ($M = 2.94$, $SE = 0.66$) was greater than for deceptive responses ($M = 2.19$, $SE = 0.59$). The main effect of valence was significant, [$F(1, 35) = 4.37$, $p = 0.04$, $\eta^2_p = 0.11$]; TiN ($M = 2.74$, $SE = 0.64$) elicited a larger early P3 than did TiP ($M = 2.40$, $SE = 0.59$). The main effect of electrode zone was significant, [$F(4, 140) = 84.48$, $p < 0.001$, $\eta^2_p = 0.71$], indicating that the amplitude grew greater when moving from the frontal area to the parietal area (0.23, 0.74, 1.99, 4.04, 5.85). The differences among areas are significant except the difference between frontal and frontal-central area. The interactive effect of valence and electrode zone was significant, [$F(4, 140) = 4.43$, $p = 0.03$, $\eta^2_p = 0.11$], indicating that the difference between TiN and TiP is significant in the central-parietal and parietal areas ($p = 0.003$, $p = 0.002$), marginally significant in the central area ($p = 0.06$), but not in the frontal and frontal-central areas. The other interactions were not significant. For scalp topographical maps (see **Figure 4**).

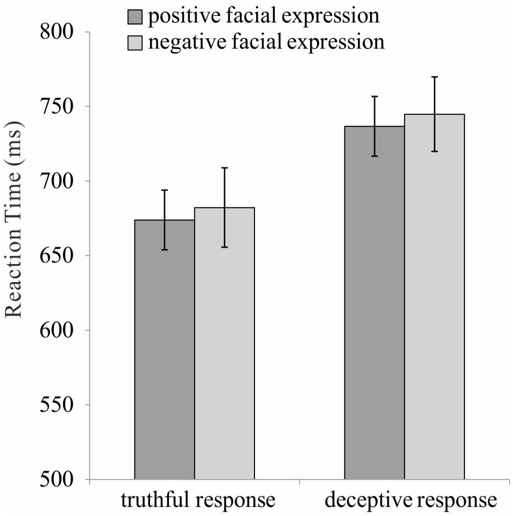


FIGURE 2 | The reaction time for positive and negative facial expression under truthful and deceptive condition.

Late P3 (450–750 ms)

For the late P3 mean amplitude, there was a significant effect of response type, [$F(1, 35) = 9.09$, $p = 0.005$, $\eta^2_p = 0.21$], with the amplitude for truthful responses ($M = 4.08$, $SE = 0.68$) being greater than for deceptive responses ($M = 3.53$, $SE = 0.66$). The main effect of valence was marginally significant, [$F(1, 35) = 3.49$, $p = 0.07$, $\eta^2_p = 009$], meaning that TiN ($M = 3.98$, $SE = 0.68$) was associated with a more positive late P3 than was TiP ($M = 3.63$, $SE = 0.66$). The main effect of electrode zone was significant, [$F(4, 140) = 15.85$, $p < 0.001$, $\eta^2_p = 0.31$], indicating that the amplitudes in the central, central-parietal and parietal areas (4.51, 5.02, 4.31) were more positive than that in the frontal and frontal-central areas (1.91, 3.26). The interactive effect of response type and valence was significant, [$F(1, 35) = 4.14$, $p = 0.05$, $\eta^2_p = 0.11$]. The amplitude for TiN ($M = 3.84$, $SE = 0.70$) was greater than for TiP ($M = 3.21$, $SE = 0.63$) under deceptive conditions, $p = 0.01$. In contrast, the late P3 amplitude for TiN and TiP did not differ under truthful conditions, $p = 0.80$ (see **Figure 5**). For scalp topographical maps (see **Figure 6**). The interactive effect of response type and electrode zone was significant, [$F(4, 140) = 7.15$, $p < 0.001$, $\eta^2_p = 0.17$], indicating that the amplitude difference between truthful responses and deceptive responses was significant in the frontal, frontal-central, central and central-parietal areas, but not in the parietal areas. The other interactions were not significant.

DISCUSSION

In the present study, we asked participants to make truthful and deceptive gender judgments of positive and negative facial expressions. An ERP was applied to investigate how individuals resolved and allocated finite attentional resources when a stimulus contained both task-relevant (i.e., gender relevance information) and task-irrelevant emotional information (positively/negatively valenced) during deceptive gender judgments. The results show that deceptive responses induced smaller parietal P3 amplitudes (both early and late P3) than did truthful responses. TiN elicited larger parietal P3 amplitudes (both early and late P3) as compared to TiP. Furthermore, the results reveal that TiN elicited larger late parietal P3 amplitude than did TiP in deceptive but not in truthful responses.

In the present study, significantly reduced early and late parietal P3 amplitudes were observed during deceptive responses in comparison to truthful responses. This result is consistent with previous deception researches suggesting that smaller P3 amplitude (both early and late) reflects attention directed away

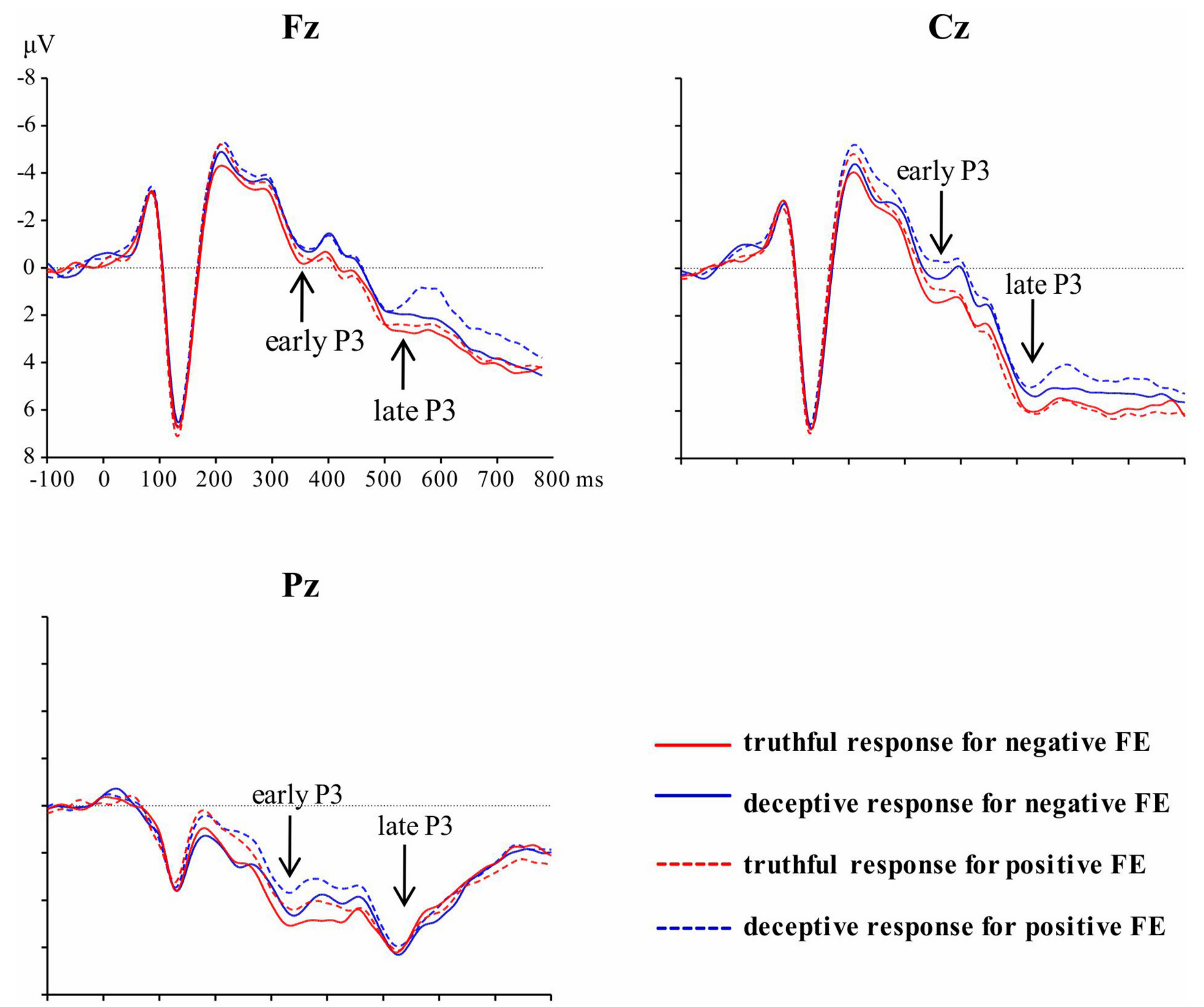


FIGURE 3 | Grand average waveforms at Fz, Cz, and Pz for each of the four experiment conditions.

to other processes present during deception (Johnson et al., 2003, 2005; Wu et al., 2009). In line with recent meta-analysis (Leue and Beauducel, 2019), reduced parietal P3 amplitude in deception can be interpreted in terms of mental effort. Taken together, the results in the present study indicate more mental effort was required when the individuals were cued to make deceptive rather than truthful responses. Significantly larger early and late parietal P3 amplitudes for the TiN trials in comparison to the TiP trials further revealed the influence of task-irrelevant emotional information. Numerous studies have observed differences in parietal P3 amplitude among emotional stimuli of different valences. Specifically, larger parietal P3 amplitude has commonly been reported in negatively valenced stimuli, as compared to positively valenced stimuli (Ito and Cacioppo, 2000; Ferrari et al., 2008; Vaish et al., 2008). In general, previous findings suggest that negatively valenced stimuli have higher saliency than other affectively neutral or even positive stimuli (Ogawa and Suzuki, 2004). Our ERP results mirror those of previous ERP studies, with the results of early and late P3 amplitudes indicating that TiN was more salient compared to TiP.

On top of the main effect for both response type and valence, is the observed interaction between response type and valence for late P3 amplitude. In the present study, larger late parietal P3 amplitude was observed for TiN than for TiP when participants made deceptive responses, whereas no differences in late parietal P3 amplitude were found between TiN and TiP when participants made truthful responses. Previous findings have suggested that late parietal P3/LPP activity indicates the amount of attentional resources being allocated to the processing of a stimulus (Ferrari et al., 2008). In accordance with previous findings, the former result suggests that the amount of attentional resources being allocated to process TiN was greater than for TiP when individuals made deceptive responses. This finding is consistent with the majority of previous findings, which reported that

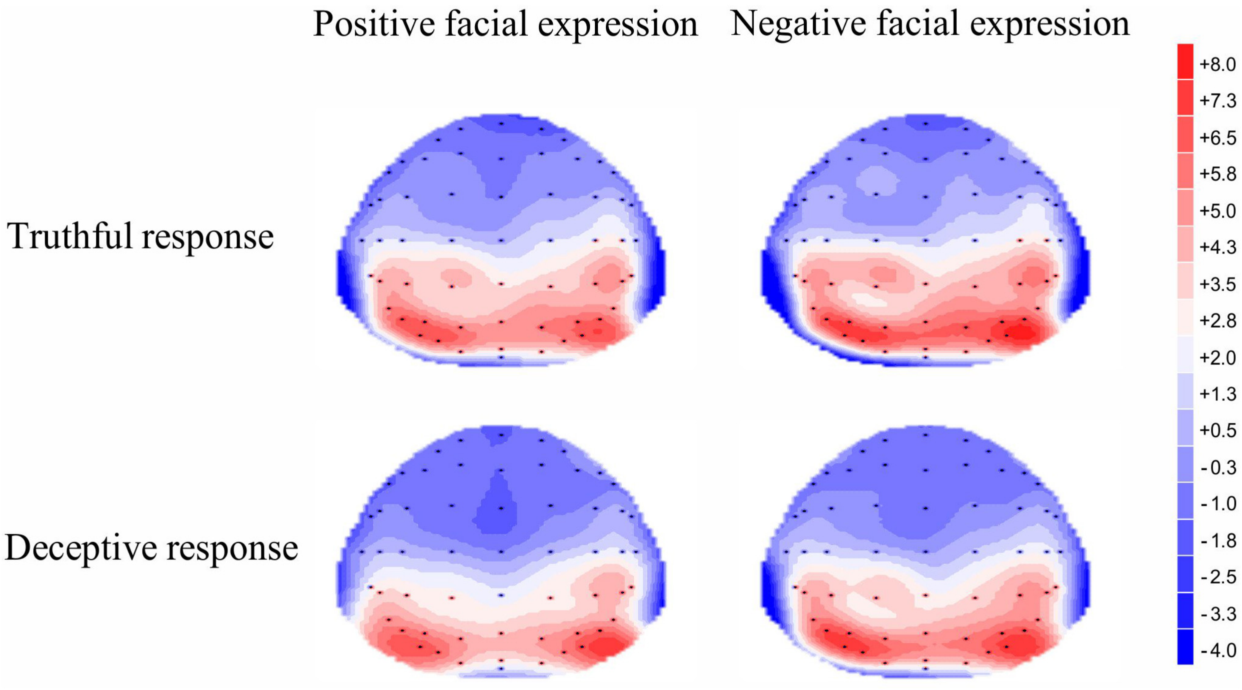


FIGURE 4 | Topographic maps of the mean voltage amplitudes for the four experimental conditions in the 300–400 ms time window.

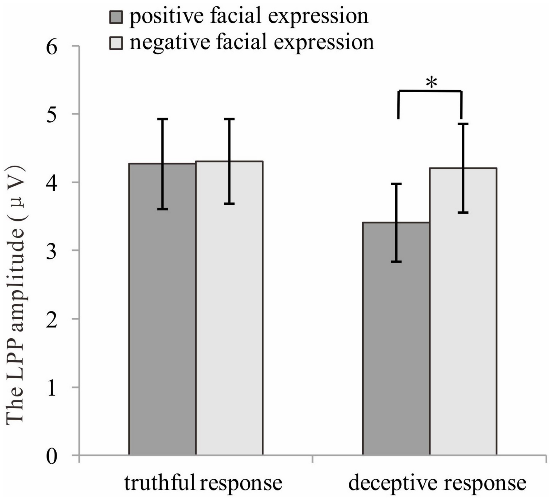


FIGURE 5 | The late P3 amplitudes for truthful and deceptive responses to positive and negative facial expressions. *$p < 0.05$.

task-irrelevant negative stimuli were more salient and thus demanded more attention (due to processing bias) than task-irrelevant positive stimuli. Yet the latter result suggests otherwise, indicating that the amount of attentional resources allocated to process TiN did not differ from that of TiP when the individuals made truthful responses. Our ERP results imply that when attentional resources are strained (for example, in deceptive responses and when attention needs to be spread simultaneously across numerous cognitive processes), the amount of attentional resources or effort allocated to process TiN is greater than when TiP is present. This indicates that negatively valenced emotional information, despite being task-irrelevant, competed for attentional resources and successfully acquired more attentional resources when those attentional resources were strained during deceptive responses; less attentional resources were allocated to TiP during deceptive responses. In contrast, when attentional resources were not as strained (for example, in truthful response conditions), both TiN and TiP competed for attentional resources and perhaps received equal allocations. Based on the results of the present study, it can be speculated that TiN was able to consistently obtain attentional resources. In contrast, the attentional resources or effort allocated to TiP depended on whether attentional resources were constrained or not.

These results provide a glimpse into how task-irrelevant emotional information influences individuals' deceptive and truthful responses. Deception detection researches have unveiled numerous findings underscoring the importance of both emotion and cognitive load in deception (Vrij et al., 2012, 2017). The findings of the present study demonstrate how the two equally important processes involved in deception simultaneously compete for and are allocated finite attentional resources. These conclusions have implications for future deception detection research relying on imposed cognitive load to magnify differences between liars and truth-tellers. Numerous studies have reported how additional cognitive loads can effectively aid in deception detection (Frank and Svetieva, 2012; Vrij et al., 2012, 2017). According to that theory and previous findings, additional

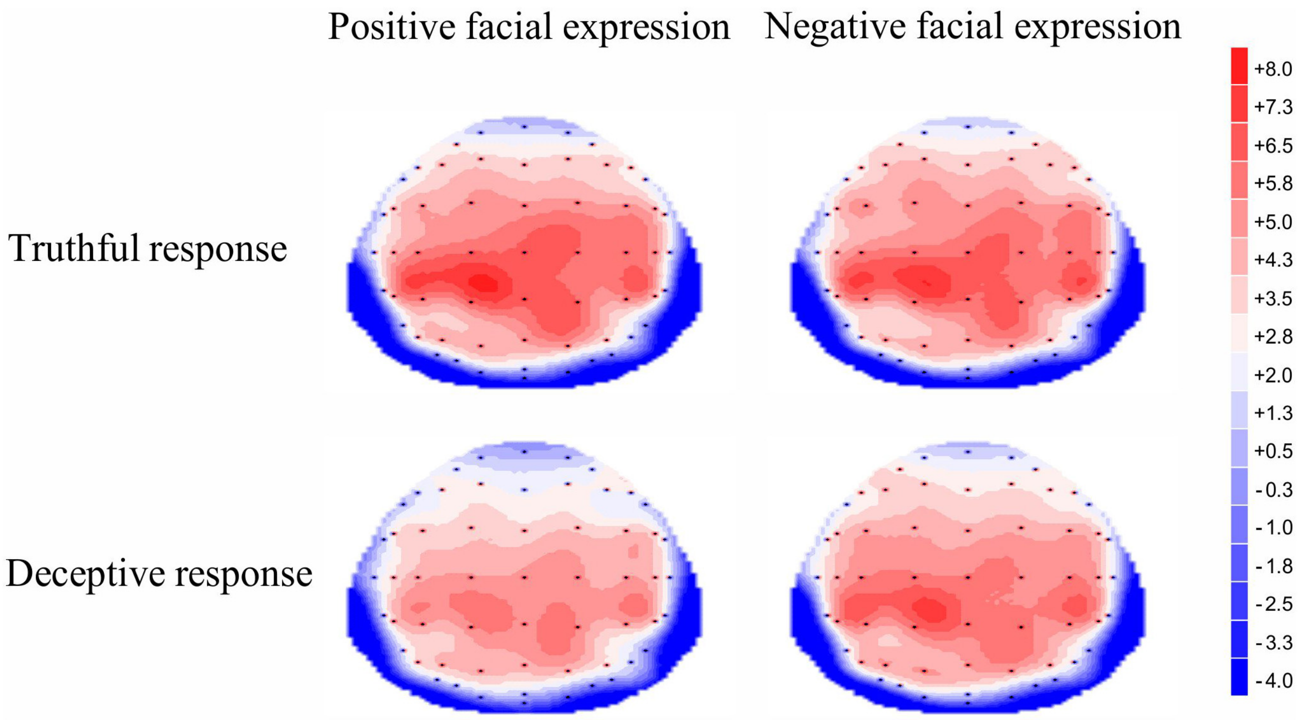


FIGURE 6 | Topographic maps of the mean voltage amplitudes for the four experimental conditions in the 450–750 ms time window.

cognitive load imposed on liars whose cognitive resources have already been partially depleted by the cognitively demanding task of lying, further strain their available resources. As a result, such loading has a particularly debilitative effect on a liar's attempt to present a plausible, detailed, and convincing depiction of an event.

Based on the findings of the present study, the presence of TiP and TiN should be considered in the process of deception detection. Our results indicate that TiN consistently competed for and exerted strain on attentional resources, regardless of the response type (deceptive vs. truthful). In contrast, the strain TiP imposed on attentional resources in a manner that did not facilitate the goal of overloading and debilitating liars by imposing additional strain. Our results suggest that less strain on attentional resources was imposed when individuals made deceptive responses if TiP was present. Conversely, the strain on attentional resources imposed when individuals made truthful responses did not differ between trials where TiN or TiP was present. These results suggest that the presence of TiP in deception detection may lessen rather than magnify the differences between truth-tellers and liars. However, caution is advised when interpreting results obtained in the present study. Deceptive responses were dictated by the experiment's instructions, rather than freely undertaken by the participants in our study. Hence, the experiment conditions differed from the real-world high stakes lies considered in previous deception detection research, where the motivation to deceive was strong. A study examining the effect of willingness toward honest and deceptive responses, reported a significant main effect of willingness, observing larger P3 amplitude elicited by self-determined responses compared to forced responses (Wu et al., 2009). Another limitation of this research is the lack of control over possible individual differences in cognitive capacity. Study has reported a relationship between individual differences in executive functions (cognitive capacity) and the accuracy as well as latency of deceptive responses. Results showed that set-shifting and inhibition were directly related to deception accuracy and speed, respectively. However, enhanced underlying working memory skills (both verbal and spatial) were negatively associated with deception speed (Visu-Petra et al., 2012).

CONCLUSION

The present study highlighted how emotional and cognitive processes involved in deception simultaneously compete for finite attentional resources. The results show that deceptive responses induced smaller early and late P3 amplitudes than did truthful responses. Additionally, TiN elicited larger early and late P3amplitudes than did TiP. Furthermore, the results reveal that TiN elicited larger late P3 amplitude than did TiP in deceptive but not truthful responses. The results suggest that attentional resources were directed away to deception-related cognitive processes. Moreover, TiN but not TiP, was consistently able to compete for and obtain attentional resources during deception. The results reveal a unique effect of stimulus saliency (in the form of task irrelevant negatively valenced emotional information) during deceptive responses but not truthful responses.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the IRB in Ludong University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JL conceived and designed the experiments. W-JY and JL performed the experiments. JL, Y-HC, Q-NR, and K-KF wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the National Natural Science Foundation of China (No. 31900761), Shandong Provincial Natural Science Foundation (ZR2018PC034), Zhejiang Province Higher Education Teaching Reform Funding (JG20190471), and Shandong Province Higher Educational Science and Technology Program (J17RB179). This research was also supported by the Joint Training Base for Postgraduates of Applied Psychology by Ludong University and Qingdao Ruiyang Xinyu Applied Psychology Technology.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.698877/full#supplementary-material

REFERENCES

Abe, N. (2009). The neurobiology of deception: evidence from neuroimaging and loss-of-function studies. *Curr. Opin. Neurol.* 22, 594–600. doi: 10.1097/WCO.0b013e328332c3cf

Buller, D. B., and Burgoon, J. K. (1996). "Emotional expression in the deception process," in *Handbook of Communication and Emotion*, eds P. A. Andersen and L. K. Guerrero (San Diego, CA: Academic Press), 381–402. doi: 10.1016/B978-012057770-5/50016-3

Burgoon, J. K., and Buller, D. B. (2015). "Interpersonal deception theory," in *The International Encyclopedia of Interpersonal Communication*, eds S. W. Littlejohn, and K. A. Foss (Hoboken, NJ: John Wiley & Sons), 1–6. doi: 10.1002/9781118540190.wbeic170

Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., and McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. *Cereb. Cortex* 19, 1557–1566. doi: 10.1093/cercor/bhn189

Damasio, A. R., Tranel, D., and Damasio, H. (1991). "Somatic markers and the guidance of behavior: theory and preliminary testing," in *Frontal Lobe Function and Dysfunction*, eds H. B. Levin, H. M. Eisenberg, and A. L. Benton (Oxford: Oxford University Press), 217–229.

Debey, E., Ridderinkhof, R. K., De, H. J., De, S. M., and Verschuere, B. (2015). Suppressing the truth as a mechanism of deception: delta plots reveal the role of response inhibition in lying. *Conscious. Cogn.* 37, 148–159. doi: 10.1016/j.concog.2015.09.005

Debey, E., Verschuere, B., and Crombez, G. (2012). Lying and executive control: an experimental investigation using ego depletion and goal neglect. *Acta Psychol.* 140, 133–141. doi: 10.1016/j.actpsy.2012.03.004

Ferrari, V., Codispoti, M., Cardinale, R., and Bradley, M. M. (2008). Directed and motivated attention during processing of natural scenes. *J. Cogn. Neurosci.* 20, 1753–1761. doi: 10.1162/jocn.2008.20121

Frank, M. G., and Svetieva, E. (2012). Lies worth catching involve both emotion and cognition. *J. Appl. Res. Mem. Cogn.* 1, 131–133. doi: 10.1016/j.jarmac.2012.04.006

Hajcak, G., and Olvet, D. M. (2008). The persistence of attention to emotion: brain potentials during and after picture presentation. *Emotion* 8, 250–255. doi: 10.1037/1528-3542.8.2.250

Hajcak, G., MacNamara, A., and Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: an integrative review. *Dev. Neuropsychol.* 35, 129–155. doi: 10.1080/87565640903526504

Ito, T. A., and Cacioppo, J. T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *J. Exp. Soc. Psychol.* 36, 660–676. doi: 10.1006/jesp.2000.1430

Johnson, R., Barnhardt, J., and Zhu, J. (2003). The deceptive response: effects of response conflict and strategic monitoring on the late positive component and episodic memory-related brain activity. *Biol. Psychol.* 64, 217–253. doi: 10.1016/j.biopsycho.2003.07.006

Johnson, R., Barnhardt, J., and Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: an event-related brain potential study. *Cogn. Brain Res.* 24, 386–404. doi: 10.1016/j.cogbrainres.2005.02.011

Kayser, J., Bruder, G. E., Tenke, C. E., Stewart, J. W., and Quitkin, F. M. (2000). Event-related potentials ERPs/to hemifield presentations of emotional stimuli: differences between depressed patients and healthy adults in P3 amplitude and asymmetry. *Int. J. Psychophysiol.* 36, 211–236. doi: 10.1016/S0167-8760(00)00078-7

Lee, T. M., Lee, T. M., Raine, A., and Chan, C. C. (2010). Lying about the valence of affective pictures: an fMRI study. *PLoS One* 5:e12291. doi: 10.1371/journal.pone.0012291

Leue, A., and Beauducel, A. (2015). Effects of injustice sensitivity and sex on the P3 amplitude during deception. *Biol. Psychol.* 109, 29–36. doi: 10.1016/j.biopsycho.2015.04.004

Leue, A., and Beauducel, A. (2019). A meta-analysis of the p3 amplitude in tasks requiring deception in legal and social contexts. *Brain Cogn.* 135:103564. doi: 10.1016/j.bandc.2019.05.002

Leue, A., Lange, S., and Beauducel, A. (2012). Have you ever seen this face?—Individual differences of deception and event-related potentials. *Front. Psychol.* 3:570. doi: 10.3389/fpsyg.2012.00570

Li, H., Yuan, J., and Lin, C. (2008). The neural mechanism underlying the female advantage in identifying negative emotions: an event-related potential study. *NeuroImage* 40, 1921–1929. doi: 10.1016/j.neuroimage.2008.01.033

Liang, J., Chen, Y. H., Yan, W. J., Qu, F., and Fu, X. (2018). Effects of task-irrelevant emotional information on deception. *Cogn. Emot.* 32, 1265–1274. doi: 10.1080/02699931.2017.1404967

Luck, S. J. (2005). *An Introduction to Event-Related Potential Technique*. Cambridge, MA: The MIT Press.

Meijer, E. H., Klein Selle, N., Elber, L., and Ben-Shakhar, G. (2014). Memory detection with the concealed information test: a meta-analysis of skin conductance, respiration, heart rate, and P300 data. *Psychophysiology* 51, 879–904. doi: 10.1111/psyp.12239

Ogawa, T., and Suzuki, N. (2004). On the saliency of negative stimuli: evidence from attentional blink. *Jpn Psychol. Res.* 46, 20–30. doi: 10.1111/j.1468-5884.2004.00233.x

Suchotzki, K., Crombez, G., Smulders, F. T., Meijer, E., and Verschuere, B. (2015). The cognitive mechanisms underlying deception: an event-related potential study. *Int. J. Psychophysiol.* 95, 395–405. doi: 10.1016/j.ijpsycho.2015.01.010

Suchotzki, K., Verschuere, B., Van Bockstaele, B., Ben-Shakhar, G., and Crombez, G. (2017). Lying takes time: a meta-analysis on reaction time measures of deception. *Psychol. Bull.* 143, 428–453. doi: 10.1037/bul0000087

Vaish, A., Grossmann, T., and Woodward, A. (2008). Not all emotions are created equal: the negativity bias in social-emotional development. *Psychol. Bull.* 134, 383–403. doi: 10.1037/0033-2909.134.3.383

Visu-Petra, G., Miclea, M., and Visu-Petra, L. (2012). Reaction time-based detection of concealed information in relation to individual differences in executive functioning. *Appl. Cogn. Psych.* 26, 342–351. doi: 10.1002/acp.1827

Vrij, A., Fisher, R., and Blank, H. (2017). A cognitive approach to lie detection: a meta-analysis. *Legal Criminol. Psychol.* 22, 1–21. doi: 10.1111/lcrp.12088

Vrij, A., Hartwig, M., and Granhag, P. A. (2019). Reading lies: nonverbal communication and deception. *Annu. Rev. Psychol.* 70, 295–317. doi: 10.1146/annurev-psych-010418-103135

Vrij, A., Leal, S., Mann, S., and Fisher, R. (2012). Imposing cognitive load to elicit cues to deceit: inducing the reverse order technique naturally. *Psychol. Crime Law* 18, 579–594. doi: 10.1080/1068316X.2010.515987

Wu, H., Hu, X., and Fu, G. (2009). Does willingness affect the N2-P3 effect of deceptive and honest responses? *Neurosci. Lett.* 467, 63–66. doi: 10.1016/j.neulet.2009.10.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Psychology

ORIGINAL RESEARCH
published: 15 October 2021
doi: 10.3389/fpsyg.2021.744400

Reward Expectation Differentially Modulates Global and Local Spatial Working Memory Accuracy

*Qingjie Zhou, Zanzan Jiang and Jinhong Ding**

Beijing Key Laboratory of Learning and Cognition, College of Psychology, Capital Normal University, Beijing, China

Although it has been suggested that reward expectation affects the performance of spatial working memory tasks, controversial results have been found in previous experiments. Hence, it is still unclear to what extent reward expectation has an effect on working memory. To clarify this question, a memory-guided saccade task was applied, in which participants were instructed to retain and reconstruct a temporospatial sequence of four locations by moving their eyes in each trial. The global- and local-level spatial working memory accuracies were calculated to determine the reward effect on the global and local level of processing in spatial working memory tasks. Although high reward expectation enhanced the encoding of spatial information, the percentage of trials in which the cued location was correctly fixated decreased with increment of reward expectation. The reconstruction of the global temporospatial sequence was enhanced by reward expectation, whereas the local reconstruction performance was not affected by reward. Furthermore, the improvements in local representations of uncued locations and local sequences were at the cost of the representation of cued locations. The results suggest that the reward effect on spatial working memory is modulated by the level of processing, which supports the flexible resource theory during maintenance.

Keywords: reward expectation, spatial working memory, global-level processing, local-level processing, high cognitive load

Edited by:
Hong Xu,
Nanyang Technological University, Singapore

Reviewed by:
Supriya Ray,
Allahabad University, India
Kerstin Fröber,
University of Regensburg, Germany

Correspondence:
Jinhong Ding
dingjh@cnu.edu.cn

Specialty section:
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

Received: *20 July 2021*
Accepted: *15 September 2021*
Published: *15 October 2021*

Citation:
Zhou Q, Jiang Z and Ding J (2021) Reward Expectation Differentially Modulates Global and Local Spatial Working Memory Accuracy. Front. Psychol. 12:744400. doi: 10.3389/fpsyg.2021.744400

INTRODUCTION

Working memory plays an essential role in human adaptive behavior and lies at the core of cognitive psychology research since its birth. The limited capacity of working memory has been an important concern since the insightful research of Miller on “the magical number seven” (Miller, 1956), and it has been demonstrated that the capacity of working memory is even less than seven (Cowan, 2001; Kawasaki and Yamaguchi, 2013; Marchette et al., 2015). However, the total capacity of working memory is not fixed; it varies across individuals (Just and Carpenter, 1992; Barrett et al., 2004) and can even be improved by motivation within the same individual (Krawczyk et al., 2007; Heitz et al., 2008; Kawasaki and Yamaguchi, 2013).

As an extrinsic motivation, monetary reward is a powerful modulator of attention. The expectation of a monetary reward narrows down the scope of attention (Filetti et al., 2019), allocates cognitive resources (Wei and Kang, 2014; Su et al., 2021), and alters cognitive executive function (Qu et al., 2013). It has been proposed that reward expectation improves working memory capacity by encouraging participants to make more efforts to fulfill the working memory tasks (Gilbert and Fiez, 2004; Hopstaken et al., 2016).

However, empirical results have revealed a complex relationship between reward and working memory. Under some circumstances, reward did improve working memory performance (Gilbert and Fiez, 2004; Taylor et al., 2004; Rowe et al., 2008; Beck et al., 2010; Savine et al., 2010; Marquand et al., 2011; Kawasaki and Yamaguchi, 2013; Sandry et al., 2014; Choi et al., 2015; Hammer et al., 2015; Fairclough and Ewing, 2017; Hefer and Dreisbach, 2017; Heritage et al., 2017; Klink et al., 2017; Allen and Ueno, 2018; Anna and Anna, 2018; Thurm et al., 2018; Gaillard et al., 2019; Magis-Weinberg et al., 2019; Manga et al., 2020; Sandry and Ricker, 2020), while other studies did not find the reward effect on working memory accuracy (Pochon et al., 2002; Krawczyk et al., 2007; Beck et al., 2010; Hager et al., 2015; Infanti et al., 2017; Smith et al., 2017; Fairclough et al., 2018; Di Rosa et al., 2019). A potential factor mediating the reward effect is working memory load. For example, the reward effect was pronounced in studies of complex tasks, such as remembering 20 locations (Cho et al., 2018) or maintaining spatial and other features simultaneously (Rowe et al., 2008; Kawasaki and Yamaguchi, 2013; Klink et al., 2017; Allen and Ueno, 2018; Anna and Anna, 2018; Gaillard et al., 2019). In contrast, when the task was relatively simple, such as the eight-arm maze task (Smith et al., 2017), the influence of reward expectation on spatial working memory disappeared. Some researchers directly manipulated working memory load and/or compared performances of participants with different working memory abilities (Taylor et al., 2004; Esteban et al., 2015; Hammer et al., 2015; Thurm et al., 2018; Gaillard et al., 2019). For example, Hammer et al. (2015) required children with attention-deficit/hyperactivity disorder (ADHD) and normally developed children to complete 2-back visual working memory tasks. The behavioral results revealed the reward expectation effect on working memory performance only in the ADHD group but not in the normally developed group. It was proposed that the disappearance of the reward effect was due to the ceiling effect (Savine et al., 2010; Esteban et al., 2015; Hammer et al., 2015).

However, the reward effect is not merely mediated by working memory load (Pochon et al., 2002; Heritage et al., 2017; Fairclough et al., 2018; Gaillard et al., 2019), and participants with better working memory ability have shown the reward effect, while others did not (Thurm et al., 2018; Manga et al., 2020). Hence, working memory load is not sufficient to explain the inconsistent results of the previous studies. There are other factors involved in the relationship between reward and working memory, such as processing level (Ahmed and Fockert, 2012) and other factors (Beck et al., 2010; Savine et al., 2010; Choi et al., 2015; Hammer et al., 2015; Heritage et al., 2017; Fairclough et al., 2018; Magis-Weinberg et al., 2019).

According to the load theory, the effectiveness of voluntary attention is impaired when working memory load is high, because of exhausting cognitive resources (Lavie et al., 2004). However, according to Ahmed and Fockert (2012), this theory is valid only when the task requires relative local-level processing, while the effect reverses when global-level processing is required. Ahmed and Fockert (2012) suggested that the ability to effectively concentrate attention to relevant local visual fields is reduced when working memory load is high. In typical spatial working memory studies, increment of spatial working memory load always couples with a more localized requirement of processing. Compared with low spatial working memory load condition, the visual field in high working memory load condition is divided into relatively small areas, and attention must be constrained to a more local visual field to take in distinct representations of multiple locations (Saarinen, 1988), while the effectiveness of voluntary attention deteriorates when the working memory load is high. Thus, the increase in working memory requirement is not only a burden to the maintenance of spatial information but also weakens the encoding of spatial locations.

The interaction between the effect of reward and the effect of working memory load may reflect the unstable reward effect on working memory. The reward effect improves working memory performance by enhancing voluntary attention (Gilbert and Fiez, 2004; Pessoa, 2009), but a high working memory load undermines the precision of spatial representations (Ahmed and Fockert, 2012). Based on these findings, we hypothesized that reward enhances spatial working memory performance at the global level, but working memory performance at the local level benefits less from reward or, even worse, the reward effect vanishes.

To verify this hypothesis, we applied a sequential memory-guided saccade task under different reward conditions. Compared with other delay-match tasks, the memory-guided saccade task is more flexible and accountable (Funahashi et al., 1993) and has been widely applied in studies of spatial working memory (Funahashi et al., 1993; Park et al., 1995; Sawaguchi and Iba, 2001; Johnston and Everling, 2008; Tsujimoto and Postle, 2012). The accuracy of saccade relies on precise spatial representation (Vergilino and Beauvillain, 2001; Theeuwes et al., 2009). The memory of sequential locations consists of location and sequence information. Sequential information is organized in either time or space, which are compatible (Fischer-Baum and Benjamin, 2014). Serial-order memory is highly connected with spatial working memory, and sequence information is represented in the form of space (van Dijck et al., 2013; Antoine et al., 2018).

In this study, spatial working memory performance was assessed both at the global and local level of similarity between stimuli sequence and scan path. In addition to the saccade landing point, fixation duration was calculated as a measurement of cognitive effort devoted to the task. Eye movements reflect the status of attention (Rayner, 1978, 2009; Theeuwes et al., 2009), and fixation duration is a valid indicator of attention (Rayner et al., 2007; Papageorgiou et al., 2014). Friedman and Liebelt (1981) found that fixation duration was prolonged when the gazed object was unusual or was required to be remembered. They suggested that the prolonged fixation duration reflects additional attention allocated to the gazed object. Considering that fixation duration covers multiple cognitive procedures, such as intake of foveal information and saccade planning (Rayner, 1998; Ludwig et al., 2014), and is influenced by factors other than cognitive effort (Rayner, 1998; Ludwig et al., 2014), comparing the fixation durations of reward cue and non-cue can provide more details about the impact of reward expectation on working memory.

METHODS

Participants

Twenty right-handed college students (10 men and 10 women) participated in this experiment. All of them had a normal or correct-to-normal vision and color vision. They were compensated after the experiment with a basic amount of money plus a bonus depending on their performance.

Equipment and Materials

Stimuli were presented on a 19-inch CRT monitor with a refresh rate of 120 Hz and a resolution of 1,024 × 768 pixels *via* Visual Basic. The viewing distance was 60 cm. A chin and forehead rest was used to reduce head movements. Eye positions were recorded by SMI Hi-speed eye tracker (SensoMotoric Instruments GmbH, Teltow, Germany) at a sampling rate of 350 Hz. The spatial resolution of the eye tracker was 0.1° of visual angle. A standard nine-point calibration and validation were conducted at the beginning of each block to ensure the eye data quality. Memory arrays consisted of four items, three of which were gray disks, while the fourth, a reward cue, was either a 1-Fen or 1-Yuan or blurred Chinese coin (1 Yuan equals 100 Fen in Chinese currency). The locations of reward cues in memory arrays were randomized. All stimuli were in the size of 5.59° visual angle and had the same luminance (**Figure 1B**). The reward expectation level was assigned to none, low, and high corresponding to reward cues of blurred, 1-Fen, and 1-Yuan Chinese coins (reward cues), respectively.

The screen was divided into 16 grids and equally distributed in four quadrants (four squares in each quadrant). Items (disks or reward cues) were displayed in squares in accordance with the experimental design.

Procedure

A sample test paradigm was used in the experiment. During the encoding phase, a "+" appeared at the center of the screen and disappeared until the subjects fixated it steadily for at least 800 ms. Four pictures, including three gray disks and a reward cue (blurred/1-Fen/1-Yuan Chinese coins), were presented in each quadrant at one of four grids for 1,000 ms, sequentially. The stimulus positions were described by the sequential order of quadrants (denoted by bigger numbers) and the number of grids (labeled by letters; as shown in **Figure 1A**) in which the stimuli were displayed. The participants were asked to gaze at the pictures and to remember their locations and sequential order.

There was a delay of 800 ms before the recall phase. During the recall phase, the participants were instructed to reconstruct the spatial location sequence as precisely as possible in 5,000 ms by gazing on a blank screen. The eye scan path was described by two series of positions of different processing levels. One was the quadrant sequence, indicating the sequential order of eye positions in different quadrants, such as "1234" (denoted by letters in **Figure 1A**). The other one was the sequential positions at grids in each of the quadrants, denoted by letters "ABCD." The recall performance of spatial working memory was calculated by comparing the sequences of stimulus in the encoding phase and eye scan path in the retrieval phase. Similarities of each trial were calculated according to Brandt and Stark (1997) and Eddy (2004), including global similarity (GS), which was based on the quadrant, and local similarity (LS), which was based on the grid. Both GS and LS refer to degrees of similarity between sequences of stimuli positions in the study phase and eye fixation positions in the recall phase, ranging from 0 (totally different) to 1 (the same). GS and LS indicate the recall accuracy of the global and local positions of stimuli, respectively. LS has a finer spatial scale than GS and reflects a more rigid requirement of spatial resolution.

At the end of each trial, a feedback of reward amount (money in Chinese Yuan) was shown for 1,000 ms. Coin pictures were used to indicate locations to be remembered and reward cues (Hager et al., 2015; Di Rosa et al., 2019; Manga et al., 2020). The amount of reward depended on the performance (i.e., GS and LS) and reward conditions of participants and was calculated by the formula $W^*(\text{GS}+\text{LS})/2)$, where $W = 0$ for No-reward, $W = 1$ for Low-reward, and $W = 10$ for High-reward. The unit of it is the Chinese Yuan.

A total of 192 trials in the whole experiment were randomized and assigned into 12 blocks. There were 48 trials in each of the reward expectation conditions (No, Low, and High). The procedure of one trial is shown in **Figure 2**.

Data Preprocessing

Dependent variables in the present research were derived from the fixation data. Fixations were defined using a temporal threshold of 100 ms and a spatial threshold of 2° visual angle, which were calculated offline. Then, the mean fixation number and the mean fixation duration were generated and grouped by the reward expectation conditions and location of fixation. The percentage of trials in which the cued grid was correctly re-fixed was used as the index of reward-cue memory performance. The spatial working memory task performance was evaluated by the sequence similarity proposed by Eddy (2004). Specifically, locations with the same temporal order of stimuli in the study phase and of eye fixation in the recall phase were compared, and each pair of overlapped locations scored 1 point. Sequence similarity was calculated by dividing the total score of four pairs of locations by 4, with a range from 0 (totally wrong) to 1 (totally correct).

RESULTS

Eye Fixations During the Study Phase

During the study phase, three gray disks and a reward cue were presented sequentially one by one. The subjects were instructed to remember their spatial and temporal locations. We segregated the numbers of eye fixations and their durations by whether they were on the cue or non-cue (gray disk; as shown in **Figure 3**).

There was no significant effect of reward expectation or fixation position (on non-cue or cue) on the number of fixations. However, the fixation duration was significantly affected by reward expectation [$M_{\text{No-reward}} = (362 \pm 14)$ ms; $M_{\text{Low-reward}} =$

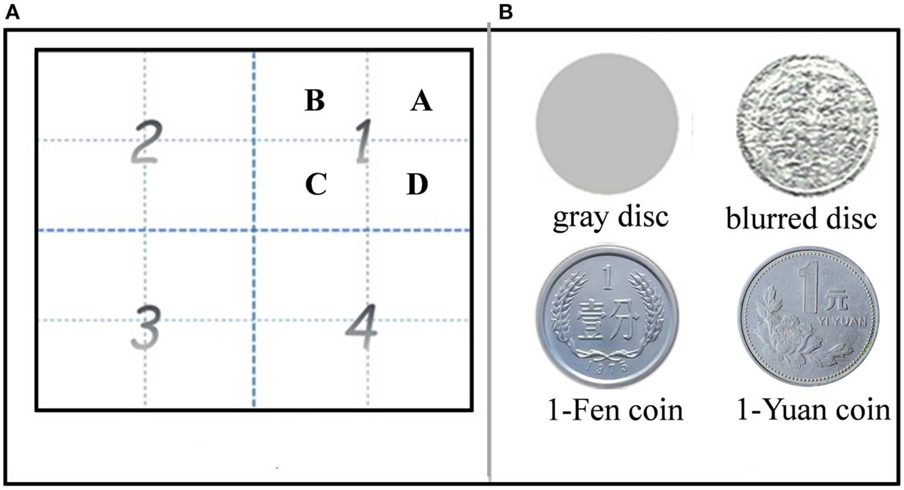


FIGURE 1 | Partition of the screen for stimulus presentation **(A)** and stimuli used in the experiment **(B)**. **(A)** Four quadrants (marked by numbers) were divided into a total of 16 equal-size grids (four per quadrant, labeled by letters). **(B)** Stimuli were gray disk (top left), blurred disk (top right), 1-Fen (bottom left), and 1-Yuan (bottom right) Chinese coins.

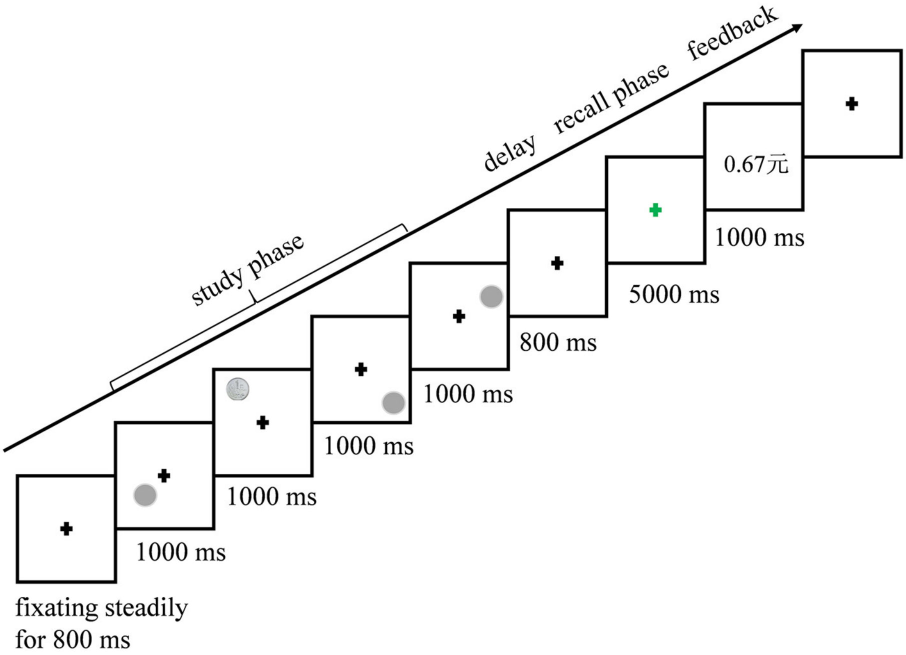


FIGURE 2 | An example of the sequence of events and presentation times.

(355 ± 13) ms; $M_{\text{High-reward}} = (410 \pm 17)$ ms; $F_{(2,\ 38)} = 25.736$, $p < 0.001$, $\eta_p^2 = 0.575$], fixation position [$M_{\text{cue}} = (421 \pm 17)$ ms; $M_{\text{disk}} = (331 \pm 11)$ ms; $F_{(1,\ 19)} = 91.351$, $p < 0.001$, $\eta_p^2 = 0.828$], and their interaction [$F_{(2,\ 38)} = 15.087$, $p < 0.001$, $\eta_p^2 = 0.443$]. Simple effect analysis showed significant differences in fixation duration between the fixations on locations of reward cue and those on locations of non-cue disk under No-reward [$t_{(19)} = 8.714$, $p < 0.001$, Cohen's $d = 1.949$], Low-reward [$t_{(19)} = 4.810$, $p < 0.001$, Cohen's $d = 1.075$], and High-reward expectation conditions [$t_{(19)} = 7.005$, $p < 0.001$, Cohen's $d = 1.566$]. Reward expectation effects were significant for fixation durations of both reward cue [$F_{(2,\ 38)} = 6.020$, $p < 0.001$, $\eta_p^2 = 0.241$] and non-cue disks [$F_{(2,\ 38)} = 23.120$, $p < 0.001$, $\eta_p^2 = 0.549$]. For the non-cue disks, fixation durations of High-reward cue were significantly longer than those of Low- [$t_{(19)} = 2.174$, $p = 0.043$, Cohen's $d = 0.486$] and No-reward cues [$t_{(19)} = 4.077$, $p < 0.001$, Cohen's $d =$

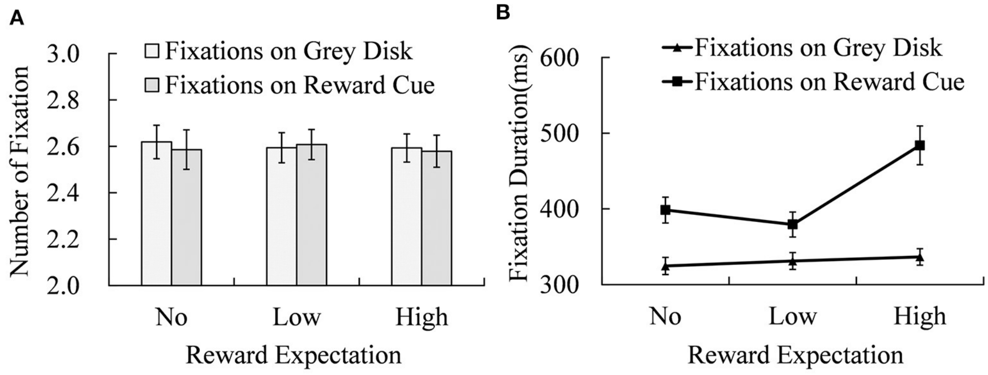


FIGURE 3 | The numbers of fixation **(A)** and mean fixation duration **(B)** on cue and non-cue under different reward expectation conditions. Error bars indicate the standard error of mean (SEM).

0.912]. The difference in fixation durations between No-reward and Low-reward cues was marginally significant [$t_{(19)} = 2.057$, $p = 0.054$, Cohen's $d = 0.460$]. When reward cues were fixated, the same pattern appeared. Fixation durations of High-reward cues were significantly greater than those of No-reward [$t_{(19)} = 4.266$, $p < 0.001$, Cohen's $d = 0.954$] and Low-reward cues [$t_{(19)} = 5.596$, $p < 0.001$, Cohen's $d = 1.251$], and there was no significant difference between No- and Low-reward cues.

Eye Fixations on the Location of Reward Cue During the Retrieval Phase

During the retrieval phase, the percentage of cued grids that were correctly fixated (as shown in **Figure 4A**) and the fixation durations of fixated or unfixated cued grids under different reward expectation conditions were calculated (as shown in **Figure 4B**).

Analysis of variance (ANOVA) of the percentage of trials in which cued grids were fixated (as shown in **Figure 4A**) revealed a significant effect of reward expectation [$M_{\text{No-reward}} = 51.70\% \pm 1.60\%$; $M_{\text{Low-reward}} = 46.40\% \pm 1.40\%$; $M_{\text{High-reward}} = 29.20\% \pm 2.50\%$; $F_{(2,\,38)} = 44.800$, $p < 0.001$, $\eta_p^2 = 0.702$]. Multiple comparisons showed that the No-reward cued grid was fixated more frequently than the Low-reward [$t_{(19)} = 2.731$, $p < 0.05$, Cohen's $d = 0.886$] and High-reward cued grids [$t_{(19)} = 9.461$, $p < 0.001$, Cohen's $d = 3.070$], while the Low-reward cued grid received more fixations than the High-reward cued grid [$t_{(19)} = 5.702$, $p < 0.001$, Cohen's $d = 1.850$].

A 2 (fixation on cued grid and non-cued grid) × 3 (No-, Low-, and High-reward cue) repeated-measures ANOVA of fixation duration revealed a significant effect of fixation location [$M_{\text{cued}} = (441 \pm 22)$ ms; $M_{\text{uncued}} = (461 \pm 19)$ ms; $F_{(1,\,19)} = 6.798$, $p = 0.017$, $\eta_p^2 = 0.263$], reward expectation [$M_{\text{No-reward}} = (470 \pm 19)$ ms; $M_{\text{Low-reward}} = (442 \pm 21)$ ms; $M_{\text{High-reward}} = (441 \pm 23)$ ms; $F_{(2,\,38)} = 5.808$, $p = 0.019$, $\eta_p^2 = 0.234$], and their interaction [$F_{(2,\,38)} = 22.793$, $p < 0.001$, $\eta_p^2 = 0.545$]. Simple effect analysis showed significant differences in fixation durations between cued and non-cued grids under No- [$F_{(1,\,19)} = 33.280$, $p < 0.001$, $\eta_p^2 = 0.637$], Low- [$F_{(1,\,19)} = 23.640$, $p < 0.001$, $\eta_p^2 = 0.554$], and High-reward expectation conditions [$F_{(1,\,19)} = 11.910$, $p < 0.001$, $\eta_p^2 = 0.385$]. Reward expectation effects were significant for mean fixation durations on both cued grid [$F_{(2,\,38)} = 11.710$, $p < 0.010$, $\eta_p^2 = 0.381$] and non-cued grid [$F_{(2,\,38)} = 15.11$, $p < 0.001$, $\eta_p^2 = 0.443$]. Duration of fixation on non-cued grid of No-reward expectation was significantly shorter than that of Low-reward [$t_{(19)} = -5.051$, $p < 0.001$, Cohen's $d = 1.639$] or High-reward expectation conditions [$t_{(19)} = -3.717$, $p = 0.001$, Cohen's $d = 1.206$]. There was no significant difference in fixation duration between High- and Low-reward expectations. For the cued grid, the fixation duration of No-reward expectation was significantly longer than that of High- [$t_{(19)} = 6.715$, $p < 0.001$, Cohen's $d = 2.179$] and Low-reward expectation conditions [$t_{(19)} = 3.825$, $p = 0.001$, Cohen's $d = 1.241$], and there was no significant difference between High- and Low- reward expectation conditions.

The Retrieval Performance of Spatial Working Memory Task

Spatial working memory performances (both GS and LS between the sequence of stimuli positions and eye fixation positions in the recall phase) under different conditions are shown in **Figure 5**.

For the GS, a 2 (cued grid fixated and unfixated) × 3 (No-, Low-, and High-reward cue) repeated-measures ANOVA revealed a significant effect of reward expectation [$M_{\text{No-reward}} = 0.601 \pm 0.012$; $M_{\text{Low-reward}} = 0.601 \pm 0.010$; $M_{\text{High-reward}} = 0.625 \pm 0.011$; $F_{(2,\,38)} = 7.078$, $p = 0.002$, $\eta_p^2 = 0.271$]. The interaction between the cued grid and reward expectation is also significant [$F_{(2,\,38)} = 4.356$, $p = 0.020$, $\eta_p^2 = 0.187$]. Simple effect analysis revealed a significant reward expectation effect when the locations of reward were successfully re-fixated [$F_{(2,\,38)} = 12.095$, $p < 0.001$, $\eta_p^2 = 0.389$] but not when the locations of reward cue were not re-fixated. Further analysis showed that, when the locations of reward cue were successfully re-fixated, there were significant differences in GS between the Low-reward condition and the High-reward condition [$t_{(19)} = 3.788$, $p = 0.002$, Cohen's $d = 0.847$] and between the No-reward condition and the High-reward condition [$t_{(19)} = 4.611$, $p < 0.001$, Cohen's $d = 1.031$]. However, there was no significant difference between the No-reward condition and the Low-reward condition. The difference

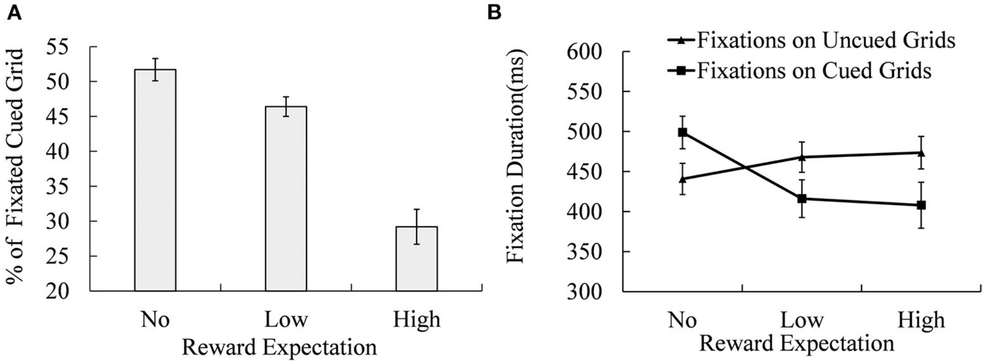


FIGURE 4 | Percentage of the fixated cued locations **(A)** and the fixation durations of both the cued grids were fixated and unfixated **(B)** under different reward expectation conditions. Error bars indicate the standard error of mean (SEM).

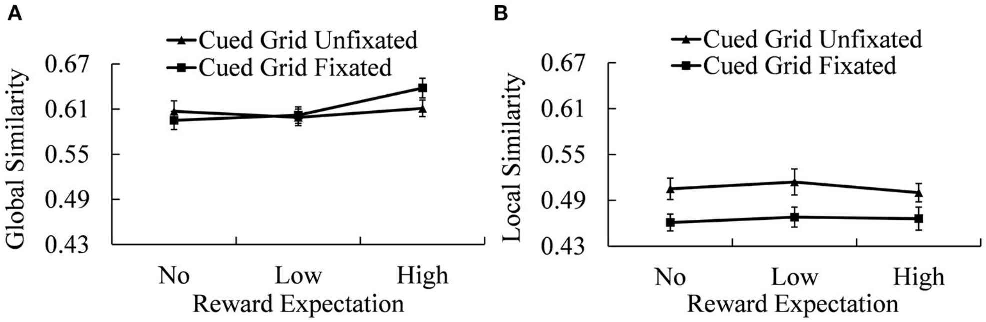


FIGURE 5 | Recall scores of GS **(A)** and LS **(B)** under different reward expectation conditions. Error bars indicate the standard error of mean (SEM).

of GS between conditions when the cued grid was fixated and when the cued grid was unfixated was significant in High-reward expectation condition [$t_{(19)} = 2.813$, $p = 0.011$, Cohen's $d = 0.629$], but not in the other two reward expectation conditions ($p > 0.100$).

Local similarity showed a different statistical pattern, that is, 2 (cued grid fixated and unfixated) × 3 (No-, Low-, and High-reward cue) repeated-measures ANOVA revealed a significant main effect of re-fixation [$M_{\text{unfixated}} = 0.521 \pm 0.012$; $M_{\text{fixated}} = 0.465 \pm 0.010$; $F_{(2,\,38)} = 26.413$, $p < 0.001$, $\eta_p^2 = 0.582$], showing that when the cued grids failed to be re-fixated, scores of local similarity went up. Other effects on LS were not significant.

DISCUSSION

The results of the current research showed that reward expectation affected the encoding and maintenance of temporally organized spatial representations. Four-location temporal sequence was encoded into the working memory, and more attention was paid to locations with high rewards in the study phase. These spatial location representations were maintained in the delay phase and retrieved successively in the reconstruction phase.

Reward Expectation Enhances Voluntary Attention in the Encoding Phase

The results of the study phase showed that stimuli, reward cues or not, in the High-reward expectation condition were fixated longer than those in No- and Low-reward expectation conditions (as shown in **Figure 5**). Considering the effect of reward expectation on recall performance in this study and the indecisive relationship between pure fixation duration and working memory performance in previous studies (Saint-Aubin et al., 2007; Oi et al., 2015), the changed fixation duration in the study phase of this study likely reflect differences in attention allocation under different reward expectation conditions. Specifically, the participants paid more attention to the task when reward expectation was high, compared with the No-reward and Low-reward conditions.

The result of fixation duration in the study phase is in line with previous studies that reported that reward expectation prompted encoding of working memory (Wallis et al., 2015; for a review, see Botvinick and Braver, 2015; Klink et al., 2017; Roberts et al., 2017). In their first experiment, Wallis et al. (2015) found that reward improved encoding of reward-associated items, and the reward effect was generalized to other items in the memory list. In this research, fixation durations on gray disks were prolonged in High-reward conditions as well. It seems that the mechanism underlying the reward effect in the encoding phase is more general than the specified encoding of the reward cue. A plausible explanation is that reward expectation enhances arousal, which in turn provides more cognitive resources for working memory encoding (Murray, 2007; Murayama and Kitagami, 2014; Unsworth and Robison, 2015), and participants are willing to make more efforts to obtain a higher reward. The null result of fixation number and the longer fixation duration of reward cue in the study phase are predictable. The design of the task in this research restricted the saccadic patterns of participants in the study phase and led to equivalent fixations under all reward expectation levels. These reward cues contained additional reward information compared with gray disks, requiring additional processes.

Reward Effects on Different Processing Levels

The main aim of this research was to explore the reward effect on spatial working memory at different processing levels. As predicted, we found a significant reward effect on GS, the indicator of global processing, but no reward effect on local similarity, the indicator of local processing. Furthermore, the percentage of trials in which the cued grid was correctly re-fixated during the retrieval phase, a relatively local indicator, decreased with the increment of reward. High reward improved spatial working memory performance at the global level but undermined the precision of spatial representations at the local level.

Mean fixation duration in the recall phase is in line with this conclusion. The mean fixation duration in memory-guided saccade tasks is affected by the strength of the memory trace, and it requires more time and effort to recall a weak memory trace (Burke et al., 2012; Haj et al., 2017; Dang et al., 2021). Therefore, when the memory of the next saccade target is weak, prolonged fixation duration is required to generate the following saccade (Meghanathan et al., 2020). Eye-tracking data of this research showed that fixation on uncued locations in the recall phase, sometimes followed by a saccade to the cued location, prolonged as the reward expectation was higher. A potential explanation is that, in this research, the memory for the cued location was degraded when the reward expectation was high, and more time was required before the saccade to the cued location could be generated.

The results of similarities at different processing levels and the negative reward effect on the representation of cued location are consistent with the hypothesis made by Ahmed and Fockert (2012), that is, the authors suggested that working memory load modulates selective attention to different levels of the same stimulus. When working memory load is high, information at a more global level is easily selected, while local-level information is ignored. It is difficult to constrain attention to the local level with a high working memory load. As a result, reward expectation prompted global similarity but did not affect local similarity in this study.

Modulation of Sustained Control on Spatial Working Memory Representation

An interesting finding of this research is the trade-off between different indicators. As mentioned above, the mean fixation duration of the recall phase reflects degraded representations of cued locations and promoted representations of uncued locations with the increment of reward. Local similarities of trials in which the cued location was falsely re-fixated were higher than local similarities of trials in which the cued location was successfully re-fixated. It seems that the improvement of representations of uncued locations and local similarities comes at the cost of representation of cued locations.

These results of cued-location memory are consistent with the proposal of flexible attention theory (Sandry et al., 2014; Sandry and Ricker, 2020), which suggests that the cognitive system assigns attention resources flexibly among items in working memory, and the elevation of working memory performance of a certain item is at the cost of performance of other items (Sandry et al., 2014; Allen and Ueno, 2018; Sandry and Ricker, 2020). In this research, to obtain a higher reward, goal-directed cognitive control may have inhibited the maintenance of reward-cue location and allocated saved resources to the maintenance of other locations in the delay phase. Moreover, higher local similarities were observed in the trials in which the cued location was falsely re-fixated.

Limitations of This Study

During the study phase of this study, all three conditions (No-, Low-, and High-) of the reward cue were randomly presented at different temporal positions. It may lead to a confounding effect. The different intervals of processing the reward cue mean that the levels of processing or the motivational states may vary accordingly. Specifically, motivational states might be identical in the three reward conditions until the presentation of the reward cue. Moreover, according to Sandry and Ricker (2020), the temporal position does affect performance in working memory tasks. Hence, the temporal position of reward cues should be considered in future studies.

CONCLUSION

By applying sequential memory-guided eye movement tasks, we reached the following conclusions regarding the reward effect on spatial working memory: (a) reward expectation enhances the encoding of spatial locations by improving voluntary attention, and (b) reward affects reconstruction only at the global level but not at the local level, in which the cognitive resource is reallocated among reward- and non-reward-associated items in working memory to maximize the reward effect.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Psychological Ethics Committee of Capital Normal University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JD and QZ discussed and developed the idea and wrote the introduction, methods, results, and discussion. ZJ designed the study with input from JD and QZ and ran the experiment on all participants. JD and ZJ were highly involved in the analysis of the results. All authors contributed to the article and approved the submitted version.

REFERENCES

Ahmed, L., and Fockert, J. W. (2012). Working memory load can both improve and impair selective attention: evidence from the navon paradigm. *Attent. Percept. Psychophys*. 74, 1397–1405. doi: 10.3758/s13414-012-0357-1

Allen, R. J., and Ueno, T. (2018). Multiple high-reward items can be prioritized in working memory but with greater vulnerability to interference. *Attent. Percept. Psychophys*. 80, 1–13. doi: 10.3758/s13414-018-1543-6

Anna, H., and Anna, S. (2018). Separate and combined effects of action relevance and motivational value on visual working memory. *J. Vis*. 18, 1–14. doi: 10.1167/18.5.14

Antoine, S., Ranzini, M., van Dijck, J., Slama, H., Bonato, M., and Tousch, A., et al. (2018). Hemispatial neglect and serial order in verbal working memory. *J. Neuropsychol*. 13, 272–288. doi: 10.1111/jnp.12145

Barrett, L. F., Tugade, M. M., and Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychol. Bullet*. 130, 553–573. doi: 10.1037/0033-2909.130.4.553

Beck, S. M., Locke, H. S., Savine, A. C., Jimura, K., and Braver, T. S. (2010). Primary and secondary rewards differentially modulate neural activity dynamics during working memory. *PLoS ONE* 5:e9251. doi: 10.1371/journal.pone.0009251

Botvinick, M. M., and Braver, T. (2015). Motivation and cognitive control: from behavior to neural mechanism. *Ann. Rev. Psychol*. 66, 83–113. doi: 10.1146/annurev-psych-010814-015044

Brandt, S. A., and Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *J. Cogn. Neurosci*. 9, 27–38. doi: 10.1162/jocn.1997.9.1.27

Burke, M. R., Allen, R. J., and Gonzalez, C. (2012). Eye and hand movements during reconstruction of spatial memory. *Perception* 41, 803–818. doi: 10.1068/p7216

Cho, Y. T., Lam, N. H., Starc, M., Santamauro, N., Savic, A., and Diehl, C. K., et al. (2018). Effects of reward on spatial working memory in schizophrenia. *J. Abnorm. Psychol*. 127, 695–709. doi: 10.1037/abn0000369

Choi, J. M., Padmala, S., and Pessoa, L. (2015). Counteracting effect of threat on reward enhancements during working memory. *Cogn. Emot*. 29, 1517–1526. doi: 10.1080/02699931.2014.993596

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci*. 24, 87–185. doi: 10.1017/S0140525X01003922

Dang, X. X., Li, L. F., Chen, Y. H., and Yang, X. J. (2021). Distinguishing between high-confidence true and false memories: evidence from eye movements. *Austr. J. Psychol*. 11, 1–12. doi: 10.1080/00049530.2020.1865779

Di Rosa, E., Brigadoi, S., Cutini, S., Tarantino, V., Dell'Acqua, and, R., Mapelli, D., et al. (2019). Reward motivation and neurostimulation interact to improve working memory performance in healthy older adults: a simultaneous tdcs-fNIRS study. *NeuroImage* 202:116062. doi: 10.1016/j.neuroimage.2019.116062

Eddy, S. R. (2004). What is dynamic programming? *Nat. Biotechnol*. 22, 909–910. doi: 10.1038/nbt0704-909

Esteban, L., Vivas, A. B., Fuentes, L. J., and Estévez, A. F. (2015). Spatial working memory is enhanced in children by differential outcomes. *Sci. Rep*. 5:17112. doi: 10.1038/srep17112

Fairclough, S. H., Burns, C., and Kreplin, U. (2018). Fnirs activity in the prefrontal cortex and motivational intensity: impact of working memory load, financial reward, and correlation-based signal improvement. *Neurophotonics* 5:035001. doi: 10.1117/1.NPh.5.3.035001

Fairclough, S. H., and Ewing, K. (2017). The effect of task demand and incentive on neurophysiological and cardiovascular markers of effort. *Int. J. Psychophysiol*. 119, 58–66. doi: 10.1016/j.ijpsycho.2017.01.007

Filetti, M., Barral, O., Jacucci, G., and Ravaja, N. (2019). Motivational intensity and visual word search: layout matters. *PLoS ONE* 14:e0218926. doi: 10.1371/journal.pone.0218926

Fischer-Baum, S., and Benjamin, A. S. (2014). Time, space, and memory for order. *Psychon. Bullet. Rev*. 21, 1263–1271. doi: 10.3758/s13423-014-0604-7

Friedman, A., and Liebelt, L. S. (1981). On the time course of viewing pictures with a view towards remembering. *Eye Mov. Cogn. Vis. Percept*. 12, 137–155. doi: 10.4324/9781315437415-12

Funahashi, S., Bruce, C. J., and Goldmanrakic, P. S. (1993). Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic "scotomas". *J. Neurosci.* 13, 1479–1497. doi: 10.1523/JNEUROSCI.13-04-01479.1993

Gaillard, C., Guillod, M., Ernst, M., Torrisi, S., Federspiel, A., and Schoebi, D., et al. (2019). Striatal responsiveness to reward under threat of shock and working memory load: a preliminary study. *Brain Behav*. 9:1397. doi: 10.1002/brb3.1397

Gilbert, A. M., and Fiez, J. A. (2004). Integrating rewards and cognition in the frontal cortex. *Cogn. Affect. Behav. Neurosci*. 4, 540–552. doi: 10.3758/CABN.4.4.540

Hager, O. M., Dirschner, M., Bischof, M., Hartmann-Riemer, M. N., Kluge, A., and Seifritz, E., et al. (2015). Reward-dependent modulation of working memory is associated with negative symptoms in schizophrenia. *Schizophr. Res*. 168, 238–244. doi: 10.1016/j.schres.2015.08.024

Haj, M. E., Nandrino, J. L., Antoine, P., Boucart, M., and Lenoble, Q. (2017). Eye movement during retrieval of emotional autobiographical memories. *Acta Psychol*. 174, 54–58. actpsy.2017.02.002 doi: 10.1016/j.actpsy.2017.02.002

Hammer, R., Tennekoon, M., Cooke, G. E., Gayda, J., Stein, M. A., and Booth, J. R. (2015). Feedback associated with expectation for larger-reward improves visuospatial working memory performances in children with ADHD. *Dev. Cogn. Neurosci*. 14, 38–49. doi: 10.1016/j.dcn.2015.06.002

Hefer, C., and Dreisbach, G. (2017). How performance-contingent reward prospect modulates cognitive control: increased cue maintenance at the cost of decreased flexibility. *J. Exp. Psychol. Learn. Mem. Cogn*. 43, 1643–1658. doi: 10.1037/xlm0000397

Heitz, R. P., Schrock, J. C., Payne, T. W., and Engle, R. W. (2008). Effects of incentive on working memory capacity: behavioral and pupillometric data. *Psychophysiology* 45, 119–129. doi: 10.1111/j.1469-8986.2007.00605.x

Heritage, A. J., Long, L. J., Woodman, G. F., and Zald, D. H. (2017). Personality correlates of individual differences in the recruitment of cognitive mechanisms when rewards are at stake. *Psychophysiology* 55:12987. doi: 10.1111/psyp.12987

Hopstaken, J. F., Dimitri, V., Bakker, A. B., Kompier, M., and Leung, Y. K. (2016). Shifts in attention during mental fatigue: evidence from subjective, behavioral, physiological, and eye-tracking data. *J. Exp. Psychol. Hum. Percept. Perform*. 42, 878–889. doi: 10.1037/xhp0000189

Infanti, E., Hickey, C., Menghi, N., and Turatto, M. (2017). Reward-priming impacts visual working memory maintenance: evidence from human electrophysiology. *Vis. Cogn*. 25:1365790. doi: 10.1080/13506285.2017.1365790

Johnston, K., and Everling, S. (2008). Neurophysiology and neuroanatomy of reflexive and voluntary saccades in non-human primates. *Brain Cogn*. 68, 271–283. doi: 10.1016/j.bandc.2008.08.017

Just, M. A., and Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychol. Rev*. 99, 122–149. doi: 10.1037/0033-295X.99.1.122

Kawasaki, M., and Yamaguchi, Y. (2013). Frontal theta and beta synchronizations for monetary reward increase visual working memory capacity. *Soc. Cogn. Affect. Neurosci*. 5, 523–530. doi: 10.1093/scan/nss027

Klink, P. C., Jeurissen, D., Theeuwes, J., Denys, D., and Roelfsema, P. R. (2017). Working memory accuracy for multiple targets is driven by reward expectation and stimulus contrast with different time-courses. *Sci. Rep.* 7:9082. doi: 10.1038/s41598-017-08608-4

Krawczyk, D. C., Gazzaley, A., and D'Esposito, M. (2007). Reward modulation of prefrontal and visual association cortex during an incentive working memory task. *Brain Res*. 1141, 168–177. doi: 10.1016/j.brainres.2007.01.052

Lavie, N., Hirst, A., de Fockert, J. W., and Viding, E. (2004). Load theory of selective attention and cognitive control. *J. Exp. Psychol. Gen*. 133, 339–354. doi: 10.1037/0096-3445.133.3.339

Ludwig, C. J. H., Davies, J. R., and Eckstein, M. P. (2014). Foveal analysis and peripheral selection during active visual sampling. *Proc. Natl. Acad. Sci. U. S. A*. 111, 291–299. doi: 10.1073/pnas.1313553111

Magis-Weinberg, L., Custers, R., and Dumontheil, I. (2019). Rewards enhance proactive and reactive control in adolescence and adulthood. *Soc. Cogn. Affect. Neurosci*. 14, 1–14. doi: 10.1093/scan/nsz093

Manga, A., Vakli, P., and Vidnyánszky, Z. (2020). The influence of anticipated monetary incentives on visual working memory performance in healthy younger and older adults. *Sci. Rep.* 10:8817. doi: 10.1038/s41598-020-65723-5

Marchette, S. A., Sever, M. W., Flombaum, J. I., and Shelton, A. L. (2015). Individual differences in representational precision predict spatial working memory span. *Spat. Cogn. Comput. Interdiscipl. J*. 15, 308–328. doi: 10.1080/13875868.2015.1078334

Marquand, A. F., De Simoni, S., O'Daly, O. G., Williams, S. C. R., Moural-Miranda, J., and Mehta, M. A. (2011). Pattern classification of working memory networks reveals differential effects of methylphenidate, atomoxetine, and placebo in healthy volunteers. *Neuropsychopharmacology* 36, 1237–1247. doi: 10.1038/npp.2011.9

Meghanathan, R. N., Leeuwen, C. V., Giannini, M., and Nikolaev, A. R. (2020). Neural correlates of task-related refixation behavior. *Vis. Res*. 175, 90–101. doi: 10.1016/j.visres.2020.07.001

Miller, G. A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol. Rev*. 63, 81–97. doi: 10.1037/h0043158

Murayama, K., and Kitagami, S. (2014). Consolidation power of extrinsic rewards: reward cues enhance long-term memory for irrelevant past events. *J. Exp. Psychol*. 143, 15–20. doi: 10.1037/a0031992

Murray, E. A. (2007). The amygdala, reward and emotion. *Trends Cognitiv*. 11, 489–497. doi: 10.1016/j.tics.2007.08.013

Oi, Y., Ikeda, Y., Okuzumi, H., Kokubun, M., Hamada, T., and Sawa, T. (2015). Fixation effects on forward and backward recall in a spatial working memory task. *Psychology* 6, 727–733. doi: 10.4236/psych.2015.66071

Papageorgiou, K. A., Smith, T. J., Wu, R., Johnson, M. H., Kirkham, N. Z., and Ronald, A. (2014). Individual differences in infant fixation duration relate to attention and behavioral control in childhood. *Psychol. Sci*. 25, 1371–1379. doi: 10.1177/0956797614531295

Park, S., Holzman, P. S., and Goldman-Rakic, P. S. (1995). Spatial working memory deficits in the relatives of schizophrenic patients. *Archiv. Gen. Psychiatr*. 52, 821–828. doi: 10.1001/archpsyc.1995.03950220031007

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends Cogn. Sci*. 13, 160–166. doi: 10.1016/j.tics.2009.01.006

Pochon, J. B., Levy, R., Fossati, P., Lehericy, S., Poline, J. B., et al. (2002). The neural system that bridges rewards and cognition in humans: an fMRI study. *Proc. Natl. Acad. Sci. U. S. A*. 99, 5669–5674. doi: 10.1073/pnas.082111099

Qu, L., Finestone, D. L., Qin, L. J., and Reena, Z. X. (2013). Focused but fixed: the impact of expectation of external rewards on inhibitory control and flexibility in preschoolers. *Emotion* 13, 562–572. doi: 10.1037/a0027263

Rayner, K. (1978). Eye movements in reading and information processing. *Psychol. Bullet*. 85, 618–660. doi: 10.1037/0033-2909.85.3.618

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bullet*. 124, 372–422. doi: 10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quart. J. Exp. Psychol*. 62, 1457–1506. doi: 10.1080/17470210902816461

Rayner, K., Li, X., Williams, C. C., Cave, K. R., and Well, A. D. (2007). Eye movements during information processing tasks: individual differences and cultural effects. *Vis. Res*. 47, 2714–2726. doi: 10.1016/j.visres.2007.05.007

Roberts, W. A., Macdonald, H., Brown, L., and Macpherson, K. (2017). Release from proactive interference in rat spatial working memory. *Learn. Behav*. 45, 263–275. doi: 10.3758/s13420-017-0263-4

Rowe, J. B., Eckstein, D., Braver, T., and Owen, A. M. (2008). How does reward expectation influence cognition in the human brain? *J. Cogn. Neurosci*. 20, 1980–1992. doi: 10.1162/jocn.2008.20140

Saarinen, J. (1988). Perception of spatial order in extrafoveal vision. *J. Psychol.* 29, 162–167. doi: 10.1111/j.1467-9450.1988.tb00787.x

Saint-Aubin, J., Tremblay, S., and Jalbert, A. (2007). Eye movements and serial memory for visual-spatial information: does time spent fixating contribute to recall? *Exp. Psychol*. 54, 264–272. doi: 10.1027/1618-3169.54.4.264

Sandry, J., and Ricker, T. J. (2020). Prioritization within visual working memory reflects a flexible focus of attention. *Attent. Percept. Psychophys*. 82, 2985–3004. doi: 10.3758/s13414-020-02049-4

Sandry, J., Schwark, J. D., and Macdonald, J. (2014). Flexibility within working memory and the focus of attention for sequential verbal information does not depend on active maintenance. *Mem. Cogn*. 42, 1130–1142. doi: 10.3758/s13421-014-0422-1

Savine, A. C., Beck, S. M., Edwards, B. G., Chiew, K. S., and Braver, T. S. (2010). Enhancement of cognitive control by approach and avoidance motivational states. *Cogn. Emot*. 24, 338–356. doi: 10.1080/02699930903381564

Sawaguchi, T., and Iba, M. (2001). Prefrontal cortical representation of visuospatial working memory in monkeys examined by local inactivation with muscimol. *J. Neurophysiol*. 86, 2041–2053. doi: 10.1152/jn.2001.86.4.2041

Smith, A. E., Dalecki, S. J., and Crystal, J. D. (2017). A test of the reward-value hypothesis. *Anim. Cogn*. 20, 215–220. doi: 10.1007/s10071-016-1040-z

Su, Z., Wang, L., Kang, G., and Zhou, X. (2021). Reward makes the rhythmic sampling of spatial attention emerge earlier. *Attent. Percept. Psychophys*. 83:5. doi: 10.3758/s13414-020-02226-5

Taylor, S. F., Welsh, R. C., Wager, T. D., Phan, K. L., Fitzgerald, K. D., and Gehring, W. J. (2004). A functional neuroimaging study of motivation and executive function. *Neuroimage* 21, 1045–1054. doi: 10.1016/j.neuroimage.2003.10.032

Theeuwes, J., Belopolsky, A. V., and Olivers, C. N. L. (2009). Interactions between working memory, attention, and eye movements. *Acta Psychol*. 132, 106–114. doi: 10.1016/j.actpsy.2009.01.005

Thurm, F., Zink, N., and Li, S. C. (2018). Comparing effects of reward anticipation on working memory in younger and older adults. *Front. Psychol.* 9:2318. doi: 10.3389/fpsyg.2018.02318

Tsujimoto, S., and Postle, B. R. (2012). The prefrontal cortex and oculomotor delayed response: a reconsideration of the "mnemonic scotoma". *J. Cogn. Neurosci*. 24, 627–635. doi: 10.1162/jocn_a_00171

Unsworth, N., and Robison, M. K. (2015). Individual differences in the allocation of attention to items in working memory: evidence from pupillometry. *Psychon. Bullet. Rev*. 22, 757–765. doi: 10.3758/s13423-014-0747-6

van Dijck, J. P., Abrahamse, E. L., Majerus, S., and Fias, W. (2013). Spatial attention interacts with serial-order retrieval from verbal working memory. *Psychol. Sci.* 24, 1854–1859. doi: 10.1177/0956797613479610

Vergilino, D., and Beauvillain, C. (2001). Reference frames in reading: evidence from visually and memory-guided saccades. *Vis. Res*. 41, 3547–3557. doi: 10.1016/S0042-6989(01)00225-5

Wallis, G., Stokes, M. G., Arnold, C., and Nobre, A. C. (2015). Reward boosts working memory encoding over a brief temporal window. *Vis. Cogn*. 23, 291–312. doi: 10.1080/13506285.2015.1013168

Wei, P., and Kang, G. (2014). Task relevance regulates the interaction between reward expectation and emotion. *Exp. Brain Res*. 232, 1783–1791. doi: 10.1007/s00221-014-3870-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Psychology

ORIGINAL RESEARCH
published: 26 October 2021
doi: 10.3389/fpsyg.2021.712102

Signalling Three-Way Intersections: Is Redundancy Better Than Only Mandatory or Prohibitory Signs?

Cristina Vargas[1]* *and Sergio Moreno-Ríos*[2]

[1] *ERI-Lectura, University of Valencia, Valencia, Spain,* [2] *The Mind, Brain and Behavior Research Center (CIMCYC-UGR), University of Granada, Granada, Spain*

At intersections, drivers need to infer which ways are allowed by interpreting mandatory and/or prohibitory traffic signs. Time and accuracy in this decision-making process are crucial factors to avoid accidents. Previous studies show that integrating information from prohibitory signs is generally more difficult than from mandatory signs. In Study 1, we compare combined redundant signalling conditions to simple sign conditions at three-way intersections. In Study 2, we carried out a survey among professionals responsible for signposting to test whether common practices are consistent with experimental research. In Study 1, an experimental task was applied ($n = 24$), and in Study 2, the survey response rate was 17%. These included the main cities in Spain such as Madrid and Barcelona. Study 1 showed that inferences with mandatory signs are faster than those with prohibitory signs, and redundant information is an improvement only on prohibitory signs. In Study 2, prohibitory signs were those most frequently chosen by professionals responsible for signposting. In conclusion, the most used signs, according to the laboratory study, were not the best ones for signposting because the faster responses were obtained for mandatory signs, and in second place for redundant signs.

Keywords: mental models, three-way intersections, mandatory sign, prohibitory sign, redundant information

Edited by:
Edwin James Burns,
Edge Hill University, United Kingdom

Reviewed by:
Christoph Hoelscher,
ETH Zürich, Switzerland
Naveen Kashyap,
Indian Institute of Technology Guwahati, India
Joachim Hass,
University of Heidelberg, Germany

Correspondence:
Cristina Vargas
cristina.vargas@uv.es

Specialty section:
This article was submitted to Cognition,
a section of the journal
Frontiers in Psychology

Received: *19 May 2021*
Accepted: *23 September 2021*
Published: *26 October 2021*

Citation:
Vargas C and Moreno-Ríos S (2021) Signalling Three-Way Intersections: Is Redundancy Better Than Only Mandatory or Prohibitory Signs? Front. Psychol. 12:712102.
doi: 10.3389/fpsyg.2021.712102

INTRODUCTION

When driving, we need to interpret mandatory and prohibitory traffic signs and make inferences to determine which direction is allowed and which is not. These inferences are made at the same time as many other cognitive activities we are engaged in. Therefore, it is not surprising that, at least under some circumstances, a large number of road accidents occur at intersections (see, for example, Pathivada and Perumal, 2019).

A potential way of reducing accidents at intersections is by applying the most suitable signposting policy to facilitate drivers' inferences. For example, when we arrive at a T-junction where a right-turn is allowed, a valid traffic sign could be a mandatory sign for the right, a prohibitory sign for the left or both signs (a mandatory right-turn sign and a prohibitory left-turn sign). Although these three signing strategies may be equally valid from a legal point of view, the inferences required to decide which route is allowed involve a different burden on the cognitive system. Cognitive theories of thinking show that some inferences call for an intuitive system, aimed at making automatic fast inferences, while others require slow, effortful, more deliberative processing (e.g., Johnson-Laird, 1983; Stanovich, 1999; Evans, 2008; Kahneman, 2011; Khemlani et al., 2018).

In this work, we present further evidence regarding the use of mandatory and/or prohibitory traffic signs at intersections by considering the results of a new laboratory experiment on inference-making. In particular, we compared the effect of mixed redundant mandatory and prohibitory information to single-type conditions. In addition, we surveyed a group of professionals responsible for signposting policies to analyse consistency in the use of mandatory or prohibitory signs across different Spanish cities and also to examine whether their decisions were consistent with the results obtained in laboratory studies.

Previous literature analysed how people interpret mandatory and prohibitory traffic signs under different conditions by using a simple laboratory task (e.g., Castro et al., 2008; Vargas et al., 2011; Roca et al., 2012). In these experimental tasks, participants were generally presented with a traffic scene in which a car approaches a T-junction, with a road to the right and another to the left. A mandatory sign (e.g., right-turn) or a prohibitory sign (e.g., no left-turn) was shown, allowing only one direction. Subsequently, a car was shown in a new scene on one of the two possible roads (e.g., on the road to the right). Participants had to decide as quickly as possible whether the manoeuvre taken was allowed or not-allowed.

This task has been successfully used to analyse how people make inferences, based, in particular, on predictions from the mental model theory (or model theory; Johnson-Laird and Byrne, 2002; see Johnson-Laird and Ragni, 2019). The model theory maintains that propositional and visual premises are converted into iconic representations called mental models. At a T-junction, a mandatory right-turn sign and a prohibitory left-turn sign may be equivalent in that both allow a right-turn (see **Figures 1A,B**), but their initial mental representations (initial models) are different. In the first case (mandatory right), the initial model represented would be

1. “Right”

while in the second case (prohibitory left), the initial model represented would be

2. “[prohibited] Left.”

However, in both cases, an explicit model (i.e., a full representation of each piece of information, including additional information) can be inferred and will be the same for both signs:

3. “Right allowed and left not-allowed.”

Thinking with the initial representations is faster and less error prone than thinking with explicit models, which requires effort (see Khemlani et al., 2018). Therefore, predictions in the experimental task described above were clear: participants would be faster in deciding that a road was allowed when it was signalled by a mandatory right-turn sign and also faster in deciding that it was not-allowed from a prohibitory left-turn sign. Results in different experiments confirmed such predictions. That is, the mandatory sign led to faster responses to the allowed road than the not-allowed one, and the prohibitory sign led to faster responses to the not-allowed road than the allowed one (Castro et al., 2008; Vargas et al., 2011; Roca et al., 2012).

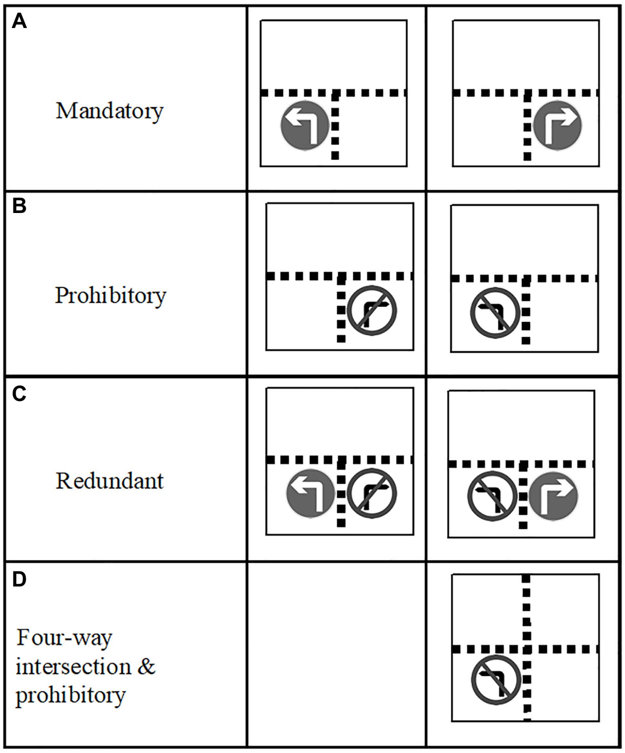


FIGURE 1 | This graph displays types of sign and intersection. **(A)** Two examples of mandatory signs at a T-junction. **(B)** Two examples of prohibitory signs at a T-junction. **(C)** Two examples of redundant information (mandatory sign and prohibitory signs) at a T-junction. **(D)** One example of prohibitory sign at a four-way intersection.

Moreover, different factors that could modulate inferences with these signs have been studied, such as the number of roads at the intersection and the number of traffic signs (Castro et al., 2008), or the exposition time of the signs (Vargas et al., 2011).

Firstly, the advantage of the signalled road persisted at an intersection with four ways (see **Figure 1D**) instead of three (Castro et al., 2008). Also, the presentation of two signs (for example, a double prohibition: for the road ahead and to the right) as two isolated signs or two signs embedded in one did not show any difference in time of response or pattern of results. More interestingly, when people had to collect information from two mandatory signs (in both conditions, two isolated signs and two signs embedded in one), it led to faster responses than from two prohibitory signs. Therefore, for both regulatory signs, it was easier to respond that a turn was allowed than it was to give a not-allowed response (see Castro et al., 2008; Experiment 2). Besides, the disadvantage of using prohibitory signs cannot be explained just in terms of interference between the perceptual arrow direction and readiness to respond to the location tested. Roca et al. (2012; Experiment 2) replicated previous results controlling the Simon effect. Moreover, Castro et al. (2008) used directional and non-directional prohibitory signs. The magnitude of the effect did not differ between them. From a theoretical point of view, the prohibition on acting implies not acting (Johnson-Laird and Ragni, 2019). Some authors have proposed that prohibition requires adding a "mental footnote" (between brackets: "[Prohibited] left") that is not present in the mandatory information (see Bucciarelli and Johnson-Laird, 2005; Vargas et al., 2011). This makes it more difficult to integrate information from premises with prohibition than with mandatory information.

Secondly, in Vargas et al. (2011), the time factor was manipulated in three experiments where encoding time (the time that the first scene was displayed) and signs exposure time (different display times for traffic signs were presented) were assessed in the first scene. The second scene, with the car in the target position, was always shown after the first scene had disappeared. Thus, participants had to make the decision as soon as the first scene disappeared. Participants showed that sign exposure time was irrelevant for making the inference: what was relevant was the total time given for processing the information (encoding time) in the first scene before having to give a response in the second scene. In this case, important differences were obtained according to whether participants had 300 ms to comprehend the sign (encoding time) or 1,000 or 2,000 ms. In all cases, the mandatory sign led to faster responses to the allowed road than the not-allowed one. Prohibitory signs led to the opposite, replicating the previous experiment in the short time condition. However, when participants had enough time (1,000 and 2,000 ms conditions), they showed faster responses to the allowed road than the not-allowed road, as happened with the mandatory signs, but more slowly in both conditions (allowed and not-allowed; Vargas et al., 2011). Results are consistent with the proposal that the negation implied by the prohibition requires (as would "falsity") a conversion of the information for taking one road to the possibility of going the other way, which takes extra time and is error prone (see Bucciarelli and Johnson-Laird, 2005; Vargas et al., 2011; Espino and Byrne, 2018; Moreno-Ríos and Byrne, 2018).

To summarise, the results of the previous laboratory studies show that in simple situations, mandatory signs are better for signing allowed roads than not-allowed roads and the opposite happens for prohibitory signs. However, in general, it is difficult to maintain such an advantage of the prohibitory sign because, if participants have enough time, they will convert the prohibitory information into mandatory information. In complex situations, the integration of prohibitory information is more difficult and takes longer due to the complexity of the tasks required.

In one of the previous studies described above (Castro et al., 2008), the combination of two traffic signs was examined, but the effect of including mixed redundant information was not tested. For example, in our initial example at a T-junction (see **Figure 1C**), both mandatory right-turn and prohibitory left-turn signs could be used to reinforce the same message: only a right-turn is allowed. Obviously, by presenting two traffic signs, the amount of information available is greater, thus increasing the processing requirements and potentially making the task more difficult. Also, participants might check only one of the two signs and the global result could be a combination of responses to the single signs. Consequently, it is uncertain whether the effect of mixed redundancy would be positive or negative in this particular situation. In contrast, T-junctions provide the simplest condition to test the redundancy effect. By using T-junctions rather than including other junctions, some factors can be more easily controlled, such as the same number of different signs (one mandatory and one prohibitory) and ensuring that all the ways are signalled (avoiding the need to infer any other way). In addition, the two signs can be shown in just one location, which is not possible with four-way junctions.

Regarding the overall traffic literature, some previous studies on redundancy have shown that when signs are unfamiliar, the inclusion of a redundant text could improve comprehension of the signs and reduce the time for interpretation (Shinar and Vogelzang, 2013). In addition, there are some mixed results regarding the effect of redundancy in Variable Message Signs, in particular, about whether it reduces compliance with the target detour message (Thomas and Charlton, 2020) or not (Harms et al., 2019).

Looking at the current traffic literature, we considered that no accurate predictions could be made at this point regarding the use of redundancy when presenting mandatory or prohibitory traffic signs at T-junctions. Therefore, we first carried out an experimental study (Study 1), in which we used a task similar to the one applied in previous studies, but now aimed at testing the potential usefulness of a mixed redundant double sign condition (Objective 1).

Second, trying to expand the results found in the laboratory to real situations, we carried out a survey (Study 2) to evaluate current policies on signposting in different cities across Spain. In particular, we tested whether there is agreement among professionals responsible for signposting when they are designing three-way and four-way intersections (Objective 2) and whether those practices are consistent with the reported experimental results (Objective 3).

STUDY 1

Previous studies have shown that the integration of two mandatory signs was easier than the integration of two prohibitory signs (Castro et al., 2008). However, no previous study has been done with redundant information (i.e., a mandatory and a prohibitory sign). From the mental model theory, the two signs lead participants to two initial representations that cannot be directly integrated because they do not share the initial representation, as shown in models (1) and (2) (see section "Introduction"). Only with the complete representation of the explicit model (3) can they be integrated, providing confirmation that both signify the same. Therefore, model theory is useful for making predictions regarding the use of redundancy when combining mandatory and prohibitory traffic signs. From this theoretical point of view, an overall delay is expected regarding the initial representation obtained with the mandatory or prohibitory signs (specifically, allowed for mandatory signs and not-allowed for prohibitory signs conditions) because the redundant condition requires the construction of two models.

In addition, when the response entails accessing the explicit model (that is, inferring the not-allowed road for mandatory signs and the allowed road for prohibitory signs conditions), faster responses would be expected in the redundant condition. These results are expected when participants process both signs, but it is important to note that with redundant information, they could also attend to just one of the two signs systematically (e.g., the mandatory sign). In this case, no differences would be expected between the redundant condition and the simple condition (mandatory or prohibitory sign).

Thus, in this study, we examine the effect of redundancy by comparing three equivalent traffic scenes (mandatory, prohibitory, and redundant signs) at a T-junction. In addition, we will analyse the effect of exposition time (300 vs. 1,000 ms) to test the robustness of the effect found by Vargas et al. (2011).

Materials and Methods

Participants

Twenty-four students (21 females) participated in the experiment. They were either Psychology or Speech and Language Therapy students at the University of Granada (Spain). Average age was 21.2 (*SD* = 4.2). They all had normal or corrected-to-normal vision and received course credits as compensation for their participation.

Stimuli

The procedure used in Study 1 was similar to Experiment 1 in Vargas et al. (2011). T-junction road traffic scenes were used in this experiment. Two consecutive screens were shown in each trial. The first scene was presented with a mandatory, a prohibitory or a redundant (one mandatory and one prohibitory) traffic sign for 300 or 1,000 ms. After that period, a second scene was displayed for a maximum of 2,000 ms or until the participant responded (**Figure 2**). In the first scene, two-thirds of the cases presented a single traffic sign (either mandatory or prohibitory) and the remaining third presented both types of sign (redundant condition). **Figures 1A–C** show the different combinations of signs in this experiment. An E-Prime software (Schneider, 2003) script was developed to control the representation of stimuli and the collection of responses on a 15-in. PC screen.

Procedure

First, the participants read and signed an informed consent form. After that, participants carried out the experiment individually, seated in front of a computer screen. Instructions for the experiment were shown. These instructions explained that the experiment consisted of evaluating the events shown in two consecutive traffic scenes. The participants were informed that the first scene always showed a car on the lower street with various roads it could take and one or two traffic signs. The second scene showed the same car arriving at one of the two other roads at the T-junction (left or right). After that, participants were asked to evaluate whether the road taken by the car was allowed or not-allowed according to information provided by the sign(s).

Participants had to press the "Z" key, labelled as "allowed", as quickly as possible if the manoeuvre was allowed. If the manoeuvre made by the car was not-allowed according to the sign(s), the "M" key, labelled as "not-allowed", had to be pressed. The response key was counterbalanced across participants. Feedback was provided about whether the correct or incorrect response had been performed.

There were 12 experimental conditions defined by combining the time of exposure for traffic signs (300 and 1,000 ms), type of sign (mandatory, prohibitory and redundant) and manoeuvre (not-allowed and allowed) as variables. After reading the instructions, participants performed a block of 48 practice trials (four trials per experimental condition) followed by four blocks of 72 experimental trials (six trials per experimental condition). Thus, the total number of experimental trials was 288 (24 per experimental condition). The order of stimuli presentation was determined randomly for each block.

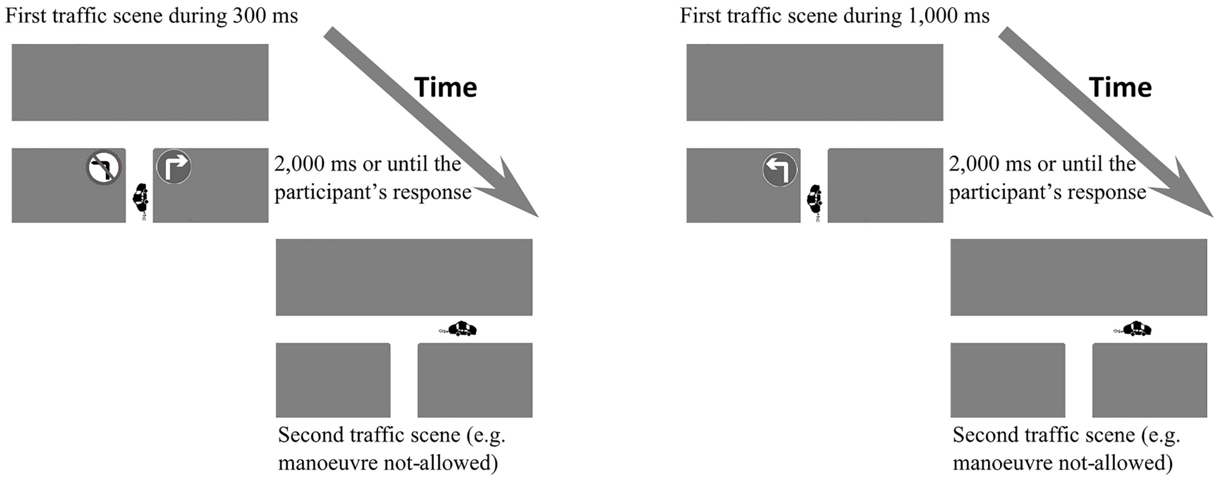


FIGURE 2 | Examples of the consecutive traffic scenes at a T-junction shown in Study 1.

TABLE 1 | Mean correct RTs (ms) and SDs (in parentheses) for each main effect in Study 1.

	Type of sign			Manoeuvre		Overall average
	Mandatory	**Prohibitory**	**Redundant**	**Not-allowed**	**Allowed**	
300 ms	604.9 (181.9)	702.3 (185.3)	642.6 (187.5)	692.7 (189.2)	607.2 (180.1)	649.9 (182.3)
1,000 ms	505.2 (166.7)	609.6 (175.6)	543.9 (178.4)	587.6 (178.6)	518.2 (163.5)	552.9 (170.0)
Overall average	555.1 (171.8)	655.9 (177.8)	593.2 (179.9)	640.2 (181.9)	562.7 (170.1)	601.4 (174.7)

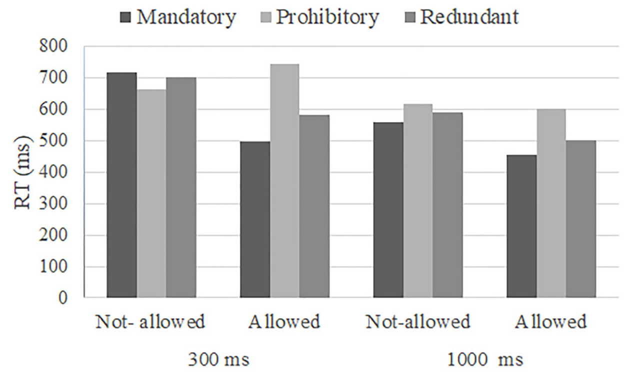


FIGURE 3 | Mean reaction time for the conditions manipulated in Study 1: time of exposure to traffic signs (300 vs. 1,000 ms), type of sign (mandatory vs. prohibitory vs. redundant) and manoeuvres (not-allowed vs. allowed).

Results and Discussion

Trials with RTs above and below three SDs across each participant and condition were excluded from the analysis. This removes the outliers that occur when participants do not follow the instructions or do not perform the task. This criterion resulted in 0.86% of the trials being eliminated from the analysis. A repeated measures ANOVA was conducted to analyse mean RTs for correct responses. The ANOVA included time of exposure to traffic signs (300 vs. 1,000 ms), type of sign (mandatory vs. prohibitory vs. redundant) and manoeuvre (not-allowed vs. allowed). Prior to the ANOVA, data were tested with Mauchly's test of sphericity and degrees of freedom modified as necessary. We used the Bonferroni test to carry out planned comparisons. All analyses were performed using the IBM SPSS Statistics 24 software.

Mean correct response times (RTs) and SDs for each main effect are shown in **Table 1**.

As expected, the second-order interaction (exposure to traffic signs × type of sign × manoeuvre) was statistically significant, $F(2,46) = 25.953$, $p < 0.001$, $\eta^2 = 0.53$ (**Figure 3** shows graphically the second-order interaction). In this context, time of exposure to traffic signs modulates the interaction between type of sign and manoeuvre. These differences are much stronger for the shorter duration (300 ms condition). In order to analyse this second-order interaction, further separate analyses were carried out for the 300 and 1,000 ms conditions, in accordance with Vargas et al. (2011).

The analysis of the 300 ms experimental condition revealed a significant interaction between the type of sign and manoeuvre, $F(1.533, 35.251) = 54.913$, $p < 0.001$, $\eta^2 = 0.71$. We carried out planned comparisons of type of sign and manoeuvre interaction. First, we reproduced previous results by simple contrast (not-allowed vs. allowed for each type of sign: mandatory, prohibitory and redundant). There were statistically significant differences for mandatory signs (*Bonferroni* = 219.314, $p < 0.001$), prohibitory signs (*Bonferroni* = −82.520, $p < 0.001$) and redundant signs (*Bonferroni* = 119.902, $p < 0.001$). On average, participants were faster when an allowed manoeuvre was presented, as compared to a not-allowed manoeuvre, but only for mandatory and redundant signs conditions. In other words, redundant signs have the same effect as mandatory signs. The opposite was true for prohibitory signs (see **Figure 3**). Thus, previous findings were replicated. Second, in order to test Objective 1 in the case study, a simple contrast for manoeuvre (mandatory vs. prohibitory vs. redundant for each manoeuvre) was performed. Results showed statistically significant differences for the allowed condition, but not for the not-allowed condition. The analysis of the allowed condition revealed statistically significant differences between the following experimental conditions: mandatory vs. redundant signs (*Bonferroni* = −87.447, $p < 0.001$), redundant vs. prohibitory signs (*Bonferroni* = −160.871, $p < 0.001$) and mandatory vs. prohibitory signs (*Bonferroni* = −248.318, $p < 0.001$). Consequently, when an allowed manoeuvre was evaluated, participants' responses were faster for mandatory signs, followed by redundant signs, and finally, prohibitory signs. Moreover, the analysis of the 300 ms experimental condition revealed main effects of type of sign [$F(2, 46) = 35.918$, $p < 0.001$, $\eta^2 = 0.61$], and of manoeuvre, [$F(1, 23) = 49.201$, $p < 0.001$, $\eta^2 = 0.68$; see **Table 1**].

Regarding the 1,000 ms experimental condition, the type of sign and manoeuvre interaction was also significant, $F(2, 46) = 9.108$, $p < 0.001$, $\eta^2 = 0.28$. To examine this interaction further, in a similar way to the analysis of the 300 ms condition, we first carried out a simple contrast test for type of sign (not-allowed vs. allowed for each type of sign). There were statistically significant results for mandatory signs (*Bonferroni* = 102.725, $p < 0.001$) and redundant signs (*Bonferroni* = 88.854, $p < 0.001$), but not for prohibitory signs. On average, participants responded faster to allowed manoeuvres for mandatory and redundant signs conditions than to not-allowed manoeuvres. However, there were no differences between manoeuvre types (not-allowed vs. allowed) for prohibitory signs (see **Figure 1**). Thus, the results of the aforementioned studies were replicated once again. Second, a simple contrast for manoeuvre (mandatory vs. prohibitory vs. redundant for each manoeuvre) was performed. Results showed statistically significant differences for both conditions of manoeuvre. For the not-allowed condition, there was a statistically significant difference between mandatory and prohibitory signs (*Bonferroni* = -61.346, $p < 0.01$). That is, participants responded faster to mandatory signs than to prohibitory signs for a not-allowed manoeuvre. Regarding the allowed condition, statistically significant results were found for all experimental comparisons: mandatory vs. redundant signs (*Bonferroni* = -45.583, $p < 0.01$), redundant vs. prohibitory signs (*Bonferroni* = -101.841, $p < 0.001$) and mandatory vs. redundant signs (*Bonferroni* = -147.424, $p < 0.001$). As in the 300 ms condition, participants showed the fastest performance on average when a mandatory sign was presented, followed by redundant signs, and finally, prohibitory signs. Regarding main effects of both type of sign and manoeuvre in the 1,000 ms condition, the analysis revealed significant main effects of the type of sign [$F(2, 46) = 34.317$, $p < 0.001$, $\eta^2 = 0.60$] and of manoeuvre, [$F(1, 23) = 67.797$, $p < 0.001$, $\eta^2 = 0.75$; see **Table 1**].

Finally, the overall main effects of all three independent variables were statistically significant: time of exposure to traffic signs [$F(1, 23) = 105.133$, $p < 0.001$, $\eta^2 = 0.82$], type of sign, [$F(2, 46) = 62.987$, $p < 0.001$, $\eta^2 = 0.73$] and manoeuvre [$F(1, 23) = 75.390$, $p < 0.001$, $\eta^2 = 0.77$]. Hence, on average, participants responded faster to the 1,000 ms than 300 ms condition, mandatory vs. redundant signs, redundant vs. prohibitory signs and allowed vs. not-allowed (**Table 1**).

There was no trade-off effect, that is, no correlation was found between reaction times and accuracy scores. We found accuracy measures with more than 86.9% of answers correct. The low frequency of errors led to a limited window for effects. Therefore, the accuracy measures are not shown in the manuscript.

According to the results of Study 1, the strategy used to process the combined condition differed depending on the manoeuvre being evaluated. For the not-allowed condition, there were no statistically significant differences between redundant and prohibitory signs (initial model) nor between redundant and mandatory signs (explicit model), which suggests that the preference was to specifically focus on one of the two signs of the combined condition.

In contrast, when the manoeuvre being evaluated was allowed, we observed statistically significant differences between the single and the combined signs conditions, which suggests that participants were actually processing both signs. This result is consistent with the predictions of mental model theory in both the 300 and 1,000 ms times of exposure to traffic signs: (a) when the initial model was represented, participants were faster for mandatory than for redundant signs; and (b) the pattern was reversed for the explicit model, that is redundant signs achieved faster responses than prohibitory signs.

In addition, the present outcome replicates previous results regarding the time for processing conditions. As in Vargas et al. (2011), when participants had little time to process the premises, the initial representation of the mandatory sign led participants to react faster for the allowed way than for the not-allowed one, and the opposite result was obtained with prohibitory signs. Results showed that when the two signs were used in this condition, results were similar to those for the mandatory sign (with slightly slower times in all average conditions). Also, as in previous studies, when participants were given longer to process the information (1,000 ms), the advantage for faster responses for the allowed road than the not-allowed remained for mandatory signs, and this also happened with the two redundant signs, but the difference between the two conditions disappeared for the prohibitory signs, which took longer than with the other two kinds of signing.

The most striking feature of these results is that, in the tested conditions, significantly faster responses were generally obtained with the mandatory signs, while there was no advantage for using mixed redundant or prohibitory signs. Regarding prohibitory signs, the performance was not significantly better than when the other two kinds of signing were used, while the redundant signs showed better results than prohibitory ones only in some particular experimental conditions (i.e., allowed condition).

STUDY 2

As suggested by the previous research, there is generally an advantage in using mandatory signs at intersections. Therefore, in Study 2, we surveyed professionals responsible for signposting in provincial capitals of Spain to analyse traffic signs used at intersections and to identify some of the factors modulating their decision-making (e.g., a recent change of direction in the road, accidents reported…). In particular, we tested whether the different professionals responsible for signing: (a) make similar decisions when signalling three and four-way intersections and (b) whether their decisions are consistent with the current experimental research.

Materials and Methods

Participants

We conducted a cross-sectional survey. The population of interest were professionals responsible for signposting in Spanish provincial capitals. An email was sent to administration staff, inviting them to take part in the study, entitled "Use of obligation and prohibition traffic signs". In a 10-week period, nine survey respondents completed all the questionnaires. Hence, the response rate was 17% (response rates to email-only surveys are seldom more than 20%, according to Dillman et al., 2014). These included among others (see section "Results and Discussion"), the most populated cities in Spain (such as Madrid, Barcelona and Valencia, with 14.9 million inhabitants in the three cities and their metropolitan areas in 2020; European Statistical Office, 2021). The average age of participants was 51.8 years (*SD* = 7.5), and they were all men. In addition, the majority of them were engineers.

Procedure and Survey

The survey was an online version, self-completed using LimeSurvey. The first version of the questionnaire was piloted with two road signing professionals (officials from Valencia City Council's Department of Traffic) and an expert in traffic research (University of Valencia). We incorporated all their suggestions. In the final questionnaire, participants provided the following information: sex, age, city, traffic regulations applied (e.g., the General Traffic Regulations), which traffic sign was used more (prohibitory or mandatory) in general, and at three- and four-way intersections, and how often they used mandatory signs in the latter circumstance. In addition, participants were asked to indicate the most frequent signs utilised to obligate drivers in five different traffic situations (e.g., road with high density of traffic). The respondents had several options for response: "prohibitory", "mandatory", "prohibitory and mandatory", "neither sign", "other (to be specified)". All survey respondents answered the same questions in the same order.

Results and Discussion

The following provincial capitals took part in the research: Barcelona, Cádiz, Castelló de la Plana, Córdoba, Logroño, Madrid, Málaga, Valencia and Vitoria-Gasteiz. Eight out of nine respondents used the General Traffic Regulations (Reglamento General de Circulación) for applying mandatory and prohibitory traffic signs. None of the participants chose the categories of response "neither sign" or "other (to be specified)", and therefore, only "prohibitory", "mandatory," and "prohibitory and mandatory" were considered. In general, the prohibitory signs were used slightly more than mandatory signs (55.6 and 44.4%, respectively). Among those who used the most prohibitory signs, the mandatory signs were utilised with a mean value of 25%. This percentage changed according to whether they applied these traffic signs to three-way or four-way intersections. In the first situation, 33.3% used prohibitory signs more frequently than mandatory signs. However, that value increased to 77.8% for four-way intersections. Finally, survey respondents applied different criteria according to the traffic conditions. For example, when there were more complicated traffic situations, the percentage of prohibitory signs used rose (**Figure 4**).

We were also interested in testing whether participants changed their choice of signalling from three-way intersections to four-way, and therefore, we evaluated the consistency of their responses across three-way and four-way intersections (**Table 2**). An exact multinomial test for paired contingency tables was applied and the results showed that it was not symmetrical ($p < 0.001$). In addition, we carried out pairwise comparisons with multiple testing adjustment. According to these analyses, there was a significant change from mandatory to prohibitory signs ($p < 0.001$; Cohen's $g = 0.5$) and from prohibitory to redundant signs ($p = 0.031$, Cohen's $g = 0.5$). The analyses were carried out using the EMT and rcompanion packages implemented in R 3.6.3 (R Core Team, 2020).

Participants responsible for signposting did not follow the same strategies when deciding whether mandatory or prohibitory signs should be used in different conditions. Some of them systematically preferred to use prohibitory signs and others mandatory ones, but an overall preference for using prohibitory signs was found.

This result contrasts with those previously reviewed, obtained in the laboratory, which showed faster processing when inferences were based on mandatory signs than on prohibitory ones (Castro et al., 2008; Vargas et al., 2011; Roca et al., 2012; and Study 1).

In the case of four-way intersections, participants changed their criteria from those used for three-way intersections, increasing the number of prohibitory signs and decreasing the number of mandatory signs (from 18 mandatory signs at three-way intersections, 13 were changed to prohibitory at four-way intersections). In all cases, more prohibitory signs were used. This made it more probable that drivers would need to integrate the prohibitory information, yet mandatory signs are the most informative, given that they eliminate a greater number of alternatives.

Participants decided to use mandatory signs more frequently than prohibitory signs in only two conditions and only at three-way intersections: roads with low traffic density and roads where there was a not-allowed turn (i.e., when the allowed way was the more probable objective of drivers).

Another interesting result is that they decided to use the two signs more often at four-way intersections than at three-way ones. As we have noted in Study 1, in a particular case, using

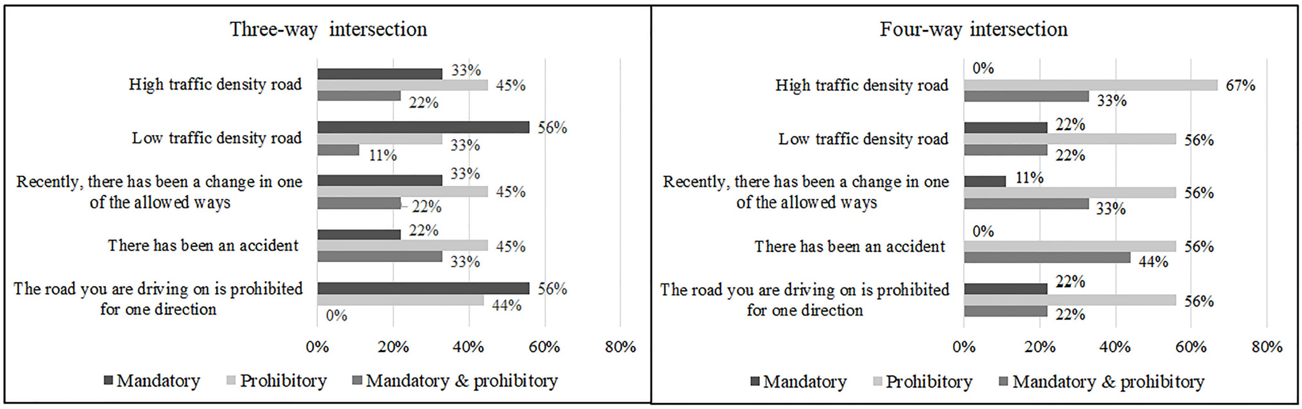


FIGURE 4 | Percentage of choices of signs used most frequently to signpost at three and four-way intersections in different conditions.

TABLE 2 | Cross-tabulation table of traffic signs changed from three-way intersections to four-way intersections.

Three-way intersection	Four-way intersection		
	Mandatory	Prohibitory	Mandatory and prohibitory
Mandatory	5	13	0
Prohibitory	0	13	6
Mandatory and prohibitory	0	0	8

redundant signs showed better results than using prohibitory signs (i.e., allowed condition). The use of these signs happens more frequently when an accident has occurred there. In those situations, it is possible that visual material effects of the accident remain at the traffic scene as potential distractors. As we have seen in the laboratory, information from the signs most often used in these situations takes longer to be integrated (Roca et al., 2012).

GENERAL DISCUSSION

One way to reduce the probability of accidents is by facilitating the interpretation of traffic signs. As we have seen, there are three ways to signal a T-junction: with a mandatory sign, with a prohibitory sign or with both (redundant condition). The result in all three cases is the same: one way is allowed and the other not. However, they are not cognitively equivalent. Previous studies (Castro et al., 2008; Vargas et al., 2011; Roca et al., 2012) in the laboratory have shown that inferences about allowed and not-allowed are faster when mandatory signs rather than prohibitory signs are used, other than in exceptional situations. Study 1 in this work showed that presenting the two signs, giving redundant information, was no better than presenting just the mandatory sign, although in some situations, when a prohibitory sign was used, adding a redundant mandatory sign could be useful. That is, a faster time was obtained in some conditions for the two signs in comparison with just the prohibitory sign and in no condition was the time longer for the combined signs. This result could contribute to restricting the context of usefulness of giving redundant information. The use of reiterative information could produce some negative effects regarding compliance with Variable Message Signs (Thomas and Charlton, 2020). In the context of evacuation signalling, Kwee-Meier et al. (2019) showed that the use of prohibition added to the allowed direction created confusion and should be avoided.

In Study 2, the results suggest low levels of agreement among those responsible for signing. Actually, some of them said they did not have a theoretical or empirical base from which to follow a clear criterion. Despite the apparent lack of agreement, the average frequency is higher for prohibitory signs and in a few cases for double signing. The differences were even greater for four-way intersections. It is important to note that these results contrast with those obtained in the laboratory, which showed faster processing when inferences were based on mandatory signs.

There are some limitations in the present work. We asked those responsible for signalling in the main cities of Spain to participate, and we think that the number was high enough to obtain a good picture, although a greater number of participants would have been desirable. Another obvious limitation is that there are many factors (external and internal) when driving that could influence the preference for using one way of signalling or the other. From the experimental research reviewed for this paper, we have mentioned some important factors that have been studied but there are other potential factors in real settings. Future research could provide new and convergent evidence about drivers' inferences with traffic signs. For example, using driving simulators would help test ecological conditions. In particular, we could test a prediction derived from our view: no differences are shown, which depends on presenting our scenes in egocentric rather than allocentric view. Our approach was based on predictions from deductive theories. These theories postulate a conversion from diagrammatic premises to symbolic representations before other inference processes run. Depending on the theory, these could be mental models, propositions or probabilities. Since the same symbolic representations are expected from allocentric and egocentric presentations (right turn not-allowed…), no differences are predicted in the inference. Also, by analysing participants' eye movements, different strategies for processing could be tested. Thus, it is possible that these measures would allow capture of the processing of the denied location such as prohibitory right (right is not-allowed), which leads to looking to the left. Finally, a mental load framework could also be added to represent better the real traffic conditions. In any case, our approach was to analyse the basic inferences involved in the processing of the two signs in the experimental conditions described.

In conclusion, our results point to some recommendations to potentially facilitate and speed up drivers' interpretation of traffic signs and the inferences they make from them, which would potentially give them some extra time that could be crucial to process other important information and reduce the probability of accidents. In particular, according to our results, whenever possible, we should use mandatory signs at T-junctions and, if prohibitory signs are used, we should add mandatory signs, even though they may be redundant.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

This study was approved by the University Ethics Committee (Comité de Ética en Investigación Humana de la Universidad de Granada: 1068/CEIH/2020). In addition, the participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CV: conceptualisation, methodology, software, investigation, writing-original preparation, and writing – review and editing. SM-R: funding acquisition, conceptualisation, methodology, software, investigation, writing-original preparation, and writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Spanish Government, Ministry of Economy and Competitiveness (PGC2018-095868-B-I00).

ACKNOWLEDGMENTS

We would like to thank Juan Carlos Camino, Ramón Mingueza, and Javier Roca for their suggestions to improve the questionnaire. We would also like to thank Antonio Pérez Peña for valuable information used to compile the questionnaire and all of the participants for their dedication and participation.

REFERENCES

Bucciarelli, M., and Johnson-Laird, P. N. (2005). Naïve deontics: a theory of meaning, representation, and reasoning. *Cogn. Psychol.* 50, 159–193. doi: 10.1016/j.cogpsych.2004.08.001

Castro, C., Moreno-Ríos, S., Tornay, F., and Vargas, C. (2008). Mental representations of obligatory and prohibitory traffic signs. *Acta Psychol.* 129, 8–17. doi: 10.1016/j.actpsy.2008.03.016

Dillman, D. A., Smyth, J. D., and Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. New Jersey, NJ: John Wiley & Sons, Inc.

Espino, O., and Byrne, R. M. J. (2018). Thinking about the opposite of what is said: counterfactual conditionals and symbolic or alternate simulations of negation. *Cogn. Sci.* 42, 2459–2501. doi: 10.1111/cogs.12677

European Statistical Office (2021). The Eurostat Dissemination Database [Data Set]. European Commission. Available at: https://ec.europa.eu/eurostat/en/ (Accessed April 7, 2021).

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.* 59, 255–278. doi: 10.1146/annurev.psych.59.103006.093629

Harms, I. M., Dijksterhuis, C., Jelijs, B., de Waard, D., and Brookhuis, K. A. (2019). Don't shoot the messenger: traffic-irrelevant messages on variable message signs (VMSs) might not interfere with traffic management. *Transp. Res. F: Traffic Psychol. Behav.* 65, 564–575. doi: 10.1016/j.trf.2018.09.011

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*. Cambridge: Cambridge University Press.

Johnson-Laird, P. N., and Byrne, R. M. J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychol. Rev.* 109, 646–678. doi: 10.1037/0033-295X.109.4.646

Johnson-Laird, P. N., and Ragni, M. (2019). Possibilities as the foundation of reasoning. *Cognition* 193:103950. doi: 10.1016/j.cognition.2019.04.019

Kahneman, D. (2011). *Thinking, Fast and Slow*. London: Macmillan.

Khemlani, S. S., Byrne, R. M. J., and Johnson-Laird, P. N. (2018). Facts and possibilities: a model-based theory of sentential reasoning. *Cogn. Sci.* 42, 1887–1924. doi: 10.1111/cogs.12634

Kwee-Meier, S. T., Mertens, A., and Jeschke, S. (2019). Recommendations for the design of digital escape route signage from an age-differentiated experimental study. *Fire Saf. J.* 110:102888. doi: 10.1016/j.firesaf.2019.102888

Moreno-Ríos, S., and Byrne, R. M. J. (2018). Inferences from disclosures about the truth and falsity of expert testimony. *Think. Reason.* 24, 41–78. doi: 10.1080/13546783.2017.1378724

Pathivada, B. K., and Perumal, V. (2019). Analyzing dilemma driver behavior at signalized intersection under mixed traffic conditions. *Transport. Res. F-Traf.* 60, 111–120. doi: 10.1016/j.trf.2018.10.010

R Core Team (2020). R: A Language and Environment for Statistical Computing (Version R-3.6.3) [Computer software]. Available at: https://www.R-project.org/ (Accessed June 1, 2020).

Roca, J., Castro, C., Bueno, M., and Moreno-Ríos, S. (2012). A driving-emulation task to study the integration of goals with obligatory and prohibitory traffic signs. *Appl. Ergon.* 43, 81–88. doi: 10.1016/j.apergo.2011.03.010

Schneider, W. (2003). Psychology Software Tools, Inc. (Computer software). Available at: http://www.pstnet.com/

Shinar, D., and Vogelzang, M. (2013). Comprehension of traffic signs with symbolic versus text displays. *Transport. Res. F-Traf.* 18, 72–82. doi: 10.1016/j.trf.2012.12.012

Stanovich, K. E. (1999). *Who Is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Erlbaum.

Thomas, F. M. F., and Charlton, S. G. (2020). Inattentional blindness and information relevance of variable message signs. *Accid. Anal. Prev.* 140:105511. doi: 10.1016/j.aap.2020.105511

Vargas, C., Moreno-Rios, S., Castro, C., and Underwood, G. (2011). Encoding time and signs exposure time in the representation of diagrammatic deontic meanings. *Acta Psychol.* 137, 106–114. doi: 10.1016/j.actpsy.2011.03.006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Psychology

ORIGINAL RESEARCH
published: 27 October 2021
doi: 10.3389/fpsyg.2021.681337

The Processing of the Second Syllable in Recognizing Chinese Disyllabic Spoken Words: Evidence From Eye Tracking

Youxi Wang[1,2,3], Xuelian Zang[1,2,3], Hua Zhang[2] and Wei Shen[1,2,3]*

[1] Center for Cognition and Brain Disorders, The Affiliated Hospital of Hangzhou Normal University, Hangzhou, China,
[2] Institute of Psychological Science, Hangzhou Normal University, Hangzhou, China, [3] Zhejiang Key Laboratory for Research in Assessment of Cognitive Impairments, Hangzhou, China

In the current study, two experiments were conducted to investigate the processing of the second syllable (which was considered as the rhyme at the word level) during Chinese disyllabic spoken word recognition using a printed-word paradigm. In Experiment 1, participants heard a spoken target word and were simultaneously presented with a visual display of four printed words: a target word, a phonological competitor, and two unrelated distractors. The phonological competitors were manipulated to share either full phonemic overlap of the second syllable with targets (the syllabic overlap condition; e.g., *小篆*, *xiao3zhuan4*, "calligraphy" vs. *公转*, *gong1zhuan4*, "revolution") or the initial phonemic overlap of the second syllable (the sub-syllabic overlap condition; e.g., *圆柱*, *yuan2zhu4*, "cylinder" vs. *公转*, *gong1zhuan4*, "revolution") with targets. Participants were asked to select the target words and their eye movements were simultaneously recorded. The results did not show any phonological competition effect in either the syllabic overlap condition or the sub-syllabic overlap condition. In Experiment 2, to maximize the likelihood of observing the phonological competition effect, a target-absent version of the printed-word paradigm was adopted, in which target words were removed from the visual display. The results of Experiment 2 showed significant phonological competition effects in both conditions, i.e., more fixations were made to the phonological competitors than to the distractors. Moreover, the phonological competition effect was found to be larger in the syllabic overlap condition than in the sub-syllabic overlap condition. These findings shed light on the effect of the second syllable competition at the word level during spoken word recognition and, more importantly, showed that the initial phonemes of the second syllable at the syllabic level are also accessed during Chinese disyllabic spoken word recognition.

Keywords: second syllable, Chinese, spoken word recognition, eye tracking, printed-word paradigm

OPEN ACCESS

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Jie Wang,
The Education University
of Hong Kong, Hong Kong SAR,
China
Li-Chih Wang,
National Tsing Hua University, Taiwan

****Correspondence:***
Wei Shen
shen_wei@yahoo.com

Specialty section:
This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: *16 March 2021*
Accepted: *07 October 2021*
Published: *27 October 2021*

Citation:
Wang Y, Zang X, Zhang H and Shen W (2021) The Processing of the Second Syllable in Recognizing Chinese Disyllabic Spoken Words: Evidence From Eye Tracking. Front. Psychol. 12:681337. doi: 10.3389/fpsyg.2021.681337

INTRODUCTION

Humans can understand speech quickly and almost without effort. Successful speech comprehension requires the segmentation of the continuous speech stream into discrete spoken words, and the mapping of spoken words onto the corresponding lexical representations in the mental lexicon. After decades of research, current spoken word recognition models have reached a common consent that a set of phonologically similar word candidates compete for activation as the speech signal unfolds. With the incremental availability of more disambiguating phonemic information, candidates that no longer match the speech signal are inhibited while the target word is activated until it finally wins the competition. However, current spoken word recognition models debate the size of the competitor set and the time point at which word competition occurs. For example, the cohort model (Marslen-Wilson and Tyler, 1980) posits that only competitors sharing similar word-initial phonemes with targets (e.g., "cloud" vs. "clothes") are activated for competition in the earlier phrase of speech perception. Other types of phonologically similar words such as rhymes (i.e., words that share ending phonemes such as "cloud" vs. "proud") are not involved in the competition process. By contrast, the Neighbor Activation Model (NAM; Luce, 1986; Luce and Pisoni, 1998) and the TRACE model (McClelland and Elman, 1986) allow for competition among a much broader set of candidates such as rhymes and other phonological neighbors. The NAM proposes that word candidates that differ in no more than one phoneme are neighbors competing for recognition. But the hypothesis regarding the speech temporary is not considered in the NAM model. The TRACE model assumes that word recognition is a continuous mapping process in which competitors are activated continuously. The TRACE model makes an explicit assumption that cohorts and rhymes both participate and compete in the process of spoken word recognition. Moreover, TRACE also predicts that the cohort competitors are activated strongly and earlier while rhyme competitors are activated weakly and later due to the temporal properties of speech information.

Although a growing number of studies have provided compelling evidence for the existence of cohort competition during spoken word recognition (Marslen-Wilson and Zwitserlood, 1989; Allopenna et al., 1998; Desroches et al., 2006, 2009; Simmons and Magnuson, 2018), the rhyme competition effect has been found to be weak and less reliably detected than the cohort effect in alphabetical languages like English (Allopenna et al., 1998; Simmons and Magnuson, 2018; Hendrickson et al., 2020) and elusive in non-alphabetical languages such as Chinese (Liu et al., 2006; Malins and Joanisse, 2010; Zhao et al., 2011). The aim of the current study is to investigate the role of rhyme (i.e., second syllable) in Chinese spoken disyllabic word recognition. Specifically, we aimed to examine whether rhyme competition is actually involved, and more importantly, whether the initial phonemes of the second syllable, is activated during Chinese disyllabic spoken word recognition.

Existing evidence from eye-tracking studies has demonstrated that rhymes compete in spoken word recognition but the rhyme competition effect is weaker and less stable than the cohort competition effect in alphabetical languages (Allopenna et al., 1998; Desroches et al., 2006, 2009; Brouwer and Bradlow, 2016; Simmons and Magnuson, 2018; Hendrickson et al., 2020). In the seminal visual-world paradigm study conducted by Allopenna et al. (1998), participants were presented with a visual display of four objects: a target (e.g., a beaker), a cohort competitor (e.g., a beetle), a rhyme competitor (e.g., a speaker) and an unrelated distractor (e.g., a carriage). Participants followed a spoken instruction (e.g., "please pick up the beaker") as eye movements were recorded. Results of the fixation probability showed more fixations on the cohort and rhyme competitors than on the distractors. More importantly, the cohort competition effect was found to be stronger and occurred earlier than the rhyme competition effect. Similarly, another eye-tracking study generalized the cohort and rhyme competition effect from adults to normally developing children (Desroches et al., 2006). However, the comparison of the normal group and a dyslexic group of children showed that both groups exhibited a significant cohort competition effect (i.e., cohort competitors attracted more fixations than the baseline condition), but only the normal group directed more visual attention to the rhyme competitors (e.g., sandal) than to the unrelated distractor when recognizing target words (e.g., candle), while such a rhyme competition effect was absent in the dyslexic children. This finding suggests that the rhyme competition effect is less likely to be observed than the cohort effect and is more vulnerable to other factors such as the phonological deficits of the dyslexic children. Further evidence for the rhyme competition effect was provided by Event-Related Potential (ERPs) studies. For example, Desroches et al. (2009) used ERPs to examine the time course of phonological competition in English spoken word recognition. In the picture-word matching task, participants were required to judge whether a spoken word matched a visual picture. They found that both cohort competitors (e.g., CONE vs. comb) and rhyme competitors (e.g., CONE vs. bone) elicited significantly larger N400 effects than did the complete match condition (e.g., CONE vs. cone). In addition, results from the *post hoc* analysis showed that the N400 effect in the rhyme condition was much weaker than that in the cohort condition. Together, the above findings are in line with the assumption of the TRACE model that rhymes compete later and much more weakly than do cohorts.

While the above-mentioned studies proposed that rhymes do compete for recognition in alphabetical languages, the role of rhymes in Chinese spoken word perception is rather elusive, given that some studies found rhyme competition effect while other did not. For example, in an eye-tracking study, Meng (2014) investigated rhyme competition in disyllabic Chinese words by manipulating rhyme competitors sharing entire phonemes of the second syllable in an eye-tracking study. Participants were presented with spoken target words (e.g., *桥洞*, *qiao2dong4*, "arches") in spoken sentences and were simultaneously presented with a visual display of four words: a cohort competitor (e.g., *樵夫*, *qiao2fu1*, "woodman"), a rhyme competitor (e.g., *果冻*, *guo3dong4*, "jelly"), a tonal competitor (e.g., *麻袋*, *ma2dai4*, "gunny-bag") and an unrelated distractor. The results showed longer total viewing time on the cohort competitors than on distractors, suggesting the activation of cohort competitors. However, they did not find any interference effect induced by rhyme competitors and the total viewing times for rhymes were even shorter than those for the distractors. On the other hand, an ERP study conducted on disyllabic words by Liu et al. (2006) found evidence for rhyme activation. Liu et al. (2006) found that rhyme competitors (e.g., *水池*, *shui3chi2,* "water pool"; target words: *电池*, *dian4chi2,* "battery") elicited earlier N400 amplitudes than did cohort competitors (e.g., *电路*, *dian4lu2,* "electric stove"; target words: *电池*, *dian4chi2,* "battery"). Moreover, the survival time of the N400 amplitude was much shorter for the rhyme competitor than for the unrelated condition. This indicates that the N400 effect elicited by the rhyme competitor disappeared earlier than that in the unrelated condition. These findings provide evidence for the activation of both cohort and rhyme information in Chinese, and the rhyme effect occurred later in the processing time course than did the cohort effect.

To summarize, evidence from previous studies for rhyme processing in Chinese remains elusive. We assumed that the inconsistent findings between Liu et al. (2006) and Meng (2014) could be due to that the degree of sentence constraints differs greatly across the two studies. Liu et al. (2006) examined rhyme processing in the context of highly predictive sentences (the mean predictability score was 6.33 on a 7-point scale), while the spoken sentences used in Meng (2014) were neutral ones. The predictive context contains highly constraining contextual cues for the upcoming words while the neutral context or the isolated word situation cannot provide any valid contextual cues. Studies have shown that phonological processing of words can be affected by the sentence predictability. For example, the lexical tone exhibits stronger constraining role during Chinese spoken word recognition in the predictive context compared to the neutral context or isolated word recognition; and a similar modulating role of lexical tone was found between the neutral context or isolated word recognition (Liu and Samuel, 2007; Shen et al., 2020). The mechanism underlying the prediction is assumed to be the pre-activation, that is, the activation occurs earlier before the presentation (Otten et al., 2007; Laszlo and Federmeier, 2009). It was found that both phonological/form and semantic information of words can be pre-activated in highly constraining contexts (Ito et al., 2016). Therefore, it is possible that the rhyme information was pre-activated in highly predictive sentences, which would result in a larger and earlier rhyme competition effect compared to the neutral context or isolated word recognition.

However, the neutral context or the isolation word recognition cannot fully account for the absence of the rhyme competition effect in Meng's study given that the role of rhyme is consistently observed in the isolated word recognition in alphabetical languages (Allopenna et al., 1998; Desroches et al., 2006, 2009). We speculated that the absence of rhyme competition effect could be due to the inappropriate experimental design and the data analysis in Meng's study. First, presenting more than one competitor in a same visual display (see the experimental manipulation in Meng, 2014) could lead to cognitive competition among different competitors and thus weaken the observed the rhyme competition effects. It has been shown that the competitor effect to be reduced and occurred later with the increasing number of visual referents (Sorensen and Bailey, 2007), which was explained by the working memory capacity limitation. The representations of visual referents would become faded in the working memory in the larger visual arrays compared to the small visual arrays (Sorensen and Bailey, 2007; Huettig et al., 2011). Second, the semantic plausibility between the spoken sentences and the rhyme competitors was not controlled. For example, the rhyme competitors (e.g., *果冻*, *guo3dong4*, "jelly") were implausible for substitution into sentences like "*从这里穿出_*" ("Walking out from _"). A prior study found that the visual objects which were semantically plausible extension of the preceding context attracted more visual attention than the objects that were semantically implausible (Frassinelli and Keller, 2012). This is because the words in sentences can generate semantic properties concerning with the concept and such information can be used in allocating visual attention to the well-matched visual objects (Frassinelli and Keller, 2012). Therefore, the uncontrolled semantic plausibility could have resulted in fewer fixations on the rhyme competitors in Meng's study. Third, no time course analysis of the rhyme processing was conducted in Meng (2014) and the indexes of fixation durations (*first fixation duration* and *total fixation duration*) and the *number of fixations* cannot necessarily reflect the rhyme competition occurred in the specific processing stage during spoken perception (e.g., a relatively later rhyme processing found in alphabetical languages; Allopenna et al., 1998; Desroches et al., 2006).

Based on the above debates, the current study aimed to address two research questions. First of all, we aimed to clarify whether a rhyme (i.e., the second syllable[1]) is activated for competition in disyllabic Chinese words using a visual world paradigm with eye tracking. Noted that Chinese syllables differ from alphabetical languages in the following aspects. First, Chinese syllable can refer to individual morphemes and can be used as mono-syllabic words (Zhou and Marslen-Wilson, 1994). This unique property enables that the rhymes (i.e., the second syllable in the disyllabic words) in Chinese are more salient than alphabetical languages like English. Second, Chinese syllables are composed of segmental information (i.e., the phonemes) and suprasegmental information (i.e., lexical tone). Thus, the syllable processing is more complicated than in alphabetical languages. Current spoken word recognition models (TRACE and NAM models) are proposed on the basis of the alphabetical languages and it remains less clear that the extent to which the rhyme information across languages is processed differently or similarly. Given the unique property and structure of the Chinese syllables, we assumed that the role of rhyme information in Chinese may be different from that in the alphabetical languages. In

---

[1] In previous alphabetical languages studies, rhymes were defined as sharing similar ending phonemes in the final syllable between two or more words (e.g., *"beaker"* – *"speaker"* in Allopenna et al., 1998). In Chinese disyllabic words studies, rhymes were defined by the second syllable (e.g., *桥洞, qiao2dong4,* "arches," – *果冻, guo3dong4,* "jelly" in Meng, 2014; *电池, dian4chi2,* "battery"- *水池, shui3chi2,* "water pool" in Liu et al., 2006). Similarly, in the current study, we also considered the second syllable to be the shared rhyme between the target and competitor words.

addition, our current study focused on word but not character processing is because disyllabic word plays a predominant role in Chinese (73% of words in Chinese are disyllabic words; Zhou and Marslen-Wilson, 1994) which is practically and theoretically more significant to examine the rhyme processing. In addition, the target words used in prior alphabetical studies are usually disyllabic words (Allopenna et al., 1998; Desroches et al., 2006; Simmons and Magnuson, 2018); thus, examining rhyme processing in the disyllabic Chinese words allows us to explore the similarities and differences of rhyme processing across between Chinese and English.

The second question of the current study was to examine whether an embedded cohort of the rhyme, that is initial phonemes of the second syllable in Chinese words, is activated for competition during spoken word recognition. Note that more than 73% of the Chinese words are disyllabic and each syllable in the disyllabic words can be used independently as a free monosyllabic word (Zhou and Marslen-Wilson, 1994). Therefore, the second syllable of the disyllabic words is not only the rhyme at the whole word level (Liu et al., 2006; Meng, 2014), but also contains its own initial and end parts at the monosyllabic level. For example, in the disyllabic word *"gong1zhuan4"* (*公转*, "revolution"), the second syllable "*zhuan4*" is the rhyme which contains monosyllabic-initial part "zhu-" and monosyllabic-end part "-uan." The TRACE model assumes that cohort and rhyme within a word are accessed in temporal order and compete during spoken word recognition, but it makes no assumptions about whether there is also a further competition from the monosyllabic-initial/end. Chinese syllables are ideal material to examine this issue.

Regarding the processing of monosyllabic word, previous studies showed that the cohort part plays a more important role than the rhyme part. For example, in an eye-tracking study, Malins and Joanisse (2010) presented participants with a spoken Chinese monosyllabic word (e.g., *床,*[2] *chuang2*, "bed") and a visual display of four objects simultaneously: a target (e.g., *chuang2*, "bed"), a cohort competitor (e.g., *船, chuan2*, "ship")/a rhyme competitor (e.g., *黄, huang2*, "yellow") and two phonologically unrelated distractors. Participants were instructed to select the corresponding referent of the target in the visual display. The results showed a significant cohort competition effect with cohort competitor attracted more fixations than the distractors. In comparison, comparable fixations were found between the rhyme competitor and distractors, suggesting that rhyme information was not involved (as compared to cohort) in lexical competition during spoken word recognition. A similar pattern of stronger and sustained role of cohort competitor compared to the rhyme competitor was also observed in ERP studies (e.g., Malins and Joanisse, 2012). Given that the cohorts have more weight than the rhymes during spoken word processing, in this study, we investigated the issue that whether the embedded cohort of the rhyme (i.e., second syllable) could also be activated during spoken perception in Chinese.

In the following experiments, we adopted a printed-word paradigm with eye tracking to investigate the processing of the second syllable processing in Chinese spoken word perception (Huettig and McQueen, 2007; McQueen and Viebahn, 2007). The visual world paradigm has been widely used to explore phonological processing in visual word recognition (Shatzman and McQueen, 2006; Huettig and McQueen, 2007; McQueen and Viebahn, 2007; Weber et al., 2007; Ito et al., 2018; Shen et al., 2018). Compared to the ERP technique, the visual world paradigm with eye tracking has several advantages: (1) In ERPs studies, some explicit responses such as eye blinks and moving eyes would cause a great deal of electrical noise on the EEG signals (Rayner and Clifton, 2009), while the eye movements recording can occur implicitly without the interferences from explicit responses (Huettig and McQueen, 2007); (2) In ERP studies, incongruent spoken sentences are usually constructed when investigating the spoken comprehension (e.g., Liu et al., 2006), the eye movements recording can be recorded in a more natural language comprehension environment with the normal spoken sentence as stimuli; thus has higher ecological validity than the ERP technique. The printed-word paradigm adopted in the current study is a variation on the visual world paradigm in which the printed words replace visual pictures (McQueen and Viebahn, 2007). In addition, to avoid potential confounding that may harm the observation of rhyme competition effect, in the current study we adopted the following manipulations: (a) Only one type of phonological competitors (i.e., the rhyme competitor) was presented in each visual display. This could reduce the potential interference from any other competition effects. (b) Target words were presented in isolation (without sentence context) to avoid any possible prediction or semantic plausibility effect from sentence context; (c) A detailed time course analysis was conducted to tap into the rhyme competition in Chinese.

In the current study, participants viewed a display of printed words with simultaneous verbal presentation of target words. For a given target word, the competitors either shared the full phonemes of the second syllable with the targets (hereafter called "the syllabic overlap condition"), or shared partial (i.e., initial) phonemes of the second syllable (hereafter called "the sub-syllabic overlap condition") with targets. The cohort model posits that rhyme information is not accessed during word recognition, and thus no competition effect should be observed in the syllabic overlap condition. On the other hand, both TRACE and NAM predict a significant phonological competition effect in the syllabic overlap condition, that was, a rhyme competition effect should be observed. In addition, based on the prior findings according to monosyllabic words that cohorts have more weight than the rhymes during spoken word recognition (Malins and Joanisse, 2010, 2012), we hypothesized that if the second syllable (i.e., rhyme) can be activated, the initial phonemes of the second syllable would also be activated to some extent.

Furthermore, the manipulation of two conditions also allows us to test the hypothesis of "phonological similarity." According to the TRACE model, "global similarity" plays an important role in mapping spoken words onto lexical representations (McClelland and Elman, 1986). This assumption predicts that the degree of word activation varies with phonological similarity.

---

[2] Chinese characters of the examples in Malins and Joanisse (2010) were not given in the original article. We added those Chinese characters to ease the understanding for readers.

Using the eye-tracking technique, Shen et al. (2018) manipulated the phonological similarity of the first syllable in Chinese disyllabic words and found that more fixations were allocated to high-similarity competitors (sharing full phonemic overlap with targets) compared to low-similarity competitors (sharing partial phonemic overlap with targets). If the mapping rule of "global similarity" also applies to second syllable processing in Chinese, we would observe a larger phonological competition effect in the syllabic overlap condition while a small effect in the sub-syllabic overlap condition.

EXPERIMENT 1

Method

Participants

Forty undergraduates (13 men, 27 women) from Hangzhou Normal University were randomly recruited and participated in the experiment. Their ages ranged from 18 to 27 years (*Mean age* = 21.2 years) and they were all native Mandarin Chinese speakers who had normal or corrected-to-normal vision and normal hearing. A monetary compensation was paid to each participant after the experiment. The research protocol reported here was approved by the ethics committee of the Institute of Psychological Sciences from Hangzhou Normal University.

Materials and Design

All spoken target words were recorded by a native Chinese female speaker at a normal speaking speed on the software Praat at a sampling rate of 44.1 kHz. All spoken target words were embedded in a spoken instructional carrier, "*请点击*" ("please click on"), and presented to participants through headphones.

Fifty-four Chinese disyllabic words were selected as target items. Each visual display included a target word, a phonological competitor, and two distractors that were neither semantically nor phonologically related with target word. For each target word, there were two types of corresponding phonological competitors: a syllabic overlap competitor and a sub-syllabic overlap competitor. For the syllabic overlap condition, the phonological competitor shared all phonemes of the second syllable with the target words (e.g., *"小篆," xiao3zhuan4*, "calligraphy" vs. *"公转," gong1zhuan4*, "revolution"). For the sub-syllabic overlap condition, the competitor shared cohort-part phonemes of the second syllable with targets (e.g., *"圆柱," yuan2zhu4*, "cylinder" vs. *"公转," gong1zhuan4*, "revolution"). To maximize the possibility of observing the phonological competition effect, all phonological competitors were matched in the lexical tone of the second syllable with the targets. (See **Figure 1** for sample stimuli of Experiment 1.) Word frequency and number of strokes were carefully matched across the four printed words in the two conditions (syllabic overlap condition: $Fs < 1$, $ps > 0.60$; sub-syllabic overlap condition: $Fs < 1$, $ps > 0.82$; see **Table 1** for the lexical properties of experimental stimuli; word frequency data from the Chinese Linguistic Data Consortium, 2003). All phonological competitors were carefully selected to share no semantic association or orthographic association with target words. Another twenty participants (who did not participate in the following eye-tracking experiment) were recruited and were asked to rate the semantic relatedness between targets and competitors on a 5-point scale. Results showed no significant difference between the two types of competitors (targets and syllabic overlap competitors: *mean score* = 1.15, *SD* = 0.25; targets and sub-syllabic overlap competitors: *mean score* = 1.13, *SD* = 0.21; $t = 0.45$, $p = 0.65$). In addition, two research assistants examined the material to ensure there was no orthographic association between targets and competitors (i.e., sharing no radicals). All critical stimuli were split into two lists. Each list contained 27 syllabic overlap items and 27 sub-syllabic overlap items with no repetition of target words. Each participant was randomly assigned to perform one list only. See the **Appendix** for all materials used in critical trials.

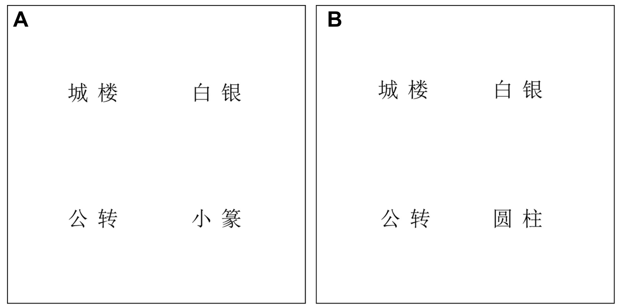


FIGURE 1 | An example of a printed-word display in Experiment 1. For the spoken target word *公转*, *gong1zhuan4*, "revolution," the printed-word display consisted of an identical target word, a phonological competitor word in the syllabic overlap condition **(A)** *小篆*, *xiao3zhuan4*, "calligraphy" or in the sub-syllabic overlap condition **(B)** *圆柱*, *yuan2zhu4*, "cylinder" and two unrelated distractors *城楼*, *cheng2lou2*, "gate tower" and *白银*, *bai2yin2*, "silver" in four corners of the display.

Apparatus

Participants' eye movements during the experiment were recorded using an EyeLink1000 Desktop tracker (SR Research, Mississauga, ON, Canada), with sampling at a rate of 1000 Hz. The experimental task was programmed using Experimental Builder software (SR Research Ltd). Auditory stimuli were presented to the participant via headphones (Sennheiser, PC 230). Visual stimuli were presented on a 21-inch monitor (resolution: 1024 × 768; refresh rate: 85 Hz) of a Dell computer. The visual stimuli were displayed in black (RGB: 0, 0, 0) against a white background (RGB: 255, 255, 255). Participants were seated about 57 cm away from the display screen and 1 cm on the screen subtended a visual angle of approximately 1°. A chin rest was used to control the participant's head position. Although the viewing was binocular, only eye movements of the right eye were recorded during the entire procedure.

Procedure

Before the experiment, each participant was given a brief introduction to the experiment. The eye-tracker was then calibrated and validated via 9-point calibration prior to the beginning of the experiment. A drift check was performed before the start of each trial. Each trial started with a blank screen

TABLE 1 | Lexical properties of the experimental materials in Experiment 1 and 2.

	Target word	Phonological competitor (the syllabic overlap condition)	Phonological competitor (the sub-syllabic overlap condition)	Distractor	Distractor
Mean word frequency	2.39 (2.53)	2.32 (2.61)	2.74 (3.70)	2.20 (2.04)	2.28 (2.11)
Mean number of strokes	16.81 (4.60)	16.94 (4.76)	17.37 (4.19)	17.37 (4.46)	17.81 (4.75)

Mean word frequency was calculated as occurrences per million. The standard deviants are presented in brackets.

displayed for 500 ms, and then followed by a visual display consisting of four printed words. The auditory stimuli were presented 200 ms[3] after the visual display onset. The four printed words were arranged at the four corners of the screen and the positions of each printed word were randomized across trials[4]. The size of each disyllabic word was approximately 1.5° × 3° and located about 10° away from the screen center. Participants were asked to find the target word (that was same as the auditory stimulus) and click on it using a computer mouse. The visual display remained on the screen until a response was made. Participants were required to make the responses as quickly and accurately as possible.

There were seven practice trials prior to the formal experiment to ensure that participants were familiar with the task procedure. Additional 54 filler trials including a target and three distractors were added into the experiment, which were constructed to avoid participants being aware of the manipulations in the critical trials. Critical trials and filler trials were intermixed and randomly presented. The duration of the whole experiment was approximately 10–15 min.

Results and Discussion

Accuracy Data

Filler trials, practice trials, and critical trials with wrong responses were all excluded from the data analysis. Participants' accuracy in the critical trials was 99.9%, suggesting that they paid sufficient attention to the task.

Data Coding

A square region of 10° × 10°centered around each printed word was designated as the region of interest (ROI) for eye tracking. Only fixations that fell into the ROI were defined as "fixating on the current word" and those fixations that did not fall on the printed words were considered as falling on the background. All fixations were coded as "0" (not fixed) or "1" (fixed) for every 100 ms bin starting from 200 ms before the onset of the auditory stimulus presentation.

Eye Movement Data

A logit mixed model (Jaeger, 2008; Quené and van den Bergh, 2008; Ferreira et al., 2013) was employed to analyze the eye movement data in the R software (R Core Team, 2020). A glmer() function in the lme4 package (Version 1.1-23; Bate et al., 2015) was used to build the mixed effect models. For the base model, a random intercept for participants and items was added. Then, the fixed effects (*"the word type"*: competitor vs. distractor and *"the competitor type"*: the syllabic overlap competitor vs. the sub-syllabic overlap competitor) were added into the models one by one, followed by the interaction of the fixed effects as well as the by-participants random slope for the fixed factor (Cunnings, 2012; Barr et al., 2013). To test whether adding a factor improved the model fit, anova() was used to conduct the model comparison. For any significant interaction effect, contrast analysis was performed to compare the effect between the competitor and the distractors.

Figure 2 presents the proportion of the fixations to the targets, the phonological competitors, and the distractors[5] in the syllabic overlap and sub-syllabic overlap conditions, respectively. As seen in **Figure 2**, the fixation proportion curve of the target showed a significant separation trend compared to that of the phonological competitors and distractors after 300 ms from the onset of the spoken words. However, in both conditions, the fixation proportion curves of the phonological competitors and the distractors were almost overlapping except for a very tiny difference in the time window bin of 400–500 ms vs. 500–600 ms. The data for the two time windows were thus analyzed to test whether the difference reached statistical significance. Results of the logit mixed model showed that adding fixed factors or interaction did not improve model fit for either the time window of 400–500 ms [χ^2 (2) = 0.24, p = 0.89] or 500–600 ms [χ^2 (2) = 0.89, p = 0.64].

In Experiment 1, no evidence was found for a phonological competition effect on the second syllable (i.e., the rhyme), in agreement with the findings of Malins and Joanisse (2010). This result stands in contrast to the cohort competition effect found in Chinese using a very similar design (Shen et al., 2018). We assume that rhyme processing may be relatively weak compared to cohort processing in Chinese. Therefore, a competitive situation (e.g., presenting a target word in the visual display) may make the subtle rhyme effect less observable. Some previous studies have already shown that the semantic competition effects were larger when the target was not presented in a visual display (i.e., a target-absent design) compared to when the target was presented (i.e., a target-present design) (Huettig and Altmann, 2005). This is because the presence of the target will attract the most attentional resources in the visual field, thus leading to less attention being directed to other visual referents. For this reason, the use of the target-present design may have reduced the likelihood of observing the rhyme competition effect in

---

[3] This choice of preview time prevents participants from reading the words strategically.

[4] For each spoken target word, the positions of printed targets, competitors, and distractors were the same in the syllabic overlap and sub-syllabic overlap conditions.

[5] The curve of distractors refers to the average fixation proportions of the two distractors in **Figures 2**, **4**.

Experiment 1. Therefore, in Experiment 2, we adopted a target-absent display of the visual-world paradigm to further test the role of the second syllable (i.e., the rhyme) competition effect during Chinese spoken word recognition.

EXPERIMENT 2

Method

Participants

Forty undergraduates (15 men, 25 women) from the same participant pool were randomly recruited to participate in Experiment 2; none of them had participated in Experiment 1. Their ages ranged from 17 to 23 years (*Mean age* = 19.95 years) and they were all native Mandarin Chinese speakers who had normal or corrected-to-normal vision and normal hearing. A small compensation was paid to each participant after the experiment. The research protocol reported here was approved by the university ethics committee.

Materials and Apparatus

Materials and apparatus were the same as those used in Experiment 1.

Procedure

The procedure was modified from Experiment 1, in that the visual referents of the spoken target words for the critical trials were not presented on the display screen. Additionally, the spoken target words were presented to participants without the preceding carrier phrase. For each trial, three printed words were presented, arranged as a V-shape or an inverted V-shape on the screen. The positions of the printed words were randomized across trials[6] (Tsang and Chen, 2010). Each word was located about 8° away from the screen center (see **Figure 3** for experimental stimuli of Experiment 2). Participants were presented with spoken target words and instructed to determine whether or not the referent of the spoken target word was on the screen by pressing corresponding buttons on a keyboard. An equivalent number of filler trials were constructed to balance the responses. The referents of spoken targets were presented in the visual display only for filler trials. Thus, "YES" responses were expected in the filler trials, and "NO" responses were expected in the critical trials.

Results and Discussion

Accuracy Data

Filler trials, practice trials, and critical trials with wrong responses were all excluded from the further data analysis. Participants' accuracy in the critical trials was 99.26%, suggesting that they paid sufficient attention to the task.

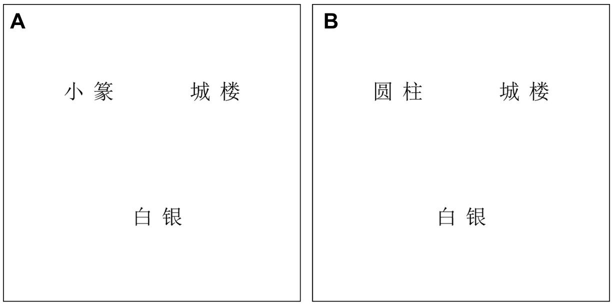


FIGURE 3 | An example of a printed-word display in Experiment 2. For the spoken target word *公转*, *gong1zhuan4*, "revolution," the printed-word display consisted of a phonological competitor word in the syllabic overlap condition **(A)** *小篆*, *xiao3zhuan4*, "calligraphy" or in the sub-syllabic overlap condition **(B)** *圆柱*, *yuan2zhu4*, "cylinder" and two unrelated distractors *城楼*, *cheng2lou2,* "gate tower" and *白银*, *bai2yin2*, "silver". In this example, the three printed words were arranged as a V-shape.

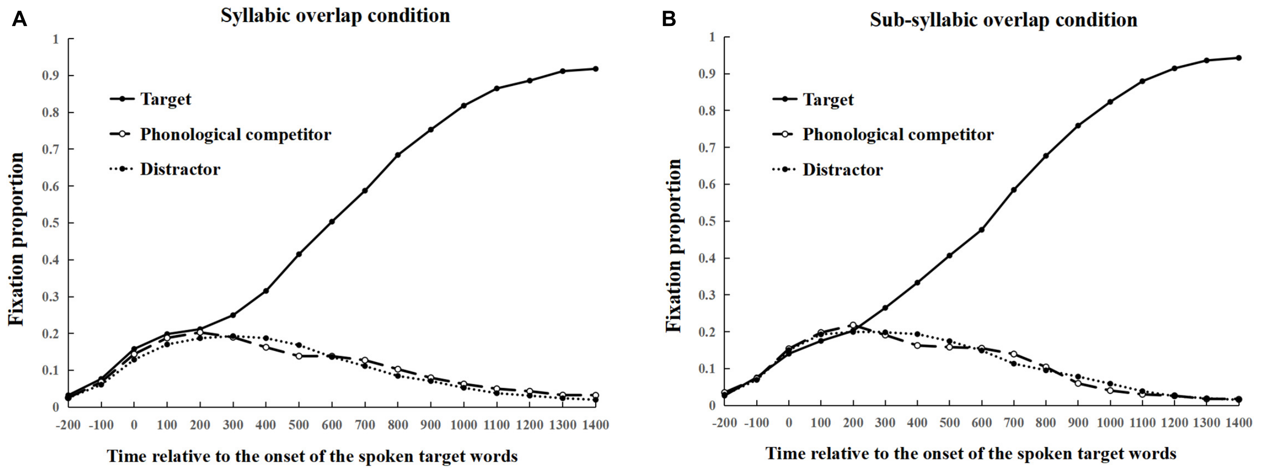


FIGURE 2 | Proportion of fixations on the target, the phonological competitor and the distractors from 200 ms before the onset of the spoken target word in the syllabic overlap condition **(A)**, and sub-syllabic overlap condition **(B)**, respectively in Experiment 1.

---

[6] For each spoken target word, the positions of competitors and distractors were the same in the syllabic overlap and sub-syllabic overlap conditions.

Eye Movement Data

The data coding and statistical analysis were the same as that in Experiment 1. **Figure 4** presents the proportion of the fixations to the phonological competitors and the distractors in the syllabic overlap and sub-syllabic overlap conditions, respectively. The fixation proportion curves of the phonological competitor and the distractor had a clear disassociation from 800 ms until 1300 ms after the onset of the spoken targets. Numerous fixations fell on the phonological competitors compared to the distractors in both conditions. But the degree of divergence in the syllabic overlap condition was larger than that in the sub-syllabic overlap condition. To test whether the differences approached statistical significance, the time windows from 900 to 1300 ms were analyzed using the logit mixed model (see **Table 2**). The results showed that the model was significantly improved by the interaction of fixed effects and the by-participant random slope of the *"the competitor type"* $[\chi^2\ (5) = 16.62,\ p < 0.01]$ in the time window of 900–1000 ms. The pattern of interaction effects was sustained for the time window from 1000 to 1300 ms ($\chi^2 s > 8.01$, *ps* < 0.05). For the time window of 900–1000 ms, contrast analysis showed that phonological competitors attracted more fixations compared to the distractors in both the syllabic overlap condition ($b = 0.59$, $SE = 0.09$, $Z = 6.30$, $p < 0.001$) and the sub-syllabic overlap condition ($b = 0.32$, $SE = 0.10$, $Z = 3.41$, $p < 0.001$). Moreover, fixation proportions on syllabic overlap competitors were significantly larger than those on the sub-syllabic overlap competitors ($b = -0.45$, $SE = 0.10$, $Z = -4.59$, $p < 0.001$), while no difference was found for the distractors ($Z = 0.70$, $p = 0.49$). The significant phonological competition effects suggested that the second syllable (i.e., the rhyme) was indeed competing and activated during processing of the spoken word perception. More importantly, the phonological activation of sub-syllabic overlap competitors also suggested a significant phonological competition effect of the initial phonemes of the second syllable. In addition, the interaction effects suggested that the activation of phonological competitors was sensitive to phonemic overlapping such that the activation degree of the phonological competitor was varied as a function of the full/partial phonemic overlap. Thus, more visual attention was directed to the syllabic overlap competitors than to the sub-syllabic overlap competitors.

Unexpectedly, we also found a main effect of *"the word type"* in the time window of 300–500 ms. The fixation proportions on phonological competitors were significantly higher than those on distractors in both conditions (χ^2s > 6.86, *ps* < 0.04). However, no interaction of fixed effects was found (χ^2s < 1, *ps* > 0.70). We assume that the mismatched word frequency for some items may account for this unexpected result. After a careful check of the material, we found that the word frequency of item 28 did not match perfectly across conditions such that the word frequency of the sub-syllabic overlap competitor (i.e., *"爆炸," bao4zha4*, *Word frequency* = 23.31 occurrences per million) was much higher than the word frequencies of the distractors (i.e., *"陵墓," ling2mu4*, *Word frequency* = 2.14 occurrences per million, *"座椅," zuo4yi3*, *Word frequency* = 2.19 occurrences per million). When a further data analysis excluding item 28 was conducted, results showed that the unexpected main effect in the time window of 300–500 ms disappeared ($\chi^2\ s > 3.45$, *ps* > 0.13), and our main interaction effects in the time windows of 900–1300 ms remained significant ($\chi^2\ s > 6.91$, *ps* < 0.05). In addition, the mean word frequency and number of strokes between targets and the competitors and distractors remained matched (syllabic overlap condition: *Fs* < 1, *ps* > 0.69; sub-syllabic overlap condition: *Fs* < 1, *ps* > 0.54)[7].

GENERAL DISCUSSION

Two experiments were conducted to determine whether the second syllable compete for word recognition and whether the initial-part of the second syllable is involved in the processing of

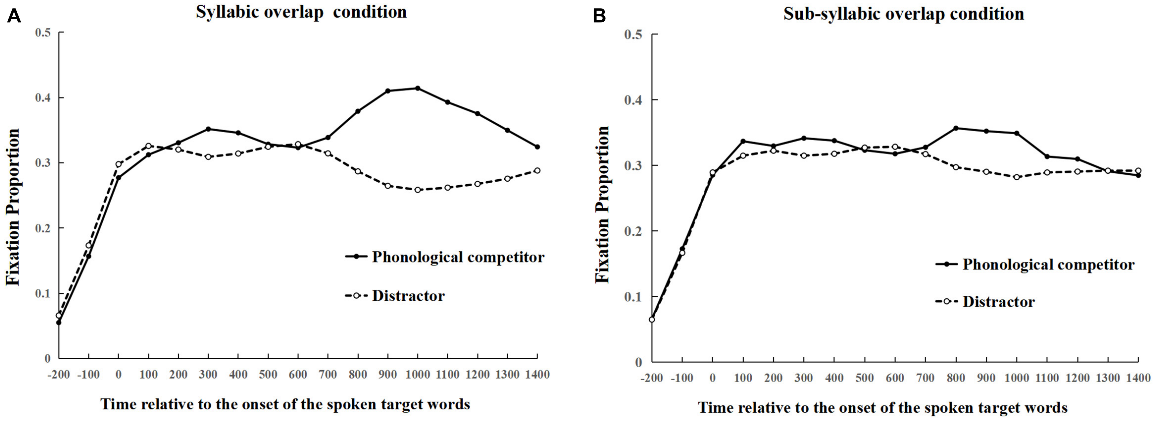


FIGURE 4 | Proportion of fixations on the phonological competitor and the distractors from 200 ms before the onset of the spoken target word in syllabic overlap condition **(A)**, and sub-syllabic overlap condition **(B)**, respectively in Experiment 2.

---

[7] Mean word frequency: syllabic overlap competitors, $M = 2.36$, $SD = 2.62$; sub-syllabic overlap competitors, $M = 2.35$, $SD = 2.38$; distractor, $M = 2.20$, $SD = 2.06$; distractor, $M = 2.28$, $SD = 2.13$; Mean number of strokes: syllabic overlap competitors, $M = 16.94$, $SD = 4.81$; sub-syllabic overlap competitors, $M = 17.17$, $SD = 3.96$; distractor, $M = 17.26$, $SD = 4.43$; distractor, $M = 17.74$, $SD = 4.76$.

TABLE 2 | Time windows analysis for results of logit mixed models in the syllabic overlap and sub-syllabic overlap conditions in Experiment 2.

Time window from spoken target word (ms)	Predictor	Syllabic overlap condition				Sub-syllabic overlap condition			
		Estimate	Standard error	*Z* values	Corrected *p*-values	Estimate	Standard error	*Z* values	Corrected *p*-values
900–1000 ms	(Intercept)	−1.01	0.07	−14.00	< 0.001***	−1.03	0.07	−13.92	< 0.001***
	Distractor	−0.07	0.10	−0.69	0.49	0.06	0.10	0.63	0.53
	Phonological competitor	0.59	0.09	6.34	< 0.001***	0.32	0.10	3.41	< 0.001***
1000–1100 ms	(Intercept)	−1.29	0.09	−13.91	< 0.001***	−1.28	0.09	−13.71	< 0.001***
	Distractor	0.05	0.10	0.47	0.64	0.10	0.10	0.93	0.35
	Phonological competitor	0.70	0.10	7.10	< 0.001***	0.33	0.10	3.17	< 0.001***
1100–1200 ms	(Intercept)	−1.57	0.11	−13.96	< 0.001***	−1.54	0.12	−13.10	< 0.001***
	Distractor	0.16	0.11	1.44	0.15	0.10	0.11	0.95	0.34
	Phonological competitor	0.62	0.11	5.88	< 0.001***	0.17	0.11	1.54	0.12
1200–1300 ms	(Intercept)	−1.71	0.14	−12.61	< 0.001***	−1.85	0.14	−13.19	< 0.001***
	Distractor	−0.19	0.12	−1.58	0.11	0.07	0.12	0.55	0.58
	Phonological competitor	0.31	0.11	2.76	< 0.01**	0.09	0.12	0.79	0.43

***p < 0.01, ***p < 0.001.*

Chinese disyllabic word recognition using a printed-word version of the visual-world paradigm. We manipulated the competitors sharing full phonemes of the second syllable with the targets (i.e., "the syllabic overlap condition") or sharing the cohort-part phonemes of the second syllable with the targets (i.e., "the sub-syllabic overlap condition"). No main effect or any interaction effect was found to be significant when the targets were presented in the visual display in Experiment 1. However, a significant phonological competition effect was observed under both conditions when the target referents were removed from the visual display in Experiment 2 (i.e., the fixation proportions on the phonological competitors were higher than those on distractors) in both syllabic overlap and sub-syllabic overlap conditions, suggesting that both second syllable (i.e., the rhyme) and the initial part of the second syllable were accessed for competition. In addition, we also found a larger competition effect in the syllabic overlap condition than that in the sub-syllabic overlap condition in Experiment 2.

In the current study, we observed clear and robust evidence that the second syllable (which was usually deemed as a rhyme in previous Chinese studies, Liu et al., 2006; Meng, 2014) are involved in competition in Chinese spoken disyllabic word recognition using a target-absent display of the visual-world paradigm. Prior studies regarding rhyme processing in Chinese have yielded contradictory results. Some studies have found no evidence for rhyme competition while other studies have observed a rhyme competition effect (Zhao et al., 2011; Meng, 2014). To revisit this issue and maximize the possibility of observing a compound rhyme competition effect on disyllabic words, we altered some aspects of our original experimental design, i.e., presenting only one competitor type and presenting targets in isolation, using the printed-word paradigm. The results showed an important role of the second syllable (i.e., the rhyme) in Chinese spoken word recognition, in that the syllabic overlap competitors attracted more fixations than did the distractors. The significant rhyme effect in Chinese is consistent with the role of rhyme for disyllabic words in alphabetical languages, showing a universal competitive role of rhymes across languages. However, the rhyme competition effect in Chinese was only observed in the target-absent situation in Experiment 2 and the effect disappeared in the target-present situation in Experiment 1. These results seem to indicate that the rhyme effect for disyllabic words in Chinese is relatively weak, such that a competitive environment (such as presenting targets on a screen) weakens the rhyme effect to the point of producing a null effect in Experiment 1.

More importantly, we also observed the activation of the initial phonemes of the second syllable (i.e., rhyme) during spoken word recognition. As noted, most studies focused on the separate role of cohort/rhyme and few studies have ever investigated the role of embedded cohort in the rhyme. The significant competition effect observed in the sub-syllabic overlap condition suggested that the embedded cohort part within a rhyme was also activated and the rhyme processing in the disyllabic word was accessed in a multiple-layer way: the whole-word layer and the monosyllabic layer. In addition, we also observed a modulating role of phonemic overlapping proportion on the activation degree of the phonological competitors; a larger competition effect was found for syllabic overlap competitors and a relatively small competition effect was found for sub-syllabic overlap competitors, demonstrating that the mapping of spoken signals onto the inner lexical representation was determined by the degree of phonological similarity. Therefore, syllabic overlap competitors with full phoneme overlapping with targets are activated to a higher degree than the sub-syllabic overlap competitors with only partial phoneme overlapping with targets. This is reflected in eye movement behaviors as more fixations are directed to the syllabic overlap competitors than to the sub-syllabic overlap competitors.

Given that the second syllable was considered as the rhyme in prior Chinese disyllabic word processing (Liu et al., 2006;

Meng, 2014), the current findings also contribute to refinements of the rhyme processing assumptions in the spoken word recognition models. Existing spoken word recognition models make different assumptions regarding rhyme processing. The second syllable (i.e., rhyme) competition effect observed in the present study is in line with the assumptions of the TRACE model (McClelland and Elman, 1986) which assumes that spoken word recognition is a continuous process with the competition among rhymes being involved. The TRACE model also predicts that rhymes and cohorts are processed with different weights, with a weaker and later activation of the rhyme processing but an earlier and stronger activation for the cohort processing. Our current finding of an absence of a rhyme competition effect in the target-present situation stands in contrast to the significant cohort effect found in the target-present situation in our prior studies (Shen et al., 2018). The combined results across two studies provide some indication that the rhyme effect is relatively weaker than the cohort effect in Chinese. In addition, both the TRACE and TTRACE model (proposed by Tong et al., 2014) make no assumptions about how the embedded cohort/rhyme are processed in Chinese spoken word recognition. Our current findings provided the first piece of evidence that the embedded cohort part of the rhyme can also be activated during Chinese disyllabic word recognition. The significant phonological competition effect observed in the sub-syllabic overlap condition in Experiment 2 also shed some lights on the uniqueness and complication of the rhyme processing in Chinese. The rhyme processing in Chinese is likely to be accessed in a multiple-layer manner: the whole word layer and the monosyllabic layer, and the rhymes are possibly to be accessed in parallel or serially at these two different levels.

It is well known that compared to alphabetical languages such as English, most syllables in Chinese can be mapped onto morphemes and can stand alone as monosyllabic words (Zhou and Marslen-Wilson, 1994; Zhao et al., 2011). This specific linguistic property may lead to different rhyme representations in Chinese (especially in disyllabic Chinese words) as compared to that in alphabetical languages such that rhyme representation in Chinese may be more salient than that in alphabetical languages. However, the saliency of rhyme representation in Chinese may be affected by word frequency. Prior studies have found that low-frequency words are more likely to be represented as separated morphemic entries while high-frequency words are more likely to be represented as whole-word entries in the mental lexicon (Pollatsek et al., 2000; Yan et al., 2006). Thus, it is possible that the separated morphemic representations of low-frequency words may also increase the saliency of the linking phonological representations (i.e., rhyme representation) compared to high-frequency words. In our current study, the mean word frequency of the target words is 2.39 occurrences per million. Based on a Chinese Linguistic Data Consortium (2003), only 22.86% words' frequencies are higher than the targets. Therefore, target words in current study are relatively high-frequency Chinese words and the rhyme representations of those words are not that salient. In this study, we did not manipulate the word frequency of targets directly, and thus it remains unclear how the word frequency would influence rhyme processing and representations. More studies need to be conducted to further examine this issue.

One may argue that the phonological competition effect observed in the target-absent design in Experiment 2 may have resulted from task-specific strategies. For example, participants may allocate more attentional resources to the target absent displays because the task was to search for a matched target. However, this explanation is less plausible because: (1) the visual display was previewed for 200 ms before the onset of the auditory words. Given that 200 ms is usually assumed to be the retrieval time for phonological information from printed words (Huettig and McQueen, 2007), this preview time prevented listeners from searching for visual words strategically and accessing the phonological code of printed words based on the phonological information of the spoken targets. The setting of the preview time allows that the phonological information of the spoken word and the printed words was accessed concurrently and the eye movement measures of the printed words reflected the ongoing cognitive processing during spoken word recognition rather than a later search effect after the target words had been activated. (2) In addition, participants in the current study were instructed to search for a target word in the visual display and responded by pressing keys. It should be noted that no explicit phonological processing was necessary in order to complete the task. If the phonological competition effect had resulted from a visual search strategy, then the fixation proportion under both syllabic overlap and sub-syllabic overlap conditions should be the same since both the syllabic overlap and sub-syllabic overlap competitors were not targets and there was no reason for participants' eyes to fixate on these visual referents. However, the significant competition effect suggested that the phonological information of the second syllable was indeed activated. Based on this fact, we argue that the phonological information of the second syllable was activated automatically to a larger extent during the visual search and less likely to have been the result of the task. More future studies need to be conducted to further examine whether the phonological competition effect on rhymes and on partial phonemes of rhymes still exist under a more general and natural language processing situation.

Our study has several limitations worth noting. First, the lexical tone of the phonological competitors was not manipulated in the current study. One of our prior studies showed that lexical tone affects the degree of activation of cohort competitors (Shen et al., 2020), and thus it remains unclear how the lexical tone of rhymes might modulate the activation degree of rhyme competitors. This should be investigated in future studies. Second, the current study did not include a cohort competitor type in the word level, and thus it was unable to directly compare the processing differences in time course and activation degree between the two types of phonological candidates on same level during spoken word recognition. Third, in the current study, we only considered the processing of word-initial phonemes (i.e., cohort-part) in the syllable level (i.e., second syllable). Future studies need to be conducted to further explore whether the word-final phonemes (i.e., rhyme-part) of the second syllable can also be activated. These studies have important implications in uncovering the uniqueness of rhyme processing in Chinese. Fourth, current study did not intentionally control or manipulate the participants' cognitive/meta-linguistic abilities or the familiarity degree of Chinese printed words. It is possible that those factors may also exert some confounding influence on our current findings. Future studies need to be conducted to further investigate the role of these factors on the syllable processing in Chinese. Lastly, more future studies need to be designed to investigate how context predictability may influence rhyme processing in Chinese spoken perception.

Taken together, the current study confirmed that not only the second syllable (i.e., rhyme) at the word level is activated, but also the initial phonemes of the second syllable at the syllabic level are also activated for competition in Chinese spoken disyllabic word recognition.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author via shen_wei@yahoo.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Institute of Psychological Sciences from Hangzhou Normal University. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YW and WS conceived and designed the experiments. YW and HZ performed the experiments. YW and WS analyzed the data. YW, XZ, and WS wrote the manuscript. All the authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by grants from the National Natural Science Foundation of China (Grant No. 31800918) and a Start-up funding from Hangzhou Normal University (Grant No. RWSK20201034) awarded to WS.

REFERENCES

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439. doi: 10.1006/jmla.1997.2558

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bate, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Brouwer, S., and Bradlow, A. R. (2016). The Temporal Dynamics of Spoken Word Recognition in Adverse Listening Conditions. *J. Psychol. Res.* 45, 1151–1160. doi: 10.1007/s10936-015-9396-9

Chinese Linguistic Data Consortium (2003). 现代汉语通用词表 [Chinese lexicon] (CLDC-LAC-2003-001). Beijing, China: Tsinghua University, State Key Laboratory of Intelligent Technology and Systems, and Chinese Academy of Sciences, Institute of Automation. Available online at: http://www.chineseldc.org.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Lang. Res.* 28, 369–382. doi: 10.1177/0267658312443651

Desroches, A. S., Joanisse, M. F., and Robertson, E. K. (2006). Specific phonological impairments in dyslexia revealed by eye tracking. *Cognition* 100, B32–B42. doi: 10.1016/j.cognition.2005.09.001

Desroches, A. S., Newman, R. L., and Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *J. Cognit. Neurosci.* 21, 1893–1906. doi: 10.1162/jocn.2008.21142

Ferreira, F., Foucart, A., and Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *J. Mem. Lang.* 69, 165–182. doi: 10.1016/j.jml.2013.06.001

Frassinelli, D., and Keller, F. (2012). The plausibility of semantic properties generated by a distributional model: Evidence from a visual world experiment. *Proc. Annu. Meet. Cognit. Sci. Soc.* 34, 1560–1565.

Hendrickson, K., Spinelli, J., and Walker, E. (2020). Cognitive processes underlying spoken word recognition during soft speech. *Cognition* 198:104196. doi: 10.1016/j.cognition.2020.104196

Huettig, F., and Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: semantic competition effects and the visual world paradigm. *Cognition* 96, B23–B32. doi: 10.1016/j.cognition.2004.10.003

Huettig, F., and McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *J. Mem. Lang.* 57, 460–482. doi: 10.1016/j.jml.2007.02.001

Huettig, F., Olivers, C. N., and Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychol.* 137, 138–150. doi: 10.1016/j.actpsy.2010.07.013

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., and Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *J. Mem. Lang.* 86, 157–171. doi: 10.1016/j.jml.2015.10.007

Ito, A., Pickering, M. J., and Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *J. Mem. Lang.* 98, 1–11. doi: 10.1016/j.jml.2017.09.002

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007

Laszlo, S., and Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *J. Mem. Lang.* 61, 326–338. doi: 10.1016/j.jml.2009.06.004

Liu, S., and Samuel, A. G. (2007). The role of Mandarin lexical tones in lexical access under different contextual conditions. *Lang. Cognit. Proces.* 22, 566–594. doi: 10.1080/01690960600989600

Liu, Y., Shu, H., and Wei, J. (2006). Spoken word recognition in context: Evidence from Chinese ERP analyses. *Brain Lang.* 96, 37–48. doi: 10.1016/j.bandl.2005.08.007

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon. Research on Speech Perception. Technical Report No. 6*. Bloomington, IN: Indiana University.

Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear Hear.* 19, 1–36. doi: 10.1097/00003446-199802000-00001

Malins, J. G., and Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eye-tracking study. *J. Mem. Lang.* 62, 407–420. doi: 10.1016/j.jml.2010.02.004

Malins, J. G., and Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia* 50, 2032–2043. doi: 10.1016/j.neuropsychologia.2012.05.002

Marslen-Wilson, W., and Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition* 8, 1–7. doi: 10.1016/0010-0277(80)90015-3

Marslen-Wilson, W., and Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *J. Exper. Psychol. Hum. Percep. Perform.* 15, 576–585. doi: 10.1037/0096-1523.15.3.576

McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cognit. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0

McQueen, J. M., and Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Q. J. Exper. Psychol.* 60, 661–671. doi: 10.1080/17470210601183890

Meng, X. (2014). *语音信息影响中文听觉词汇识别的眼动研究* [*The influence of phonological information on spoken word recognition in mandarin Chinese: Evidence from Eye movements]* Master dissertation. Dalian: Liaoning Normal University.

Otten, M., Nieuwland, M. S., and Van Berkum, J. J. A. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neurosci.* 8:89. doi: 10.1186/1471-2202-8-89

Pollatsek, A., Hyönä, J., and Bertram, R. (2000). The role of morphological constituents in Reading finnish compound words. *J. Exper. Psychol. Hum. Percep. Perform.* 26, 820–833. doi: 10.1037/0096-1523.26.2.820

Quené, H., and van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *J. Mem. Lang.* 59, 413–425. doi: 10.1016/j.jml.2008.02.002

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rayner, K., and Clifton, C. (2009). Language processing in reading and speech perception is fast and incremental: Implications for event-related potential research. *Biol. Psychol.* 80, 4–9. doi: 10.1016/j.biopsycho.2008.05.002

Shatzman, K. B., and McQueen, J. M. (2006). The modulation of lexical competition by segment duration. *Psychon. Bull. Rev.* 13, 966–971. doi: 10.3758/BF03213910

Shen, W., Hyönä, J., Wang, Y., Hou, M., and Zhao, J. (2020). The role of tonal information during spoken-word recognition in Chinese: Evidence from a printed-word eye-tracking study. *Mem. Cognit.* 49, 181–192. doi: 10.3758/s13421-020-01070-0

Shen, W., Qu, Q., and Tong, X. (2018). Visual attention shift to printed words during spoken word recognition in Chinese: The role of phonological information. *Mem. Cognit.* 46, 642–654. doi: 10.3758/s13421-018-0790-z

Simmons, E., and Magnuson, J. (2018). Word length, proportion of overlap, and phonological competition in spoken word recognition. *Proc. Cogn. Sci. Soc.* 2018, 1064–1069.

Sorensen, D. W., and Bailey, K. G. D. (2007). The world is too much: Effects of array size on the link between language comprehension and eye movements. *Vis. Cognt.* 15, 112–115.

Tong, X., Mcbride, C., and Burnham, D. (2014). Cues for lexical tone perception in children: acoustic correlates and phonetic context effects. *J. Speech Lang. Hear. Res.* 57, 1589–1605. doi: 10.1044/2014_JSLHR-S-13-0145

Tsang, Y.-K., and Chen, H.-C. (2010). Morphemic ambiguity resolution in Chinese: Activation of the subordinate meaning with a prior dominant-biased context. *Psychon. Bull. Rev.* 17, 875–881. doi: 10.3758/PBR.17.6.875

Weber, A., Melinger, A., and Tapia, L. (2007). "The mapping of phonetic information to lexical representation in Spanish: Evidence from eye movements," in *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, eds J. Trouvain and W. J. Barry (Dudweiler: Pirrot), 1941–1944.

Yan, G., Tian, H., Bai, X., and Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *Br. J. Psychol.* 97, 259–268. doi: 10.1348/000712605X70066

Zhao, J., Guo, J., Zhou, F., and Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia* 49, 1761–1770. doi: 10.1016/j.neuropsychologia.2011.02.054

Zhou, X., and Marslen-Wilson, W. (1994). Words, morphemes and syllables in the Chinese mental lexicon. *Lang. Cogn. Proces.* 9, 393–422. doi: 10.1080/01690969408402125

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

TABLE A1 | Materials used in Experiments 1 and 2.

Id	Target	Phonological competitor (Syllabic overlap condition)	Phonological competitor (Sub-syllabic overlap condition)	Distractor	Distractor
1	电铃 (dian4 ling2)	幽灵 (you1 ling2)	山林 (shan1 lin2)	房东 (fang2 dong1)	默契 (mo4 qi4)
2	脾脏 (pi2 zang4)	土葬 (tu3 zang4)	短暂 (duan3 zan4)	嫩叶 (nen4 ye4)	麦穗 (mai4 sui4)
3	山羊 (shan1 yang2)	洛阳 (luo4 yang2)	感言 (gan3 yan2)	体操 (ti3 cao1)	媒介 (mei2 jie4)
4	豆浆 (dou4 jiang1)	边疆 (bian1 jiang1)	管家 (guan3 jia1)	温室 (wen1 shi4)	孔雀 (kong3 que4)
5	纸张 (zhi3 zhang1)	图章 (tu2 zhang1)	残渣 (can2 zha1)	水壶 (shui3 hu2)	祖辈 (zu3 bei4)
6	旧账 (jiu4 zhang4)	方丈 (fang1 zhang4)	油炸 (you2 zha4)	星空 (xing1 kong1)	拖鞋 (tuo1 xie2)
7	公转 (gong1 zhuan4)	小篆 (xiao3 zhuan4)	圆柱 (yuan2 zhu4)	城楼 (cheng2 lou2)	白银 (bai2 yin2)
8	浑浊 (hun2 zhuo2)	烧灼 (shao1 zhuo2)	翠竹 (cui4 zhu2)	情义 (qing2 yi4)	泰斗 (tai4 dou3)
9	阴阳 (yin1 yang2)	绵羊 (mian2 yang2)	精盐 (jing1 yan2)	自卑 (zi4 bei1)	气球 (qi4 qiu2)
10	臭氧 (chou4 yang3)	瞻仰 (zhan1 yang3)	眉眼 (mei2 yan3)	内科 (nei4 ke1)	绳索 (sheng2 suo3)
11	精英 (jing1 ying1)	妇婴 (fu4 ying1)	高音 (gao1 yin1)	圣旨 (sheng4 zhi3)	电车 (dian4 che1)
12	军营 (jun1 ying2)	轻盈 (qing1 ying2)	水银 (shui3 yin2)	曲调 (qu3 diao4)	早餐 (zao3 can1)
13	纸箱 (zhi3 xiang1)	丁香 (ding1 xiang1)	对虾 (dui4 xia1)	阴间 (yin1 jian1)	跳蚤 (tiao4 zao3)
14	母性 (mu3 xing4)	荣幸 (rong2 xing4)	书信 (shu1 xin4)	脸蛋 (lian3 dan4)	电量 (dian4 liang4)
15	兴旺 (xing1 wang4)	遗忘 (yi2 wang4)	手腕 (shou3 wan4)	奥秘 (ao4 mi4)	交警 (jiao1 jing3)
16	轻伤 (qing1 shang1)	智商 (zhi4 shang1)	乌纱 (wu1 sha1)	毒瘤 (du2 liu2)	独白 (du2 bai2)
17	松香 (song1 xiang1)	梦乡 (meng4 xiang1)	龙虾 (long2 xia1)	卷发 (juan3 fa4)	野营 (ye3 ying2)
18	双脚 (shuang1 jiao3)	直角 (zhi2 jiao3)	马甲 (ma3 jia3)	恩爱 (en1 ai4)	火海 (huo3 hai3)
19	辣酱 (la4 jiang4)	锁匠 (suo3 jiang4)	年假 (nian2 jia4)	巫女 (wu1 nv3)	戏服 (xi4 fu2)
20	王冠 (wang2 guan1)	双关 (shuang1 guan1)	地瓜 (di4 gua1)	舅妈 (jiu4 ma1)	孩童 (hai2 tong2)
21	芦荡 (lu2 dang4)	文档 (wen2 dang4)	鸭蛋 (ya1 dan4)	恶梦 (e4 meng4)	衣架 (yi1 jia4)
22	朝纲 (chao2 gang1)	水缸 (shui3 gang1)	猪肝 (zhu1 gan1)	牙医 (ya2 yi1)	逆贼 (ni4 zei2)
23	干冰 (gan1 bing1)	败兵 (bai4 bing1)	礼宾 (li3 bin1)	玉佩 (yu4 pei4)	黑锅 (hei1 guo1)
24	佚名 (yi4 ming2)	神明 (shen2 ming2)	烟民 (yan1 min2)	街区 (jie1 qu1)	水盆 (shui3 pen2)
25	车篷 (che1 peng2)	暖棚 (nuan3 peng2)	瓦盆 (wa3 pen2)	腮红 (sai1 hong2)	柔光 (rou2 guang1)
26	乳糖 (ru3 tang2)	澡堂 (zao3 tang2)	酒坛 (jiu3 tan2)	稀泥 (xi1 ni2)	幼芽 (you4 ya2)
27	科幻 (ke1 huan4)	祸患 (huo4 huan4)	白桦 (bai2 hua4)	桌布 (zhuo1 bu4)	街巷 (jie1 xiang4)
28	胜仗 (sheng4 zhang4)	结账 (jie2 zhang4)	爆炸 (bao4 zha4)	陵墓 (ling2 mu4)	座椅 (zuo4 yi3)
29	肾脏 (shen4 zang4)	宝藏 (bao3 zang4)	夸赞 (kua1 zan4)	美梦 (mei3 meng4)	红日 (hong2 ri4)
30	弹簧 (tan2 huang2)	女皇 (nv3 huang2)	精华 (jing1 hua2)	油菜 (you2 cai4)	旋风 (xuan2 feng1)
31	春风 (chun1 feng1)	先锋 (xian1 feng1)	高分 (gao1 fen1)	嘴巴 (zui3 ba1)	机密 (ji1 mi4)
32	勋章 (xun1 zhang1)	纸张 (zhi3 zhang1)	山楂 (shan1 zha1)	美景 (mei3 jing3)	幽灵 (you1 ling2)
33	宝剑 (bao3 jian4)	组件 (zu3 jian4)	暑假 (shu3 jia4)	群落 (qun2 luo4)	指纹 (zhi3 wen2)
34	界限 (jie4 xian4)	电线 (dian4 xian4)	属下 (shu3 xia4)	老虎 (lao3 hu3)	滋味 (zi1 wei4)
35	元帅 (yuan2 shuai4)	蟋蟀 (xi1 shuai4)	票数 (piao4 shu4)	情侣 (qing2 lv3)	下肢 (xia4 zhi1)
36	驾照 (jia4 zhao4)	口罩 (kou3 zhao4)	奸诈 (jian1 zha4)	蚊虫 (wen2 chong2)	灾荒 (zai1 huang1)
37	香肠 (xiang1 chang2)	特长 (te4 chang2)	巡察 (xun2 cha2)	粉笔 (fen3 bi3)	丝巾 (si1 jin1)
38	主线 (zhu3 xian4)	权限 (quan2 xian4)	惊吓 (jing1 xia4)	明月 (ming2 yue4)	物料 (wu4 liao4)
39	失明 (shi1 ming2)	芳名 (fang1 ming2)	灾民 (zai1 min2)	烤鸭 (kao3 ya1)	油污 (you2 wu1)
40	肉馅 (rou4 xian4)	眼线 (yan3 xian4)	初夏 (chu1 xia4)	宝刀 (bao3 dao1)	徽章 (hui1 zhang1)
41	棒冰 (bang4 bing1)	逃兵 (tao2 bing1)	来宾 (lai2 bin1)	面皮 (mian4 pi2)	家眷 (jia1 juan4)
42	行程 (xing2 cheng2)	商城 (shang1 cheng2)	清晨 (qing1 chen2)	古董 (gu3 dong3)	枝叶 (zhi1 ye4)
43	圣灵 (sheng4 ling2)	欺凌 (qi1 ling2)	松林 (song1 lin2)	长辈 (zhang3 bei4)	操场 (cao1 chang3)
44	圆锥 (yuan2 zhui1)	尾追 (wei3 zhui1)	植株 (zhi2 zhu1)	青蛙 (qing1 wa1)	佳肴 (jia1 yao2)
45	情商 (qing2 shang1)	枪伤 (qiang1 shang1)	细砂 (xi4 sha1)	问卷 (wen4 juan3)	媒婆 (mei2 po2)
46	藏青 (zang4 qing1)	左倾 (zuo3 qing1)	单亲 (dan1 qin1)	指环 (zhi3 huan2)	假牙 (jia3 ya2)
47	水缸 (shui3 gang1)	朝纲 (chao2 gang1)	标杆 (biao1 gan1)	酒窝 (jiu3 wo1)	脑瘤 (nao3 liu2)
48	赏封 (shang3 feng1)	险峰 (xian3 feng1)	秋分 (qiu1 fen1)	灵位 (ling2 wei4)	厨艺 (chu2 yi4)

(Continued)

TABLE A1 | (Continued)

Id	Target	Phonological competitor (Syllabic overlap condition)	Phonological competitor (Sub-syllabic overlap condition)	Distractor	Distractor
49	公关 (gong1 guan1)	外观 (wai4 guan1)	西瓜 (xi1 gua1)	热量 (re4 liang4)	昆虫 (kun1 chong2)
50	生姜 (sheng1 jiang1)	九江 (jiu3 jiang1)	亲家 (qing4 jia1)	北欧 (bei3 ou1)	富豪 (fu4 hao2)
51	特效 (te4 xiao4)	欢笑 (huan1 xiao4)	盛夏 (sheng4 xia4)	踪影 (zong1 ying3)	陵墓 (ling2 mu4)
52	酒精 (jiu3 jing1)	水晶 (shui3 jing1)	钢筋 (gang1 jin1)	植被 (zhi2 bei4)	舌头 (she2 tou)
53	蛛网 (zhu1 wang3)	来往 (lai2 wang3)	饭碗 (fan4 wan3)	水池 (shui3 chi2)	品格 (pin3 ge2)
54	证件 (zheng4 jian4)	弓箭 (gong1 jian4)	书架 (shu1 jia4)	月饼 (yue4 bing3)	明珠 (ming2 zhu1)

Words' pronunciations are presented in brackets.

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 November 2021
doi: 10.3389/fpsyg.2021.712309

The Mindset of Intelligence Is Not a Contributor of Placebo Effects in Working Memory Training

Peibing Liu[1†], *Xin Zhang*[1,2†] *and Renlai Zhou*[1*]

[1] Department of Psychology, Nanjing University, Nanjing, China, [2] Smart Home Solution BU of Innovation Business Group, TCL Industries Holdings Co., Ltd., Huizhou, China

Whether working memory training is effective in enhancing fluid intelligence remains in dispute. Several researchers, who doubt the training benefits, consider that placebo effects may be the reason for positive training gains. One of the vital variables that may induce the placebo effect is the mindset of intelligence. In this article, we provide a test of whether the mindset of intelligence leads to placebo effects in working memory training. Participants were overtly recruited and allocated to the growth mindset group or the fixed mindset group by Theories of Intelligence Scale scores. A single, 1 h session working memory training is the cue to introduce the placebo effects. During pre/post-testing, all participants completed tasks measuring working memory capacity (near transfer) and fluid intelligence (far transfer). Our findings show no significant difference between the two groups in both tasks. Therefore, these results suggest that the placebo effect does not exist in this study, which means individuals' mindset of intelligence may not be a contributor to the placebo effect in 1 h working memory training. This research will further help to clarify the mechanism of the placebo effect in working memory training.

Keywords: working memory training, fluid intelligence, placebo effects, mindset of intelligence, transfer effect

Edited by:
Edwin James Burns,
Edge Hill University, United Kingdom

Reviewed by:
Vanessa R. Simmering,
University of Kansas, United States
Elina Kuusisto,
Tampere University, Finland

***Correspondence:**
Renlai Zhou
rlzhou@nju.edu.cn

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cognition,
a section of the journal Frontiers in Psychology

***Received:** 20 May 2021*
***Accepted:** 23 September 2021*
***Published:** 04 November 2021*

Citation:
Liu P, Zhang X and Zhou R (2021) The Mindset of Intelligence Is Not a Contributor of Placebo Effects in Working Memory Training. Front. Psychol. 12:712309. doi: 10.3389/fpsyg.2021.712309

INTRODUCTION

Working memory is a cognitive system that plays a crucial role in keeping things in mind while performing complex attentional-cognitive control activities such as goal-directed behavior, reasoning, decision-making, comprehension, and learning (Kane and Engle, 2002; Holmes et al., 2009; Baddeley, 2010; Shahar et al., 2018). From this perspective, working memory training is assumed to improve not only working memory capacity but also a battery of related abilities. Several studies have verified the assumption that the training can enhance attention (Chein and Morrison, 2010; Kundu et al., 2013), decrease attention deficit hyperactivity disorder (ADHD)-related symptoms (Klingberg et al., 2002, 2005), and strengthen reading or language comprehension (Carretti et al., 2013, 2014; Artuso et al., 2019). In 2008, a study found a promising result, short-term working memory training can improve the fluid intelligence of healthy adults, to support that fluid intelligence is trainable (Jaeggi et al., 2008; Sternberg, 2008). Building on this initial research, more studies on working memory training on fluid intelligence have accumulated (Jaeggi et al., 2011, 2014; Hardy et al., 2015). Fluid intelligence refers to the ability to solve novel, abstract problems through insight into complex relationships without relying on previous knowledge experience (Cattell, 1963). Fluid intelligence is not only the basis for other cognitive abilities but also plays a key role in how we solve problems in daily work and life and how we adapt to new situations (Sternberg and Gastel, 1989). Although many studies have reported that the near-transfer effect (i.e., increased working memory ability) is statistically significant and lasts for several months, some controversial results remain on the far-transfer effect (i.e., improvement in other abilities), especially on the transfer to fluid intelligence (Colom et al., 2010, 2013; Owen et al., 2010; Chooi and Thompson, 2012; Bastian and Oberauer, 2013; Redick et al., 2013; Sprenger et al., 2013; Thompson et al., 2013; Bastian and Eschen, 2016; Lawlor-Savage and Goghari, 2016; Clark et al., 2017). Whether working memory training can increase fluid intelligence still lacks consistent evidence.

Considering the argument, the role of placebo effects in early positive findings, expectations may lead to post-training fluid intelligence gains, is of a central concern (Shipstead et al., 2012; Slagter, 2012; Melby-Lervåg and Hulme, 2013; Redick et al., 2013). Placebo effects are psychophysiological changes caused by the symbolic meaning of treatment rather than specific pharmacological or physiological properties (Brody, 1980; Stewart-Williams and Podd, 2004). It is well-known that in drug trials, the control group takes a placebo pill (which looks the same as the experimental group) to promise both to have the same anticipation for the pills. In medical practice, one reason for the improvement experienced by a patient after treatment is the confidence of the patient in the healer or the drug is so strong that the psychological effect of the relief alleviates their condition (Zhang et al., 2011). The improvement is not caused by the treatment itself. It is the belief and expectation (the treatment will work) that lead to the placebo effect. Psychological intervention, including working memory training, should also pay attention to the difference in expectations between groups (Boot et al., 2013). Foroughi et al. (2016) published an infusive report confirming the placebo effects in 1 h working memory training. In this research, the placebo group was recruited with a flyer that overtly advertised the cognitive enhancement effect of working memory training; the control group was recruited covertly with a visually similar flyer. During the 1 h training session, the results showed that fluid intelligence was significantly improved in the placebo group rather than the control group. These results suggest that the observed effects are due to overt recruitment (reveal the objective to induce expectancies), which challenges the true efficacy of working memory training.

It is worth noting that, in the study of Foroughi et al. (2016), the average Theories of Intelligence Scale (TIS) scores of the placebo group were significantly higher than the control group. The TIS measures the mindset of intelligence, which reflects beliefs regarding the malleability of intelligence. Students can hold different "theories" about their intelligence (Dweck, 2000). Some students who have a fixed mindset believe that their intelligence is unchangeable (an entity theory). Others who have a growth mindset believe that their intelligence can be developed through effort and persistence (an incremental theory). Individuals who have high TIS scores may consider that their intelligence can be improved. However, Foroughi et al. only attributed positive post-test outcomes to the recruitment methods. The role of the mindset of intelligence in placebo effects is ambiguous.

According to the mindset theory (Dweck, 2000), students with a growth mindset have many benefits, including adherence to learning goals rather than performance goals (Richard and Pals, 2010), positive effort belief (Tempelaar et al., 2015), and holding effort and positive strategies (Ommundsen et al., 2005). These also lead to better academic achievement (Claro et al., 2016). However, this theory has been discredited and inconsistent empirical findings were found (Dommett et al., 2013). A recent meta-analysis (Sisk et al., 2018) found that the relationship between mindset interventions and academic achievement was non-significant, which did not support Dweck's claim. This also suggests that more related research is needed. Therefore, our study (the mindset of intelligence is a critical variable) can also be seen as an answer to this need.

Several studies emphasize the influence of individual differences on training results (Jaeggi et al., 2014; Guye et al., 2017), among which mindset of intelligence is an essential factor. Individuals who think their intelligence is malleable show a greater transfer effect in training than those who think intelligence can hardly be changed through effort (Jaeggi et al., 2014). Clinical empirical studies have supported that mindset is one of the factors inducing the placebo effect. Mindset can lead to the attention and motivation of patients and affect the subjective and objective measurements of health and well-being (Crum and Zuckerman, 2017). Moreover, according to the study on stress, diet, and exercise, the mindset was related to mental and physical well-being, including blood pressure, weight loss, and cortical and hormonal responses (Crum and Langer, 2007; Crum et al., 2013; Crum and Zuckerman, 2017). Before being informed of the disease and treatment information, patients already have a certain mindset, which can interpret the information reception, affect subsequent expectancies, and induce the placebo effect (Zion and Crum, 2018). Similarly, the mindset of intelligence could be a contributor to placebo effects in working memory training. That is, the variable, mindset of intelligence, may have confounded the results reported in the Foroughi et al.'s (2016) study. Therefore, empirical research is needed to examine whether overt recruitment or the interaction between overt recruitment and mindset of intelligence contributes to the placebo effects in working memory training.

To answer this question, we adopted the same recruitment paradigm, procedure, and sample size as in the study by Foroughi et al. (2016). Additionally, we ensured no difference in TIS scores between the placebo group and the control group (Zhang et al., 2019). However, we found no improvement of fluid intelligence in either group, which failed to replicate the findings of the Foroughi et al.'s (2016) study. These results ruled out the separate role of overt recruitment in positive post-test outcomes, leading us to wonder whether the mindset of intelligence is the cause of the placebo effect. This study aims to illustrate these questions.

For this aim, in this study, we recruited both groups overtly by advertising fluid intelligence improvement. Importantly, before the formal experiments, participants filled in the TIS and were divided into the growth mindset group (higher TIS scores) and the fixed mindset group (lower TIS scores). Therefore, if the growth group rather than the fixed group has task performance enhancement (placebo effect), it would support that mindset of intelligence is the contributor to task performance. Three different versions of adaptive working memory training tasks were used in the 1 h cognitive training session. In addition to the fluid intelligence (far-transfer effect), we also used a 2-back task to measure the near-transfer effect.

MATERIALS AND METHODS

Participants

Participants were recruited at Nanjing University *via* an overt advertisement poster stating that "Various studies indicate that working memory training can improve fluid intelligence" (**Figure 1**). Participants are all adults aged between 18 and 25 years, right-handed, in good health, and not taking any drugs. All participants provided informed, written consent before the formal experiment.

We recruited a total of 50 participants and divided them into two groups of 25 persons, the same size as Foroughi et al.'s (2016) study, based on their TIS scores. The specific recruitment process is as follows: we first show our recruitment posters. If students are interested in our experiment, they can scan the QR code on the posters to fill in a questionnaire, which is the TIS. After the statistical analysis on the scale scores of participants, we will invite participants with scores >35 and <20 to participate in our formal experiments. The TIS scores in the growth mindset group are >35 ($M = 36.6$, $SD = 2.26$); in the fixed mindset group, the TIS scores are <20 ($M = 16.92$, $SD = 2.93$). **Table 1** provides detailed information on participants. The two groups did not demonstrate any significant difference in gender ratio [χ^2 (1, $N = 50$) $= 0.104$, $P = 0.747$, Cramer's $V = 0.046$] or in average age [$t_{(48)} = 1.393$, $P = 0.170$, Cohen's $d = 0.39$, 95% CI $= -0.337$, 1.857].

Procedure Overview

Figure 2 describes the procedure for the study. The experiment was divided into three parts: pretest, working memory training, and post-test. All participants attended the pretest and post-test tasks (2-back task measuring near-transfer effect; Raven's Advanced Progressive Matrices (RAPM) measuring far-transfer effect). A single, 1 h session working memory training was conducted between the pretest and post-test. Similar to Foroughi et al. (2016), the reason for choosing an hour as training duration is that training time should be adequate (roughly 20 sessions, each lasting 30–60 min) to make cognitive training effective (Shipstead et al., 2012), so the positive outcomes from 1 h training must be due to placebo effects. At the end of the experiment, all participants were compensated with money.

Training Tasks

After the pretest, participants completed a computerized working memory training. We used three different versions of adaptive working memory training tasks, including three kinds of memory materials: animals, letters, and positions (**Figure 3**). Based on the classical running memory span task, many studies have previously adopted the training tasks in cognitive training research (Zhao et al., 2011; Wang et al., 2014; Chen et al., 2018). We will take the animal training task as an example to explain the operation of training tasks in detail. In the animal training task, animals are different and presented in the center of the screen in sequence. In each trial, the number of animals varied randomly from 5 to 7, 9, and 11, and participants were asked to remember the last three animals presented in this trial. It is worth noting that participants could not predict the number of animals that would appear in each trial, they were not told, so they had to update the memory items constantly. It would train the working memory updating abilities. The other two training tasks are similar to the animal training task: participants must report the last three letters of the alphabet training task and the last three animation positions in the position training task. Each training task consists of 30 trials, which are divided into six blocks with five trials each. At the beginning of the training, the duration of each stimulus was 1,750 ms. If participants correctly reported three or more trials in this block, the duration would be decreased by 100 ms in the next block. In this study, participants completed each training task two times in a randomized order, which lasted approximately an hour.

Transfer Measurements

Working Memory Capacity (Near Transfer)

We used a computerized 2-back task to assess the near-transfer effect in working memory training. In this task, participants were asked to press key "F" if the currently presented item was the same as the item presented two steps earlier; press "J" if not. The matching and mismatching stimuli were both presented 50% in this task. A "+" was always presented in the center of the screen, and a series of numbers, ranging from 0 to 9, will appear at the top, bottom, left, and right of the "+." Participants are required to ignore the verbal information and judge whether the digital space position of the current number matches the target stimulus. Numbers and their spatial positions are both random. Each trial consisted of a fixation (200 ms), a blank screen interval (1,300 ms), target (200 s), and reaction time (until response or until 2,500 ms). Reaction time and accuracy were included in the analysis.

Fluid Intelligence (Far Transfer)

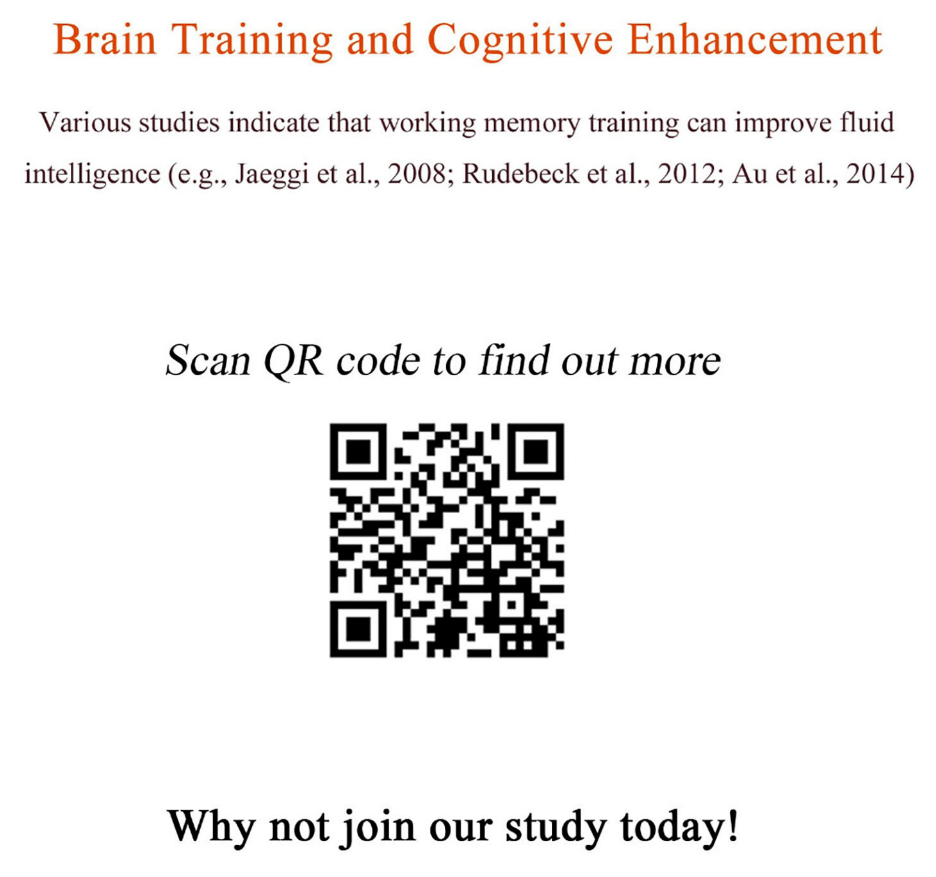
One of the far-transfer effects that researchers are most concerned with is fluid intelligence. In this study, we utilized Raven's Advanced Progressive Matrices (RAPM), commonly used in adult intelligence research, to assess fluid intelligence change (Raven et al., 1998). Referring to Jaeggi et al. (2008), we used parallel forms for the pretest and post-test by dividing the RAPM test into even and odd items.

RESULTS

All analyses were conducted using mixed-effects linear regression with restricted maximum likelihood.

Training Effects

We compared whether there were differences in training tasks between the two groups (**Figure 4**). The performance of participants in training tasks can be measured by the task difficulty level they eventually reach. The time interval between adjacent stimuli presents the difficulty of the training task. The shorter the time interval between stimuli, the more difficult the



Brain Training and Cognitive Enhancement

Various studies indicate that working memory training can improve fluid intelligence (e.g., Jaeggi et al., 2008; Rudebeck et al., 2012; Au et al., 2014)

Scan QR code to find out more

Why not join our study today!

FIGURE 1 | Poster (the original poster was in Chinese).

TABLE 1 | Characteristics of participants.

	Growth mindset group ($N = 25$)	Fixed mindset group ($N = 25$)	Group differences (p-value)
Gender (male: female)	7:18	6:19	0.747
Age (years; M ± SD)	20.76 ± 2.17	20.00 ± 1.66	0.170
TIS scores	36.6 ± 2.26	16.92 ± 2.93	<0.001

task is. The baseline interval was 2,250 ms. The maximum level of difficulty that participants attained did not differ between the two groups: $B = 72$, SE $= 87.64$, $t_{(48)} = 0.822$, $P = 0.416$, $b = 0.24$ for animal task; $B = -16$, SE $= 51.90$, $t_{(48)} = -0.308$, $P = 0.759$, $b = 0.09$ for location task; and $B = -24$, SE $= 24.77$, $t_{(48)} = -0.969$, $P = 0.338$, $b = 0.30$ for letter task, respectively.

Transfer Effects

Two-Back Task (Near Transfer)

We calculated the accuracy and reaction time difference between the two groups in 2-back task from pretest to post-test. At the pretest, the two groups did not significantly differ in accuracy [$B = -0.27$, SE $= 0.41$, $t_{(48)} = -0.670$, $P = 0.506$, $b = 0.19$] or reaction times [$B = -75.07$, SE $= 38.05$, $t_{(48)} = -1.973$, $P = 0.054$, $b = 0.58$]. After the training, there was still no significant difference in accuracy [$B = 0.0004$, SE $= 0.02$, $t_{(48)} = 0.018$, $P = 0.986$, $b = 0.005$] or reaction times [$B = -19.72$, SE $= 42.42$, $t_{(48)} = -0.465$, $P = 0.644$, $b = 0.14$]. We observed a main effect of time on the accuracy [$B = -0.13$, SE $= 0.03$, $t_{(48)} = -3.809$, $P < 0.001$, $d = 0.88$], but not on the reaction time [$B = 31.32$, SE $= 40.29$, $t_{(48)} = 0.777$, $P = 0.439$, $d = 0.16$]. Both interactions between time and group were not observed [ACC: $B = -0.03$, SE $= 0.47$, $t_{(48)} = -0.595$, $P = 0.554$, $d = 0.14$; and RT: $B = -55.34$, SE $= 56.98$, $t_{(48)} = -0.971$, $P = 0.334$, $d = 0.20$] (see **Figure 5**).

Raven's Advanced Progressive Matrices (Far Transfer)

We analyzed the performance of RAPM between the two groups. The two groups did not differ in pretest [B $= -64$, SE $= 0.50$, $t_{(48)} = -1.290$, $P = 0.203$, $b = 0.38$) or post-test ($B = -0.16$, SE $= 0.53$, $t_{(48)} = -0.304$, $P = 0.762$, $b = 0.09$). We did not observe a main effect of time on test performance [$B = 0.56$,

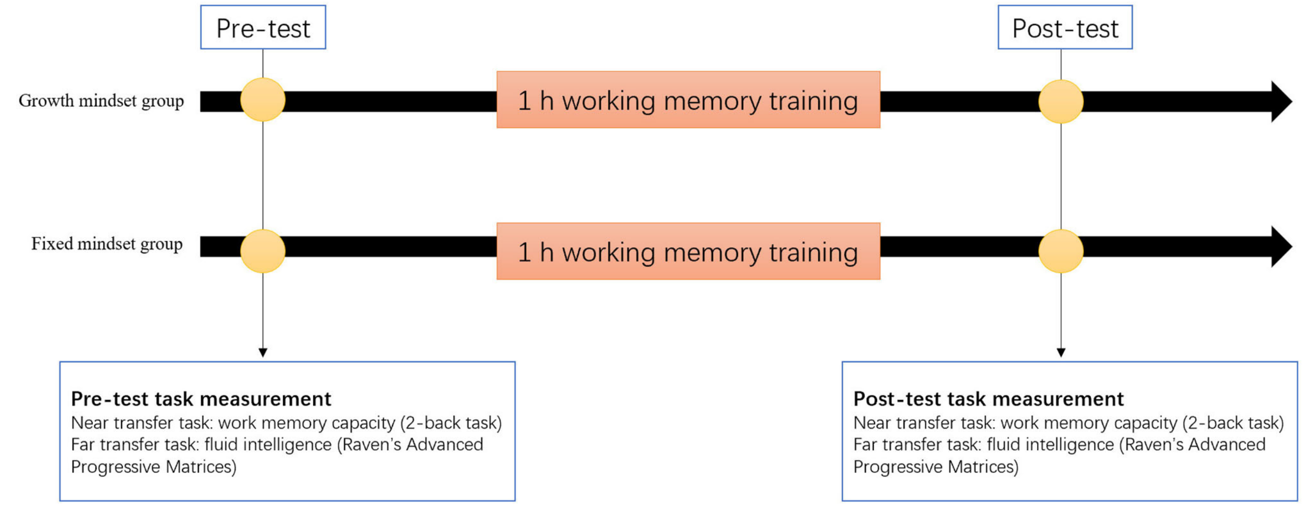


FIGURE 2 | Procedure for the training study.

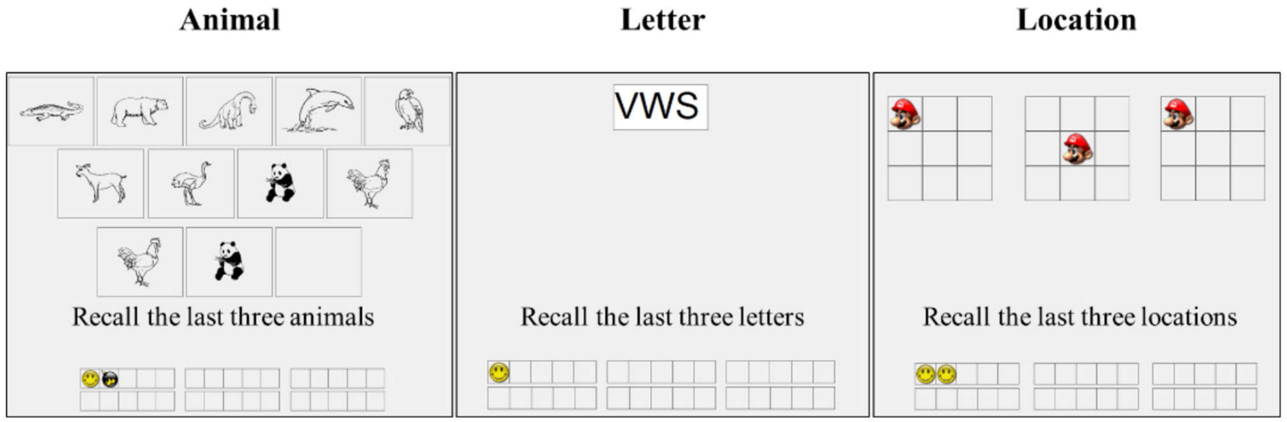


FIGURE 3 | Demonstration of three training tasks.

SE = 0.51, $t_{(48)} = 1.095$, $P = 0.276$, $d = 0.22$]. And there was not an interaction between time and group [$B = -0.48$, SE = 0.72, $t_{(48)} = -0.664$, $P = 0.508$, $d = 0.14$] (refer to **Figure 6**).

Overall, our data showed that the growth mindset group and the fixed group did not differ in training effects, near-transfer effect, and far-transfer effect.

DISCUSSION

Although evidence suggests that placebo effects exist in working memory training, no research has directly examined whether the mindset of intelligence is a critical variable. In this study, we replicated the protocol of the Foroughi et al.'s (2016) study that involved overt recruitment and a single, 1 h session of training. Importantly, we assigned participants to a growth mindset group and a fixed mindset group based on their TIS scores to avoid any confound. We tested for near-transfer effects using the 2-back task and far-transfer effects (fluid intelligence) using RAPM. However, the results showed that compared with the fixed mindset group, the performance of the growth mindset group was not significantly different from the pretest to the post-test in all tasks. It seems to suggest that mindset of intelligence does not contribute to the placebo effect in the 1 h working memory training.

Foroughi et al. (2016) recently attributed different TIS scores between two groups to their own selection of participants for the overt/covert recruitment, subjects who chose the overt flyer had higher expectancies and higher TIS scores. They ignored that the TIS score itself represents the mindset of intelligence that could cause the placebo effect, which means they confused

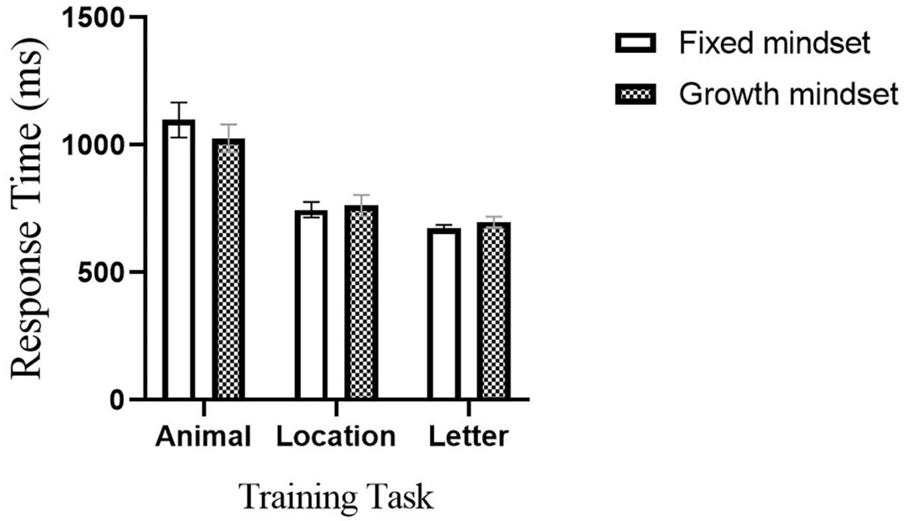


FIGURE 4 | The difference between the fixed and growth groups in working memory training tasks (animal, location, and letter).

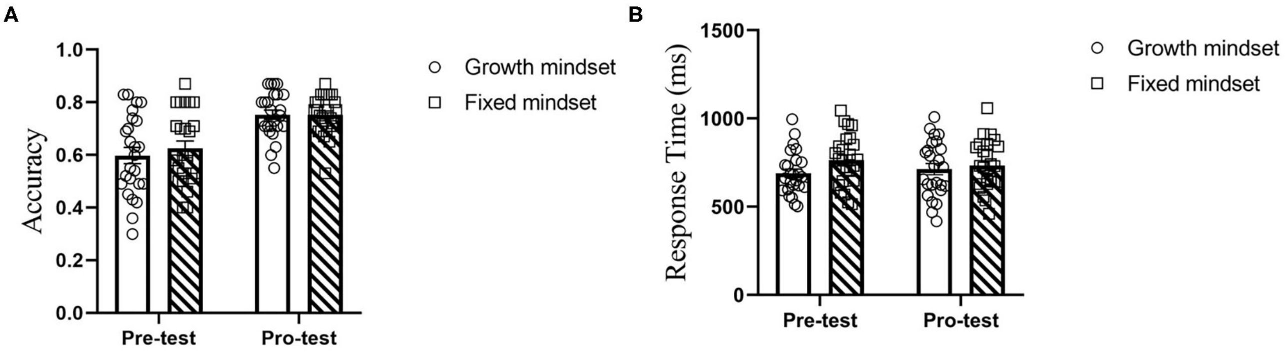


FIGURE 5 | The difference between the growth group and the fixed group in the 2-back task [accuracy **(A)** and response time **(B)**].

the role of the mindset of intelligence and recruitment methods. Another possible explanation of Foroughi et al.'s (2016) study is that the results may be due to differences in the mindset of intelligence between the placebo group and the control group. Therefore, to our knowledge, this study is the first to directly test whether the mindset of intelligence is a contributor to placebo effects. However, the answer is no. It also triggers our deeper thinking about the mindset of intelligence. Compared with western students and ethnic Chinese students growing up in western countries, Chinese students have a more fixed mindset, which may be due to differences in reasoning about intelligence (Kim et al., 2017; Sun et al., 2021). This puts the relevant research into a broader cultural context and also suggests that we should not ignore the potential influence of social context, parenting variables, and educational style. Future studies should examine what variables influence the shaping of the mindset of intelligence of children and how cultural differences in the mindset of intelligence may lead to different outcomes. However, cultural differences are not directly responsible for the discrepancy between our results and those suggested by Foroughi et al. (2016). On the one hand, there was no difference between the TIS scores of both our two groups and their two groups. On the other hand, our finding is consistent with Thompson et al.'s (2013) study (a western study), which found that relevant cognitive factors such as mindset of intelligence have a negligible effect on training results and transfer effects.

We should note that we used a different training paradigm from Foroughi et al.'s (2016). However, this does not affect the results of the experiment. According to the response expectancy view (Kirsch, 1999), when studying placebo effects, setting a training task is only a cue to trigger the expectancy or motivation of participants. The improvement in performance due to the placebo effects should not depend on the training gains; that is why these studies chose an hour as training duration (no actual training gains). Besides, the three working memory training tasks in this study are also commonly used in cognitive training research (Zhao et al., 2011; Wang et al., 2014; Chen et al., 2018). Given this evidence, we believe that the difference shown in our

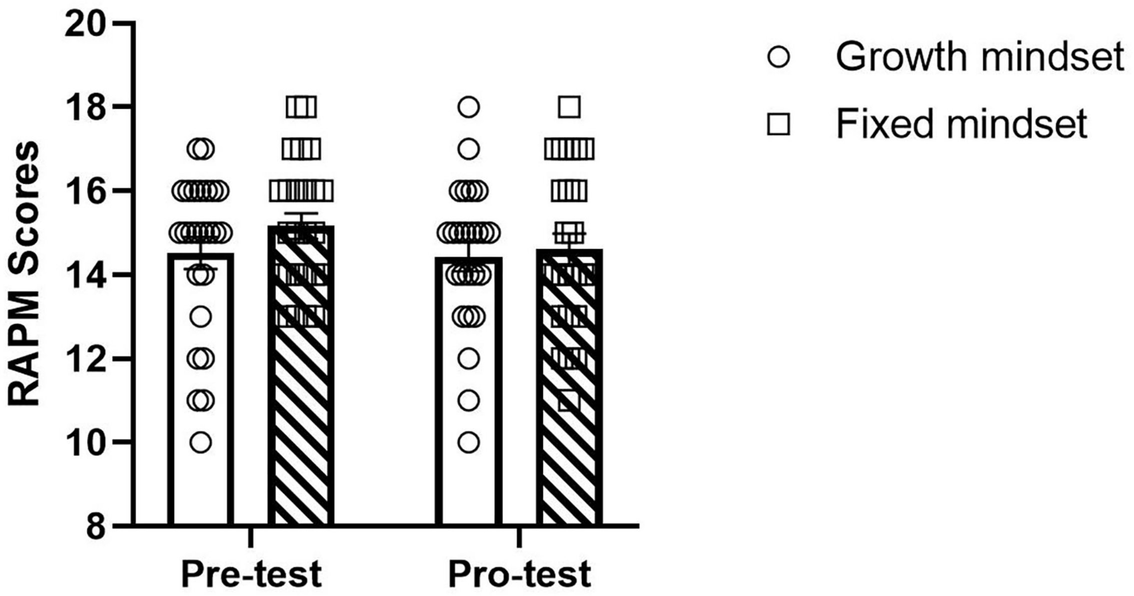


FIGURE 6 | The difference between the fixed group and the growth group in Raven's Advanced Progressive Matrices (RAPM).

research is not due to the different training paradigms. Another point worth noting is that we observed the main effect of time on the accuracy of the 2-back task. In view of the improvements in both groups, this might be a reflection of the practice effect, since the task was relatively easy for college students, and the interval was only an hour.

Overall, we found no evidence that there are placebo effects caused by the mindset of intelligence during working memory training. This study and Zhang et al.'s (2019) study reveal that neither the overt recruitment nor the mindset of intelligence contributes to placebo effects in 1 h working memory training. Two sets of explanations may account for these results.

First, do participants truly believe that 1 h of training can change their fluid intelligence? Just like researchers believe that rigorous and persistent cognitive training rather than 1 h cognitive training is practical, even the high TIS participants may think intelligence is malleable only with massive and long-duration practice; such a short training time is unrealistic. After all, the notion, fluid intelligence is hard to change, is well-known in researchers and laypeople. However, previous studies and this study ignored to examine actual expectancy of intelligence enhancement of participants in an hour, which should be addressed in future research. The goal of measuring the expectancy and motivation of participants in different intervention research stages is to infer the degree of engagement of participants in the training process (Tsai et al., 2018). The subjective report is usually used to evaluate expectancies and motivations. In the research of placebo effects, it is essential to evaluate the expectancy and motivation of participants before, during, and after the intervention. Measuring expectancy is the premise of examining its effect on positive training outcomes. However, it is unclear whether repeat measurement will expose the purpose of the experiment and weaken the expectancy or motivation of participants. Therefore, the appropriate approach to assess the expectancies and motivations needs to be further explored.

Second, can subjective expectancy improve the objective measurement of fluid intelligence? Pratkanis et al. (1994) found the illusory placebo effect: perceptions of personal improvement of participants were consistent with their expectations but inconsistent with objective measures. The illusory placebo effect also existed in working memory training, in which subjects believed that their cognitive abilities, such as intelligence, had been affected by the experiment in the absence of objective evidence (Redick et al., 2013). It challenges the placebo effect in intelligence research, which means IQ as a highly heritable ability (Plomin, 2004; Plomin et al., 2008; Sternberg, 2008) is hardly changed by subjective expectancies. To test this question persuasively, we suggest future cognitive training research to examine the relationship among expectancy, perceptions of change, and objective measurements of participants.

As far as we know, there are few studies on placebo effects in working memory training (Foroughi et al., 2016; Tsai et al., 2018; Zhang et al., 2019). Although these studies have some limitations, such as poor sample size, simple experimental design, and inconsistent results, we aimed to draw academic attention to placebo effects in cognitive training. The factors influencing the placebo effect are complex, and more empirical evidence is needed to promote the progress. Besides, the theoretical framework of placebo effects should be introduced. Previous studies lack theoretical depth in explaining the placebo effect phenomenon in working memory training and do not connect the findings with the broader field of placebo effect research. The main theoretical methods of studying placebo effects can be roughly divided into three views: classical conditioning, expectancy, and motivation (Geers et al., 2005). Interestingly, these three views are currently considered to be conflicting. Among them, the *response expectancy view*, which is widely mentioned in the perspective of expectancy, is a suitable model suggested by Foroughi et al. (2016). According to this theory, response expectancy is an automatic pre-reaction to situational and behavioral cues, a direct self-confirmation of individuals. The placebo effect is the direct and non-intermediary result of expectancy. In the study of Foroughi et al. (2016), the overt recruitment method was used to make the placebo group have stronger expectancy for the training results (intelligence can be improved); the working memory training task is only a cue to induce the placebo effect. If this theory holds, then when we use the same experimental design, no matter what type of working memory training task is used, it should always induce placebo effects. At present, the research on placebo effects in working memory training is not systematic and in-depth. We sincerely suggest that all researchers interested in this field should focus on absorbing beneficial inspiration from the classical theoretical model of placebo effects, which will promote our understanding of this field and further promote the progress of working memory training.

Although we failed to replicate the findings of Foroughi et al. (2016), we share their concern: researchers should pay more attention to the design of cognitive training experiments until substantial studies reveal the role of placebo effects. We suggest that participants should be assigned to one of three groups: training group, active control group, and no-contact control group. It is also necessary to measure the expectancies, subjective perceptions, and objective tasks of subjects. Cognitive abilities, especially fluid intelligence, can be improved is a promising finding for humans. On the one hand, we cannot exaggerate the training efficacy with placebo affecting actual training outcomes; on the other hand, we should not despise the training benefits just for concerns about placebo effects.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Evaluation of Research projects at the Department of Psychology—part of the School for Social and Behavioral Sciences at Nanjing University, China. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XZ collected the data. PL and XZ analyzed the data and wrote the current version of the manuscript. RZ administrated the whole project. All authors contributed to the article and approved the submitted version.

FUNDING

This study was financially supported by the National Defense Science and Technology Innovation Special Zone project of the Logistics Support Department of the Central Military Commission (Grant number 19-163-12-ZT-002-003-02) and the Fundamental Research Funds for the Central Universities (Grant number 2020300048).

ACKNOWLEDGMENTS

The authors would like to express their gratitude for the support of these projects and are also grateful to Mengyu Gao for assistance in the language of this work.

REFERENCES

Artuso, C., Carretti, B., and Palladino, P. (2019). Short-term training on working memory updating and metacognition in primary school: the effect on reading comprehension. *Sch. Psychol. Int.* 40, 641–657. doi: 10.1177/0143034319881671

Baddeley, A. (2010). Working memory. *Curr. Biol.* 20, 136–140. doi: 10.1016/j.cub.2009.12.014

Bastian, C. C. V., and Eschen, A. (2016). Does working memory training have to be adaptive? *Psychol. Res.* 80, 181–194. doi: 10.1007/s00426-015-0655-z

Bastian, C. C. V., and Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. *J. Mem. Lang.* 69, 36–58. doi: 10.1016/j.jml.2013.02.002

Boot, W. R., Simons, D. J., Stothart, C., and Stutts, C. (2013). The pervasive problem with placebos in psychology: why active control groups are not sufficient to rule out placebo effects. *Perspect. Psychol. Sci.* 8, 445–454. doi: 10.1177/1745691613491271

Brody, H. (1980). "Placebos and the philosophy of medicine," in *Clinical, Conceptual, and Ethical Issues*, ed H. Brody (Chicago, IL: University of Chicago Press), 164. doi: 10.1007/978-94-009-8972-6

Carretti, B., Borella, E., Zavagnin, M., and De Beni, R. (2013). Gains in language comprehension relating to working memory training in healthy older adults. *Int. J. Geriatr. Psychiatry* 28, 539–546. doi: 10.1002/gps.3859

Carretti, B., Caldarola, N., Tencati, C., and Cornoldi, C. (2014). Improving reading comprehension in reading and listening settings: the effect of two training programmes focusing on metacognition and working memory. *Br. J. Educ. Psychol.* 84, 194–210. doi: 10.1111/bjep.12022

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: a critical experiment. *J. Educ. Psychol.* 54, 1–22. doi: 10.1037/h0046743

Chein, J. M., and Morrison, A. B. (2010). Expanding the mind's workspace: training and transfer effects with a complex working memory span task. *Psychonomic Bull. Rev.* 17, 193–199. doi: 10.3758/PBR.17.2.193

Chen, X., Ye, M., Chang, L., Chen, W., and Zhou, R. (2018). Effect of working memory updating training on retrieving symptoms of children with learning disabilities. *J. Learn. Disabil.* 51, 507–519. doi: 10.1177/0022219417712015

Chooi, W. T., and Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence* 40, 531–542. doi: 10.1016/j.intell.2012.07.004

Clark, C. M., Lawlorsavage, L., and Goghari, V. M. (2017). Working memory training in healthy young adults: support for the null from a randomized comparison to active and passive control groups. *PLoS ONE.* 12:e177707. doi: 10.1371/journal.pone.0177707

Claro, S., Paunesku, D., and Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. *Proc. Natl. Acad. Sci. U. S. A.* 113, 8664–8668. doi: 10.1073/pnas.1608207113

Colom, R., Quiroga, M. Á., Shih, P. C., Martínez, K., Burgaleta, M., Martínez-Molina, A., et al. (2010). Improvement in working memory is not related to increased intelligence scores. *Intelligence* 38, 497–505. doi: 10.1016/j.intell.2010.06.008

Colom, R., Román, F. J., Abad, F. J., Pei, C. S., Privado, J., Froufe, M., et al. (2013). Adaptive n-back training does not improve fluid intelligence at the construct level: gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence* 41, 712–727. doi: 10.1016/j.intell.2013.09.002

Crum, A., and Zuckerman, B. (2017). Changing mindsets to enhance treatment effectiveness. *JAMA* 317, 2063–2064. doi: 10.1001/jama.2017.4545

Crum, A. J., and Langer, E. J. (2007). Mind-set matters: exercise and the placebo effect. *Psychol. Sci.* 18, 165–171. doi: 10.1111/j.1467-9280.2007.01867.x

Crum, A. J., Salovey, P., and Achor, S. (2013). Rethinking stress: the role of mindsets in determining the stress response. *J. Pers. Soc. Psychol.* 104, 716–733. doi: 10.1037/a0031201

Dommett, E. J., Devonshire, I. M., and Greenfield, S. A. (2013). The impact of participation in a neuroscience course on motivational measures and academic performance. *Trends Neurosci. Educ.* 2, 122–138. doi: 10.1016/j.tine.2013.05.002

Dweck, C. S. (2000). *Self-Theories: Their role in motivation, Personality, and Development*. East Sussex: Psychology Press.

Foroughi, C. K., Monfort, S. S., Paczynski, M., Mcknight, P. E., and Greenwood, P. M. (2016). Placebo effects in cognitive training. *Proc. Natl. Acad. Sci. U. S. A.* 113, 7470–7474. doi: 10.1073/pnas.1601243113

Geers, A. L., Weiland, P. E., Kosbab, K., Landry, S. J., and Helfer, S. G. (2005). Goal activation, expectations, and the placebo effect. *J. Pers. Soc. Psychol.* 89, 143–159. doi: 10.1037/0022-3514.89.2.143

Guye, S., De Simoni, C., and Von Bastian, C. C. (2017). Do individual differences predict change in cognitive training performance? A latent growth curve modeling approach. *J. Cogn. Enhanc.* 1, 374–393. doi: 10.1007/s41465-017-0049-9

Hardy, J. L., Nelson, R. A., Thomason, M. E., Sternberg, D. A., Katovich, K., Farzin, F., et al. (2015). Enhancing cognitive abilities with comprehensive training: a large, online, randomized, active-controlled trial. *PLoS ONE.* 10:e134467. doi: 10.1371/journal.pone.0134467

Holmes, J., Gathercole, S. E., and Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Dev. Sci.* 12, F9–F15. doi: 10.1111/j.1467-7687.2009.00848.x

Jaeggi, S. M., Buschkuehl, M., Jonides, J., and Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proc. Natl. Acad. Sci. U. S. A.* 105, 6829–6833. doi: 10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Jonides, J., and Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10081–10086. doi: 10.1073/pnas.1103228108

Jaeggi, S. M., Buschkuehl, M., Shah, P., and Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Mem. Cogn.* 42, 464–480. doi: 10.3758/s13421-013-0364-z

Kane, M. J., and Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differ- ences perspective. *Psychonomic Bull. Rev.* 9, 637–671. doi: 10.3758/BF03196323

Kim, J. J., Fung, J., Wu, Q., and Lau, A. S., (2017). Parenting variables associated with growth mindset: an examination of three Chinese-heritage samples. *Asian Am. J. Psychol.* 8, 115–125. doi: 10.1037/aap0000064

Kirsch, I. (1999). *How Expectancies Shape Experience*. Washington, DC: American Psychological Association. doi: 10.1037/10332-000

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., et al. (2005). Computerized training of working memory in children with ADHD - a randomized, controlled trial. *J. Am. Acad. Child Adolesc. Psychiatry* 44, 177–186. doi: 10.1097/00004583-200502000-00010

Klingberg, T., Forssberg, H., and Westerberg, H. (2002). Training of working memory in children with ADHD. *J. Clin. Exp. Neuropsychol.* 24, 781–791. doi: 10.1076/jcen.24.6.781.8395

Kundu, B., Sutterer, D. W., Emrich, S. M., and Postle, B. R. (2013). Strengthened effective connectivity underlies transfer of working memory training to tests of short-term memory and attention. *J. Neurosci.* 33, 8705–8715. doi: 10.1523/JNEUROSCI.5565-12.2013

Lawlor-Savage, L., and Goghari, V. M. (2016). Dual N-back working memory training in healthy adults: a randomized comparison to processing speed training. *PLoS ONE.* 11:e151817. doi: 10.1371/journal.pone.0151817

Melby-Lervåg, M., and Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Dev. Psychol.* 49, 270–291. doi: 10.1037/a0028228

Ommundsen, Y., Haugen, R., and Lund, T. (2005). Academic self-concept, implicit theories of ability, and self-regulation strategies. *Scand. J. Educ. Res.* 49, 461–474. doi: 10.1080/00313830500267830

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., et al. (2010). Putting brain training to the test. *Nature* 465, 775–778. doi: 10.1038/nature09042

Plomin, R. (2004). *Nature and Nurture: An Introduction to Human Behavioral Genetics*. Thousand Oaks, CA: Wadsworth.

Plomin, R., DeFries, J. C., McClearn, G. E., and McGuffin, P. (2008). *Behavioral Genetics, 5th Edn*. New York, NY: Worth.

Pratkanis, A. R., Eskenazi, J., and Greenwald, A. G. (1994). What you expect is what you believe (but not necessarily what you get): a test of the effectiveness of subliminal self-help audiotapes. *Basic Appl Soc Psychol.* 15, 251–276. doi: 10.1207/s15324834basp1503_3

Raven, J., Raven, J. C., and Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. New York, NY: Psychological Corporation.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., et al. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *J. Exp. Psychol.* 142, 359–379. doi: 10.1037/a0029082

Richard, R. W., and Pals, J. L. (2010). Implicit self-theories in the academic domain: implications for goal orientation, attributions, affect, and self-esteem change. *Self Ident.* 1, 313–336. doi: 10.1080/15298860290106805

Shahar, N., Pereg, M., Teodorescu, A. R., Moran, R., Karmon-Presser, A., and Meiran, N. (2018). Formation of abstract task representations: exploring dosage and mechanisms of working memory training effects. *Cognition* 181, 151–159. doi: 10.1016/j.cognition.2018.08.007

Shipstead, Z., Redick, T. S., and Engle, R. W. (2012). Is working memory training effective? *Psychol. Bull.* 138, 628–654. doi: 10.1037/a0027473

Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L., and Macnamara, B. N. (2018). To what extent and under which circumstances are growth mind-sets important to academic achievement? Two meta-analyses. *Psychol. Sci.* 29, 549–571. doi: 10.1177/0956797617739704

Slagter, H. A. (2012). Conventional working memory training may not improve intelligence. *Trends Cogn. Sci.* 16, 582–583. doi: 10.1016/j.tics.2012.10.001

Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J. I., Novick, J. M., Chrabaszcz, J. S., et al. (2013). Training working memory: limits of transfer. *Intelligence* 41, 638–663. doi: 10.1016/j.intell.2013.07.013

Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proc. Natl. Acad. Sci. U. S. A.* 105, 6791–6792. doi: 10.1073/pnas.0803396105

Sternberg, R. J., and Gastel, J. (1989). Coping with novelty in human intelligence: an empirical investigation. *Intelligence* 13, 187–197. doi: 10.1016/0160-2896(89)90016-0

Stewart-Williams, S., and Podd, J. (2004). The placebo effect: dissolving the expectancy versus conditioning debate. *Psychol. Bull.* 130, 324–340. doi: 10.1037/0033-2909.130.2.324

Sun, X., Nancekivell, S., and Shah, P. (2021). Growth mindset and academic outcomes: a comparison of US and Chinese students. *NPJ Sci. Learn.* 6, 21–21. doi: 10.1038/s41539-021-00100-z

Tempelaar, D. T., Rienties, B., Giesbers, B., and Gijselaers, W. H. (2015). The pivotal role of effort beliefs in mediating implicit theories of intelligence and achievement goals and academic motivations. *Soc. Psychol. Educ.* 18, 101–120. doi: 10.1007/s11218-014-9281-7

Thompson, T. W., Waskom, M. L., Garel, K. L., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., et al. (2013). Failure of working memory training to enhance cognition or intelligence. *PLoS ONE.* 8:e63614. doi: 10.1371/journal.pone.0063614

Tsai, N., Buschkuehl, M., Kamarsu, S., Shah, P., Jonides, J., and Jaeggi, S. M. (2018). (Un)Great expectations: the role of placebo effects in cognitive training. *J. Appl. Res. Mem. Cogn.* 7, 564–573. doi: 10.1016/j.jarmac.2018.06.001

Wang, Z., Zhou, R., and Shah, P. (2014). Spaced cognitive training promotes training transfer. *Front. Hum. Neurosci.* 8:217. doi: 10.3389/fnhum.0.2014.00217

Zhang, W., Yuan, L., Lu,. Y., and Luo, J. (2011). The history and development of experiment design for placebo effect studies. *Adv. Psychol. Sci.* 19, 1115–1125. doi: 10.3724/SP.J.1042.2011.01115

Zhang, X., Song, J., Ying, R., and Zhou, R. (2019). Is effectiveness of working memory training the result of placebo effect? *Stud. Psychol. Behav.* 17, 170–177.

Zhao, X., Wang, Y. X., Liu, D. W., and Zhou, R. L. (2011). Effect of updating training on fluid intelligence in children. *Chin. Sci. Bull.* 56, 2202–2205. doi: 10.1007/s11434-011-4553-5

Zion, S. R., and Crum, A. J. (2018). Mindsets matter: a new framework for harnessing the placebo effect in modern medicine. *International Rev. Neurobiol.* 138, 137–160. doi: 10.1016/bs.irn.2018.02.002

Conflict of Interest: XZ was employed by the company TCL Industries Holdings Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 19 November 2021
doi: 10.3389/fnins.2021.704880

Effects of High-Definition Transcranial Direct Current Stimulation Over the Left Fusiform Face Area on Face View Discrimination Depend on the Individual Baseline Performance

Di Wu[1†], *Pan Zhang*[2†], *Na Liu*[3], *Kewei Sun*[1] *and Wei Xiao*[1*]

[1] *Department of Medical Psychology, Air Force Medical University, Xi'an, China,* [2] *Department of Psychology, Hebei Normal University, Shijiazhuang, China,* [3] *Department of Nursing, Air Force Medical University, Xi'an, China*

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Joachim Hass,
University of Heidelberg, Germany
N. Apurva Ratan Murty,
Massachusetts Institute
of Technology, United States

***Correspondence:**
Wei Xiao
xiaowei7304@outlook.com

[†] *These authors have contributed equally to this work and share first authorship*

Specialty section:
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

Received: 04 May 2021
Accepted: 27 October 2021
Published: 19 November 2021

Citation:
Wu D, Zhang P, Liu N, Sun K and Xiao W (2021) Effects of High-Definition Transcranial Direct Current Stimulation Over the Left Fusiform Face Area on Face View Discrimination Depend on the Individual Baseline Performance. Front. Neurosci. 15:704880. doi: 10.3389/fnins.2021.704880

A basic human visual function is to identify objects from different viewpoints. Typically, the ability to discriminate face views based on in-depth orientation is necessary in daily life. Early neuroimaging studies have identified the involvement of the left fusiform face area (FFA) and the left superior temporal sulcus (STS) in face view discrimination. However, many studies have documented the important role of the right FFA in face processing. Thus, there remains controversy over whether one specific region or all of them are involved in discriminating face views. Thus, this research examined the influence of high-definition transcranial direct current stimulation (HD-tDCS) over the left FFA, left STS or right FFA on face view discrimination in three experiments. In experiment 1, eighteen subjects performed a face view discrimination task before and immediately, 10 min and 20 min after anodal, cathodal and sham HD-tDCS (20 min, 1.5 mA) over the left FFA in three sessions. Compared with sham stimulation, anodal and cathodal stimulation had no effects that were detected at the group level. However, the analyses at the individual level showed that the baseline performance negatively correlated with the degree of change after anodal tDCS, suggesting a dependence of the change amount on the initial performance. Specifically, tDCS decreased performance in the subjects with better baseline performance but increased performance in those with poorer baseline performance. In experiments 2 and 3, the same experimental protocol was used except that the stimulation site was the left STS or right FFA, respectively. Neither anodal nor cathodal tDCS over the left STS or right FFA influenced face view discrimination in group- or individual-level analyses. These results not only indicated the importance of the left FFA in face view discrimination but also demonstrated that individual initial performance should be taken into consideration in future research and practical applications.

Keywords: high-definition transcranial direct current stimulation (HD-tDCS), fusiform face area (FFA), superior temporal sulcus (STS), face view discrimination, initial performance

INTRODUCTION

As a kind of visual stimulus or complex object, the face is important to the survival and social communication of various species, including humans. It is a remarkable property for the primate visual system to recognize faces across different viewpoints in invariant views (Axelrod and Yovel, 2012). Humans, for example, can accurately recognize a face despite changes in viewpoint. In real life, we frequently need to recognize a person's face from various angles. Given that face view processing is a basic ability that people possess, methods that facilitate this kind of face processing could be valuable and attractive due to their importance.

To our knowledge, there could be two important ways to improve or recover functions related to face processing. One is perceptual learning, which refers to a phenomenon in which extensive practice of a perceptual task can boost various perceptual functions (Lu et al., 2011, 2016; Sasaki et al., 2012). Studies have confirmed that face perception can be improved through perception learning (Hussain et al., 2009; Bi et al., 2010; Mcmahon and Leopold, 2012). However, this method is time consuming because it generally needs hundreds or thousands of practice trials over days to weeks to improve. Another method is transcranial direct current stimulation (tDCS), a non-invasive brain stimulation technique that is attracting increasing attention because of its low cost, portability and feasibility (Reinhart et al., 2016; Turski et al., 2017). It has been found that tDCS not only directly boosts perceptual performance (Ding et al., 2016; Reinhart et al., 2016; Wu et al., 2020) but also facilitates perceptual learning, which produces more benefits, such as a reduction in training time, a larger magnitude of improvements and more enduring improvements (Bolognini et al., 2010; Sczesny-Kaiser et al., 2016). More importantly, tDCS contributes to exploring the causality between a certain cortical area and its corresponding functions.

Transcranial direct current stimulation transiently modulates cortical excitability by altering the membrane potential of neurons (Stagg and Nitsche, 2011; Stagg et al., 2011). The technique delivers a mild direct current (DC) between anode and cathode electrodes that are placed on the scalp of a participant. The mild intracerebral DC enters the cortex from the anode and exits the cortex to the cathode. Generally, tDCS effects are bidirectional based on the different directions in the current flow: the anodal electrode increases cortical excitability, and the cathodal electrode decreases excitability (Horvath et al., 2015; Parkin et al., 2015; Woods et al., 2016). The identification of relevant stimulus sites on the scalp is an important question to consider in tDCS studies. There is considerable evidence that humans have specific neural mechanisms for face processing (Kanwisher and Yovel, 2006). It is generally acknowledged that the "core system" for face processing currently includes the fusiform face area (FFA), the inferior occipital gyrus (occipital face area, OFA) and the superior temporal sulcus (STS; Haxby et al., 2000; Gobbini and Haxby, 2007; Fox et al., 2009). Specifically, the FFA is primarily engaged in the perceived identity of the face, whereas the OFA is apparently dedicated to the physical properties of the face stimulus (Rotshtein et al., 2005; Pitcher et al., 2007). STS is related to dynamic aspects of faces, such as their emotional expressions, gazes and viewpoints (Andrews and Ewbank, 2004).

Previous studies may have provided some insights into the neural mechanisms underlying face view discrimination. Studies involving monkeys found face view-selective neuron clustering in the inferior temporal cortex (IT) and STS (Perrett et al., 1985, 1991; De Souza et al., 2005). Early neuroimaging research in humans revealed a greater response in the STS when a face with the same identity was presented from different viewpoints (Andrews and Ewbank, 2004). In contrast, an investigation in terms of perceptual learning of face views demonstrated a close relationship between the left FFA and face view discrimination learning. Specifically, the behavioral learning effects were closely related to improved left FFA stability. Additionally, the pretraining thickness of the left FFA could predict individual behavioral learning effects (Bi et al., 2014). Similarly, an event-related potential (ERP) study also found that the trained face view rather than the untrained view significantly reduced the N170 latency in the left occipital–temporal area (Su et al., 2012). There outcomes in monkeys and humans seem to build close connections between the left cerebral hemisphere and face view discrimination. However, the essential role of the right FFA in face recognition has been extensively documented in literature over the past few decades (Kanwisher et al., 1997). There is a bulk of literature showing a larger response to faces in the right hemisphere than in the left hemisphere (Pinsk et al., 2005; Fang et al., 2007). Thus, the specific cortical region remains controversial.

Recently, tDCS research has found that the results obtained using only the group mean may mask some notable findings, suggesting that analyses of interindividual differences are necessary. For example, previous studies found that neither anodal nor cathodal tDCS over the left dorsolateral prefrontal cortex affected response inhibition measured in a go/no-go task at the group level unless interindividual differences in genetic polymorphisms (Plewnia et al., 2013; Nieratschker et al., 2015) or personality traits (Weidacker et al., 2016) were taken into consideration. Indeed, tDCS data frequently involve high levels of variability across participants, and often, there are some people who show little improvement or even opposite effects after stimulation (López-Alonso et al., 2014; Benwell et al., 2015; Li et al., 2015). Among these interindividual factors that influence tDCS effects, individuals' initial performance is worthy of attention (Li et al., 2015; Wu et al., 2020). The initial performance can be considered to be related to baseline brain excitability levels that may subsequently determine the stimulation effects on performance (London and Slagter, 2021). In a visual perception study, for example, only anodal tDCS over the primary visual cortex modulated the magnitude of change in contrast sensitivity as a function of individual baseline contrast sensitivity even though both anodal and cathodal tDCS did not influence contrast sensitivity at the group level (Wu et al., 2021). Thus, the current study not only focused on the tDCS effect on face view discrimination at the group level but also further analyzed the individual tDCS effect regarding the correlation between baseline performance and the magnitude of performance change.

Transcranial direct current stimulation is capable of modulating the excitation and inhibition of a certain brain region; therefore, it provides an effective way to tell us about the roles the different cortical areas play in processing face views at different angles. Conventional tDCS typically delivers electrical current in a relatively non-focal manner using a pair of electrodes placed on the scalp (1×1 electrode configuration; Lang et al., 2019). Recently, multielectrode configurations (referred to as high definition; HD) with individual control of current intensity at each electrode allow for unique combinations of electrode locations combined with current optimization algorithms to more focally target brain regions (Dmochowski et al., 2011).

In summary, this research aimed to examine the involvement of the left FFA, left STS and right FFA in face view discrimination using HD-tDCS. In experiment 1, face view discrimination was measured before and after anodal, cathodal and sham tDCS over the left FFA that was counterbalanced across three sessions. In experiments 2 and 3, the target brain region was the left STS or right FFA, respectively. HD-Targets software (Soterix Medical Inc., New York, United States) was employed to define the optimal electrode positions to focally stimulate the left FFA, left STS or right FFA. The modulation of face view discrimination by tDCS was analyzed at the group and individual levels.

EXPERIMENT 1

Materials and Methods

Subjects

Eighteen male subjects (mean age: 20.2 ±0.6 years) had normal or corrected-to-normal vision. Each subject signed a written informed consent form before participating, and they were all naive to the objective of this study. In particular, they were informed that we would apply mild DC on their scalp and that they needed to complete a face-related task four times before and after the application of the DC. None of them had previously participated in tDCS-related experiments. The research received approval from the local Research Ethics Committee and adhered to the principles of the Declaration of Helsinki.

Stimuli

FaceGen Modeler 3.5 was used to generate 3D face images at various in-depth rotation angles. The stimuli extended 2° × 2° of the visual angle. One block of face view discrimination included 100 trials and lasted approximately 5 min. Each subject was required to discriminate face views around the in-depth face orientation that was 30° tilted to the right. In a trial (**Figure 1A**), a 200-ms fixation was presented in the center of the screen followed by a 100-ms blank interval. Two face stimuli (30° and $30 \pm \theta°$ face views) were randomly presented in two 200-ms temporal intervals, separated by a 1400-ms blank. A brief tone appeared at the beginning of each interval. During each interval, the face stimuli were displayed in the last 100 ms. The subjects needed to make a two-alternative forced-choice (2-AFC) judgment of the orientation of the second face view relative to the first view (left or right). The step size was 0.2°, with both left and right rotations. They pressed the left button when the second face turned left relative to the first one and pressed the right button otherwise. A brief tone appeared following each response independent of its accuracy, and the next trial began 1000 ms after response.

The θ varied trial by trial and was controlled by an adaptive three-down one-up staircase method to assess subjects' face view discrimination thresholds that converged to a performance of 79.3% correct. The threshold of face view discrimination was estimated by one block of 100 trials. We recorded a reversal when the direction of the staircase changed from increasing to decreasing θ or vice versa. We deleted the first four (if the total number of reversals was even) or five (if odd) reversals. The threshold for the discrimination of the in-depth orientation of the face view was calculated by averaging the remaining reversals. The starting threshold for each staircase was set near the expected threshold based on pilot testing.

The face images were presented by a computer running MATLAB and PsychToolbox extensions. A gamma-corrected 60×34 cm monitor was used to display the face stimuli, with a spatial resolution of 1920×1080 pixels and a refresh rate of 85 Hz. The subjects viewed the displays binocularly, and their heads were placed on a chin rest to maintain stabilization. The display subtended 6.84° ×3.89° at a 5-m viewing distance. Normal vision was ensured for some subjects through optical correction.

Experimental Procedure

The study used a single-blind, sham-controlled within-subject design. All subjects took part in three sessions (anodal, cathodal and sham) with the sequence counterbalanced across subjects. The time interval between each session was at least 48 h to limit potential carryover effects. The threshold of face view discrimination was separately measured four times: before and immediately, 10 min and 20 min after tDCS (**Figure 1B**). The subjects rested during the stimulation and between the block intervals. After completing all experimental procedures, the subjects were asked to report scalp pain and uncomfortable experiences, and they could not distinguish between active and sham tDCS.

High-Definition Transcranial Direct Current Stimulation

HD-Targets software, with a finite-element model of a template adult brain to assess the current distribution, was used to confirm the stimulation sites. This software has been proven to have good effectiveness in previous studies (Nikolin et al., 2015; Hartmann et al., 2019; Lang et al., 2019). Electrode positions were selected to generate the highest current focality to the left FFA. Based on the optimized current modeling, electrodes were placed at P9, CP5, P3, AF7, and FT10 (**Figure 2A**). In the anodal stimulation, P9 served as the anode, delivering an intensity of 1.5 mA DC for 20 min (fade in/out: 30 s). The remaining electrodes receiving the return current were as follows: CP5 = -1.04 mA, P3 = -0.15 mA, AF7 = -0.18 mA and FT10 = -0.13 mA. Conductive gel was used to increase conductivity and reduce impedance. In the cathodal stimulation, the polarity of all electrodes was reversed. Electrode positions in sham conditions were counterbalanced

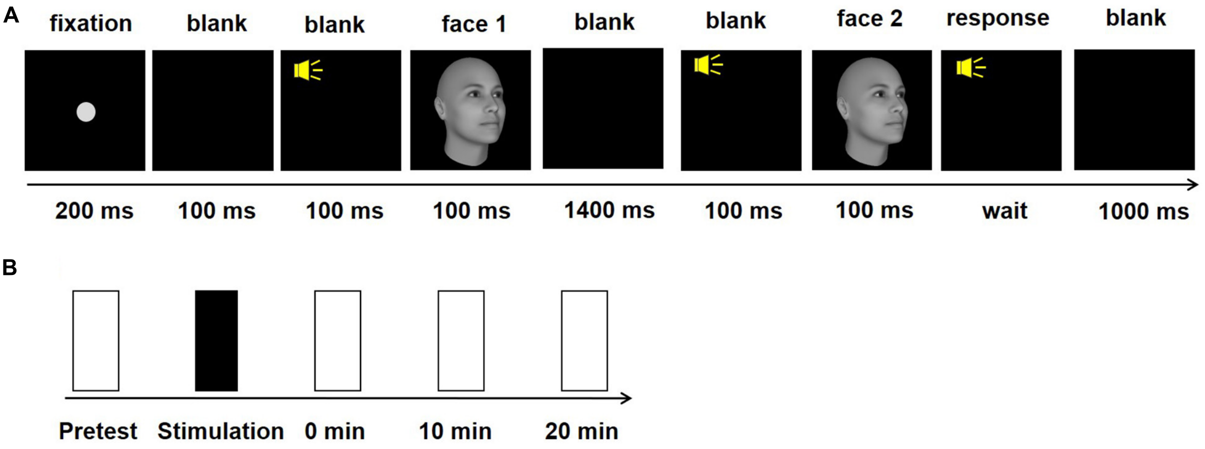


FIGURE 1 | Task and procedure in experiment 1. **(A)** Schematic description of a trial in the 2-AFC face view discrimination task. **(B)** The experimental procedure in one session. The black rectangle indicates anodal, cathodal or sham stimulation. The white rectangles indicate the four test blocks.

such that the positions corresponding to anodal and cathodal tDCS occurred at equal times. The sham current lasted only 30 s, ramping up at the beginning and down at the end of the 20-min period. **Figure 2B** depicts the calculated current flow of anodal tDCS using HD-Explore software (Soterix Medical Inc., New York, United States).

Data Analyses

SPSS statistical software was used to finish the data analyses. We conducted a two-way ANOVA on the face view discrimination threshold, with groups (anodal, cathodal and sham) and blocks (before and immediately, 10 min and 20 min after) as within-subjects factors to analyze the tDCS effect at the group level. *Post hoc* tests were performed to compare the threshold differences with Bonferroni-corrected *p*.

For the individual analyses, correlation analyses were conducted between an individual's initial performance and the magnitude of the performance change separately for anodal, cathodal and sham conditions. Furthermore, the analyses of covariance (ANCOVAs) were used to compare the slopes of the two linear models for the anodal/cathodal vs. sham models to exclude the regression effect. The above statistical analyses were performed individually for each of the different blocks.

Results

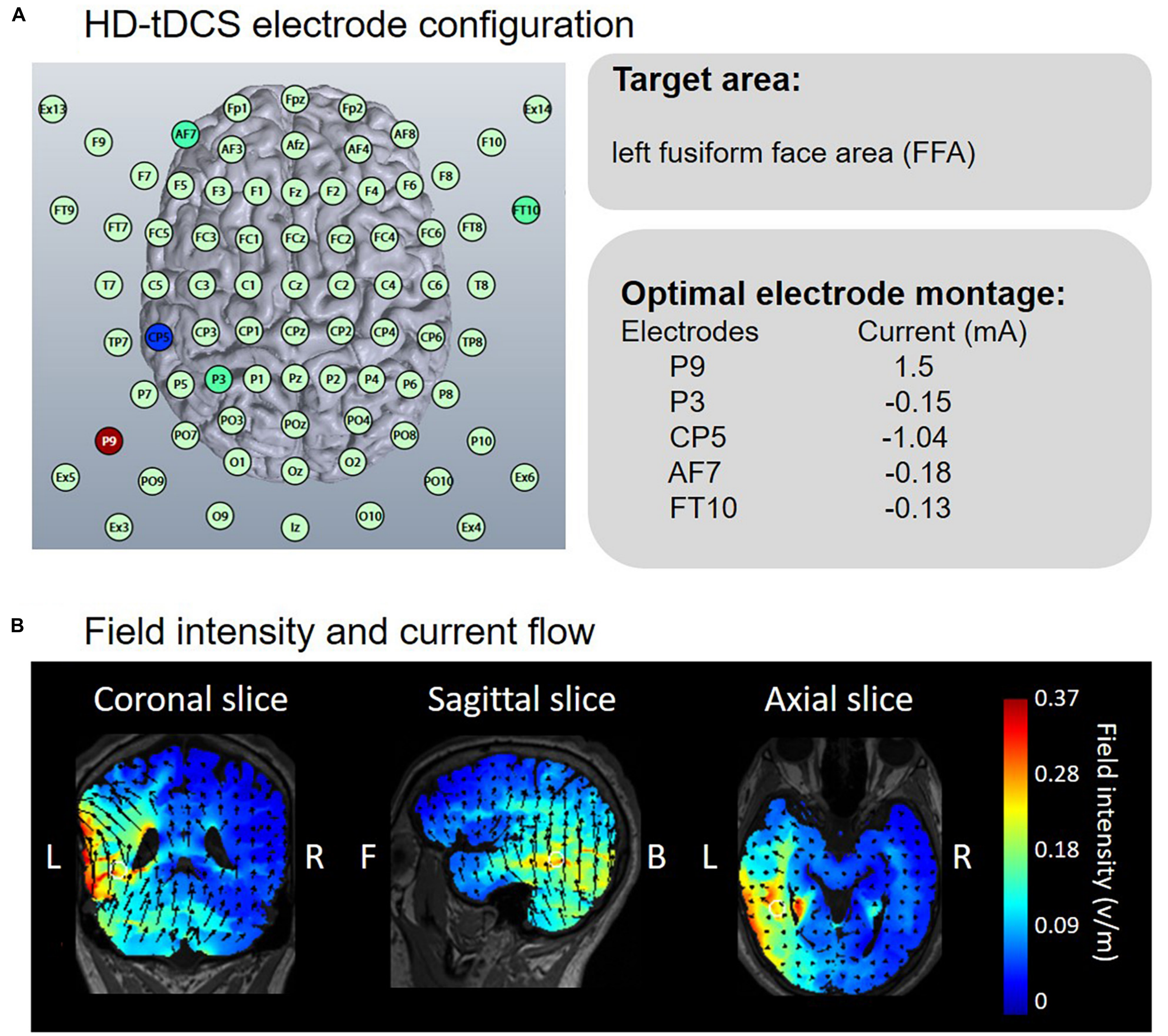
Group Analyses

Three groups and four blocks of two-way ANOVA on the face view discrimination threshold were conducted. **Figure 3A** shows a significant main effect of block, $F(3,51) = 5.76$, $p = 0.002$, $\eta^2 = 0.07$. Additionally, no significance was found for other effects, $Fs < 1$. *Post hoc* tests (with Bonferroni-corrected $p = 0.017$) showed a significantly greater threshold before stimulation than immediately after ($p < 0.001$), 10 min after ($p < 0.001$) and 20 min after ($p = 0.001$) tDCS. There were no significant differences in the thresholds for the three blocks after tDCS ($ps > 0.1$). The possible reason for the significant main effect of block was the practice effect. Theoretically, the practice effect should exist between the first and second tests, and disappear during the following tests. Indeed, we only found a difference in threshold between the first two blocks, and there was no change in the subsequent three blocks of tests. The results indicated no modulation effect of either anodal or cathodal tDCS over the left FFA on the threshold of face view discrimination at the group level.

Individual Analyses

We correlated an individual's initial threshold with the magnitude of the threshold change separately for the anodal, cathodal and sham conditions while controlling for the session order. Regarding the results immediately after tDCS, we found a significant correlation only in anodal tDCS with Bonferroni correction ($r = 0.70$, $p = 0.002$; corrected $p = 0.017$). There were no significant correlations detected in the cathodal ($r = 0.47$, $p = 0.059$) and sham conditions ($r = 0.29$, $p = 0.263$). Nevertheless, the two tests before and after stimulation may have led to a regression effect, a phenomenon in which a variable that is extreme on its first measurement will tend to be closer to the center of the distribution on later measurements. Thus, the significant correlation in anodal stimulation may have resulted from a regression effect instead of the stimulation effect. Here, the results showed a significant correlation only between the initial threshold and the change amount in the anodal condition and not the cathodal and sham conditions, suggesting the existence of a stimulation effect rather than a regression effect.

Comparisons of the slopes in the anodal (or cathodal) vs. sham linear models were conducted to further exclude the regression effect. Specifically, the best-fitting regression lines were estimated with the initial threshold and threshold change through the least square method in the anodal ($r^2 = 0.53$, $p < 0.001$), cathodal ($r^2 = 0.15$, $p > 0.10$) and sham ($r^2 = 0.05$, $p > 0.10$) conditions. ANCOVAs with groups as a fixed factor and the threshold before stimulation as a covariate were performed to compare the slopes of the two models for the



Electrodes	Current (mA)
P9	1.5
P3	-0.15
CP5	-1.04
AF7	-0.18
FT10	-0.13

FIGURE 2 | Electrode configuration and simulated electric field in anodal HD-tDCS over the left FFA. **(A)** HD-Targets software defined the optimal electrode montage to focally stimulate the left FFA. **(B)** HD-Explore software modeled the field intensity and current flow.

anodal/cathodal vs. sham models. Here, ANCOVA was formally equivalent to a moderation analysis in which the initial threshold and groups separately served as continuous and categorial independent variables. The regression effect was supported if the two models (anodal/cathodal vs. sham) were parallel; in contrast, the stimulation effect was supported if the two models were non-parallel. We detected marginally significantly different slopes of the anodal vs. sham linear models, $F(1,32) = 3.05$, $p = 0.090$, $\eta^2 = 0.07$. However, the slope differences between the cathodal and sham linear models were not significant, $F < 1$. These results demonstrated that anodal tDCS over the left FFA can modulate the threshold of face view discrimination in comparison with sham tDCS.

Furthermore, we analyzed the correlation between individual initial thresholds and the magnitude of threshold change 10 min and 20 min after tDCS using the same analytical methods. For the 10 min after the stimulation time point (**Figure 3C**), no significant partial correlation coefficients were found with Bonferroni correction (corrected $p = 0.017$) in the anodal ($r = 0.45$, $p = 0.070$), cathodal ($r = 0.35$, $p = 0.176$) and sham ($r = 0.26$, $p = 0.308$) conditions. ANCOVAs showed no significant difference in the slope between the anodal and sham models, $F(1,32) = 1.45$, $p = 0.237$, $\eta^2 = 0.04$, or between the cathodal and sham models, $F < 1$. Similarly, for the 20 min after the stimulation time point (**Figure 3D**), there were no significant partial correlations with Bonferroni correction in the anodal

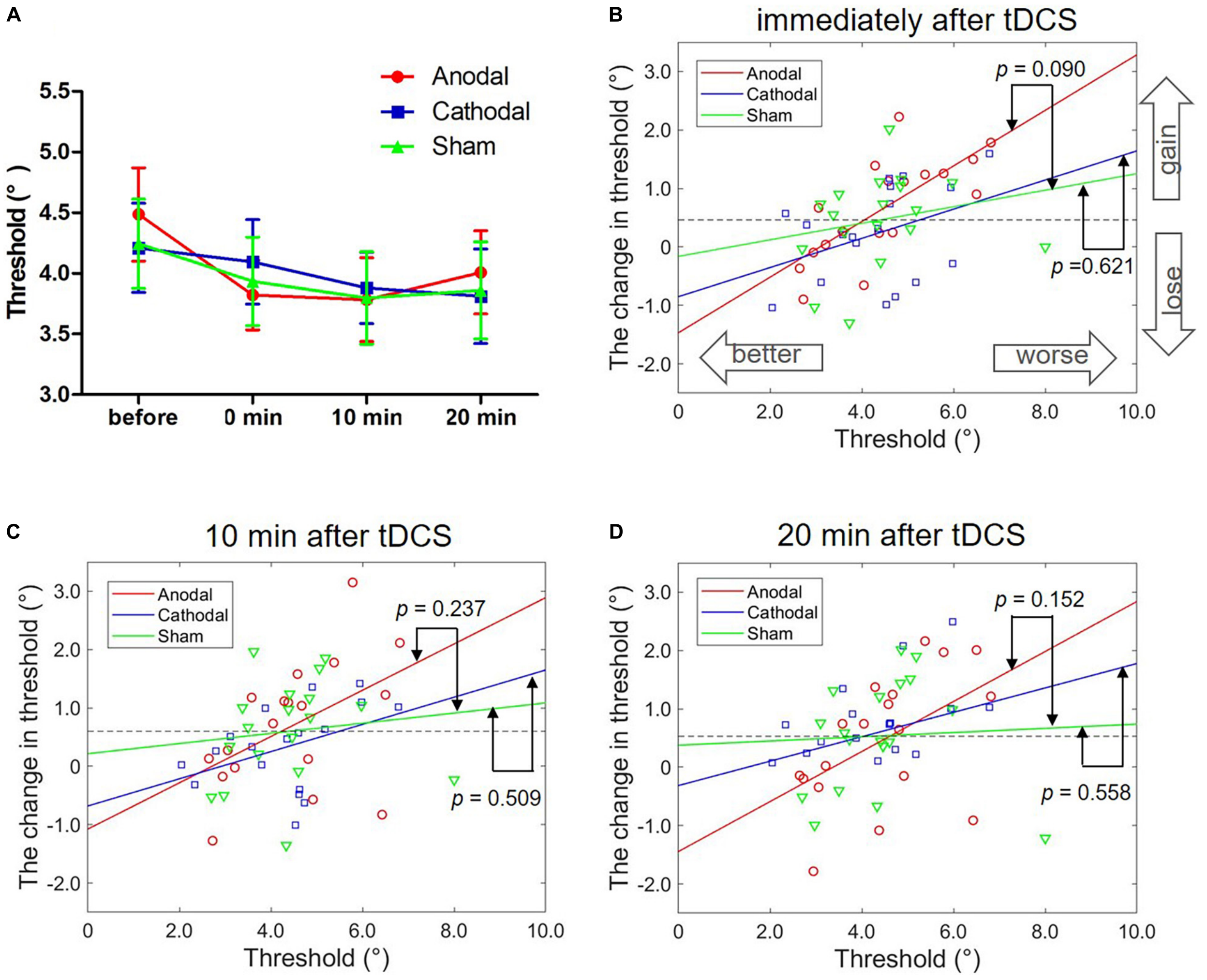


FIGURE 3 | Effects of tDCS over the left FFA on the threshold of face view discrimination. **(A)** The average thresholds are depicted at four times after anodal (red circles), cathodal (blue squares) and sham tDCS (green triangles). Error bars indicate standard errors (SE). **(B–D)** The correlation between the baseline threshold and the change amount in the threshold at different time points. *p*-values represent the significance level between the slopes of two linear models (anodal or cathodal vs. sham).

($r = 0.50$, $p = 0.042$), cathodal ($r = 0.07$, $p = 0.788$) and sham ($r = 0.31$, $p = 0.232$) conditions, and no significant difference in slope between the anodal and sham models, $F(1,32) = 2.16$, $p = 0.152$, $\eta^2 = 0.06$, or between the cathodal and sham models, $F < 1$. Although the results did not reach significance, we still observed a tendency for the slopes with the anodal vs. sham models to be larger than the slopes with the cathodal vs. sham models.

As shown in **Figure 3B**, data points in the anodal tDCS were distributed on both sides around the dashed line (averaged threshold change in the sham tDCS, 0.46°). For the better initial performers, the threshold changes following anodal tDCS went below the averaged change in sham stimulation and gradually increased with the reduction in the initial threshold. For the poorer initial performers, the threshold changes went above the mean change and improved with increases in the initial threshold. Together, anodal tDCS over the left FFA had diverse effects on face view discrimination dependent on different initial performances, which when combined, offset the group-level effect of tDCS.

EXPERIMENT 2

As mentioned above, the specific cortical region (left FFA, left STS or right FFA) involving face viewing discrimination remains controversial. Experiment 1 provided evidence of the effect of anodal tDCS over the left FFA on face viewing discrimination. Thus, experiment 2 was conducted to investigate the role of the left STS in discrimination.

Eighteen male subjects (mean age: 20.5 ± 0.4 years) with normal or corrected-to-normal vision participated in this experiment. All the experimental procedures were the same as in experiment 1. There were two changes in experiment 2:

the target brain region was the left STS (**Figure 4**); and the threshold of face view discrimination was recorded before and immediately after tDCS.

Results

Group Analyses

Similar to experiment 1, two-way ANOVA showed a significant main effect of block, $F(1,17) = 24.49$, $p < 0.001$, $\eta^2 = 0.15$. However, there was no obvious main effect of group, $F(2,34) = 1.69$, $p = 0.200$, $\eta^2 = 0.03$, or interaction effect, $F < 1$ (**Figure 5A**). The non-significant interaction effect indicated that both anodal and cathodal tDCS over the left STS did not influence the threshold of face view discrimination at the group level in comparison to sham stimulation.

Individual Analyses

Partial correlation analyses between the initial thresholds and the change amounts in the threshold were conducted in anodal, cathodal and sham tDCS (**Figure 5B**). No significant correlations were observed with Bonferroni correction in the anodal ($r = 0.42$, $p = 0.091$), cathodal ($r = 0.34$, $p = 0.187$) and sham ($r = 0.52$, $p = 0.034$) conditions. Furthermore, we estimated the best-fitting lines with the initial threshold and the threshold change amount in the anodal ($r^2 = 0.27$, $p < 0.05$), cathodal ($r^2 = 0.11$, $p > 0.10$) and sham ($r^2 = 0.26$, $p < 0.05$) conditions. ANCOVAs showed no significant differences in the slope of the linear models between the anodal (cathodal) and sham groups, $Fs < 1$. The analyses of individual differences further indicated that neither anodal nor cathodal tDCS over the left STS influenced face view discrimination.

EXPERIMENT 3

Experiments 1 and 2 revealed that the left FFA, rather than the left STS, was related to face view discrimination. Some studies have demonstrated a larger response to faces in the right hemisphere than in the left hemisphere (Pinsk et al., 2005; Fang et al., 2007). Thus, the third experiment further investigated how tDCS over the right FFA influences face view discrimination.

Twenty male subjects (mean 19.6 ± 0.6 years) with normal or corrected-to-normal vision took part in this experiment. The objective of this experiment was to further investigate the role of right FFA in face view discrimination using HD-tDCS. Based on the optimized current modeling, electrodes were placed at P10, CP6, P4, AF8, and FT9, which were symmetrical to the electrode positions of left FFA. In the anodal stimulation, the anodal electrode was placed on the P10 (1.5 mA). The remaining electrodes were as follows: CP6 = -1.04 mA, P4 = -0.15 mA, AF8 = -0.18 mA and FT9 = -0.13 mA. In the cathodal stimulation, the polarity of all electrodes was reversed. In addition, all the experimental procedures were the same as in experiment 2.

Results

Group Analyses

Three groups (anodal, cathodal and sham) and two blocks (pre- and post) of two-way ANOVA showed that the main effect of block was significant, $F(1,19) = 6.30$, $p = 0.021$, $\eta^2 = 0.11$. Additionally, the main effect of group and interaction effect were non-significant, $Fs < 1$ (**Figure 6A**), indicating that both anodal and cathodal tDCS over the right FFA had no effects on face view discrimination threshold at the group level.

Furthermore, we conducted a three-way ANOVA on the face view discrimination threshold, with groups (anodal, cathodal and sham) and blocks (before and immediately after) as within-subjects factors and stimulation sites (left and right FFA) as between-subject factors, to combine the results of Experiment 1 (left FFA) and Experiment 3 (right FFA). The results only showed a significant main effect of block, $F(1,36) = 15.66$, $p < 0.001$, $\eta^2 = 0.07$. No significance was found for other effects, $Fs < 1$. There was no significant effect regarding stimulation sites, indicating that the effect of tDCS over the left and right FFAs was not different at the group level.

Individual Analyses

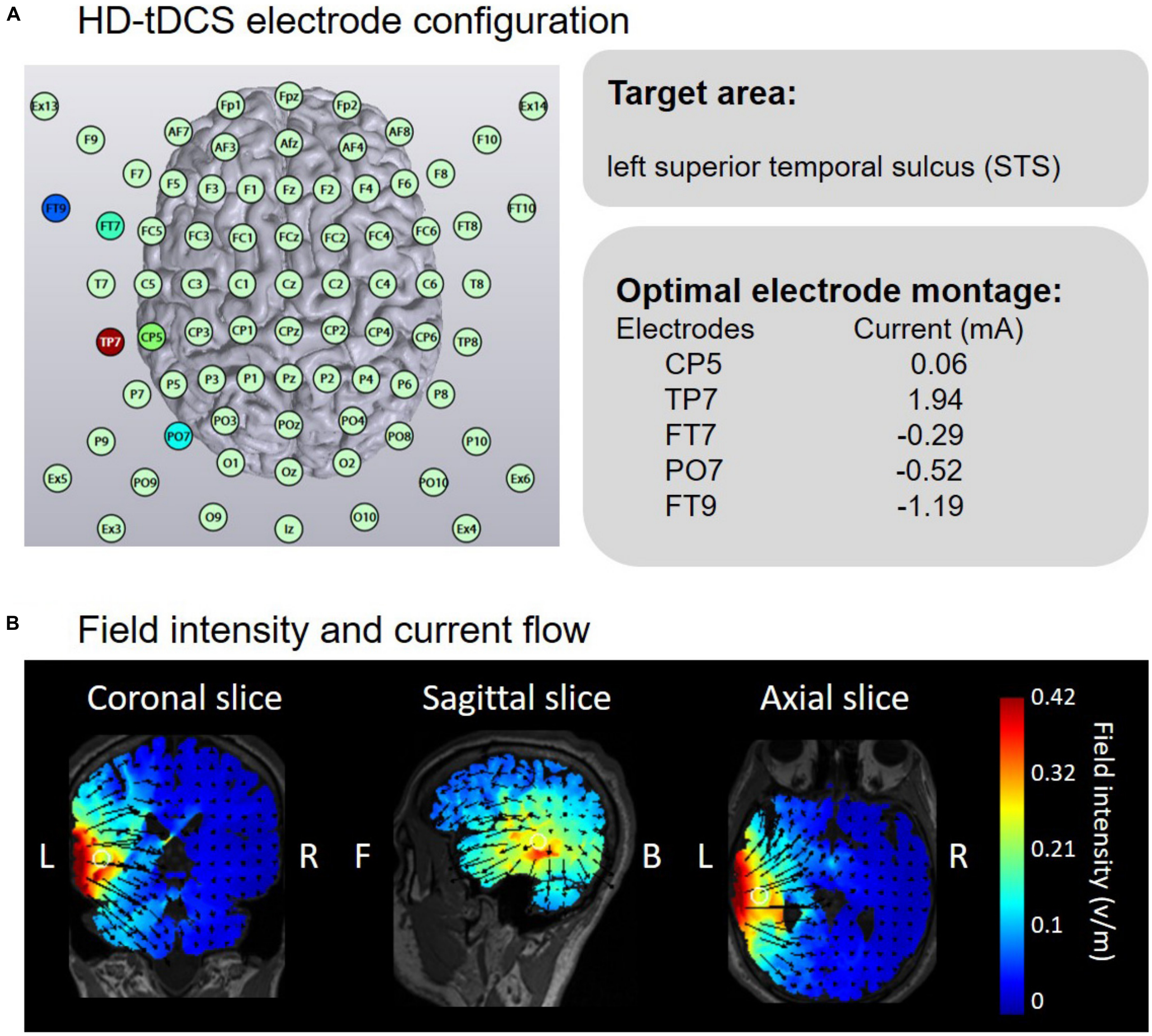
Partial correlation analyses showed non-significant relationships between the initial threshold and chance of threshold with Bonferroni correction for anodal ($r = 0.37$, $p = 0.122$), cathodal ($r = 0.51$, $p = 0.026$) and sham ($r = 0.38$, $p = 0.109$) conditions. Furthermore, the best-fitting regression lines were estimated in three tDCS groups (see **Figure 6B**). ANCOVAs were performed and revealed that the two linear models were parallel independent of anodal vs. sham outcomes, and cathodal vs. sham outcomes, $Fs < 1$, excluding the effect of stimulation. These results suggested that anodal and cathodal tDCS over the right FFA did not change face view discrimination at individual level.

Additionally, we compared the slopes of the anodal models between the left and right FFAs. ANCOVAs showed no significant difference in the slopes of the anodal models between the left and right FFAs, $F(1,34) = 1.23$, $p = 0.276$, $\eta^2 = 0.02$.

DISCUSSION

The current study used HD-tDCS over the left FFA, left STS or right FFA to modulate cortical excitability of these three brain regions and explored whether they were causally related to face view discrimination. Initially, both anodal and cathodal tDCS over the left FFA had no effects at the group level. Interestingly, anodal tDCS, but not cathodal tDCS, over the left FFA modulated the relationship between the individual initial threshold and the magnitude of the threshold change. Specifically, the degree of change after anodal tDCS relied on the initial performance, with poorer (or better) initial performers having a greater gain (or loss). In contrast, neither anodal tDCS nor cathodal tDCS over the left STS or right FFA influenced the threshold of face view discrimination at the group and individual levels. These results indicated that the left FFA seemed to be more susceptible to discriminate face views than the left STS and right FFA.

Interestingly, the effect of tDCS over the left FFA was not found at the group level but at the individual level. As shown in **Figure 3B**, the data points following anodal tDCS were distributed around the mean level of threshold change following



Target area:

left superior temporal sulcus (STS)

Optimal electrode montage:

Electrodes	Current (mA)
CP5	0.06
TP7	1.94
FT7	-0.29
PO7	-0.52
FT9	-1.19

FIGURE 4 | Electrode configuration and simulated electric field in anodal HD-tDCS over the left STS. **(A)** HD-Targets software defined the optimal electrode montage to optimize the focality on the left STS. **(B)** HD-Explore software modeled the field intensity and current flow.

sham tDCS, indicating a convergence effect. In particular, the better initial performers became worse; in contrast, the poorer initial performer improved. Finally, the differential changes averaged together, causing a non-significant change at the group level. Similar outcomes were also found in previous research regarding inhibitory control (Plewnia et al., 2013; Nieratschker et al., 2015; Weidacker et al., 2016), attentional blink (London and Slagter, 2015) and contrast sensitivity (Wu et al., 2021). For example, Wu et al. (2021) did not find modulation of anodal or cathodal stimulation over the primary visual cortex (Oz) on group-level contrast sensitivity compared with sham stimulation. However, initial contrast sensitivity was found to be negatively related to the magnitude of change (more typical at a spatial frequency of 8 c/°) only in the anodal condition, which suggested the involvement of Oz in contrast sensitivity. The convergence effect demonstrates that the various magnitudes of performance change after tDCS depend on the baseline performance. Two studies regarding visual short-term memory also revealed that low initial performers benefited from stimulation, but high performers did not (Tseng et al., 2012; Hsu et al., 2014). Similarly, in a study on attentional blink, participants with a large baseline attentional blink decreased the attentional blink after anodal tDCS, but those with a small baseline attentional blink increased the attentional blink (London and Slagter, 2015). Together, these findings suggest that individual differences in initial performance should be taken into consideration because the group mean results may cover some notable findings.

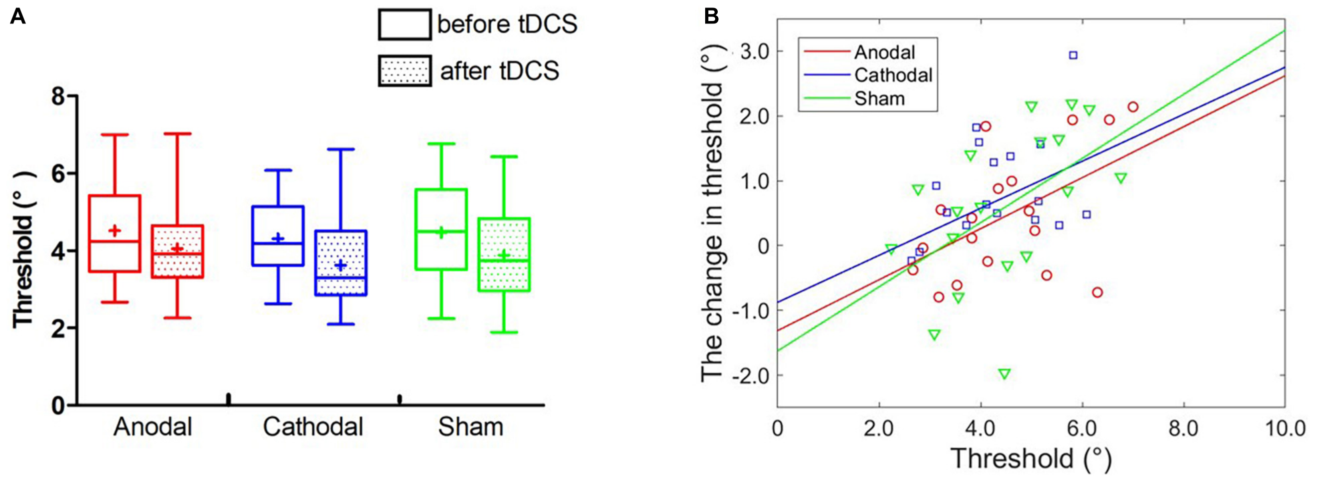


FIGURE 5 | Effect of tDCS over the left STS on the threshold of face view discrimination. **(A)** The average thresholds are depicted before and immediately after tDCS. Error bars indicate standard errors (SE). **(B)** The baseline threshold as a function of the magnitude of threshold change for each type of stimulation.

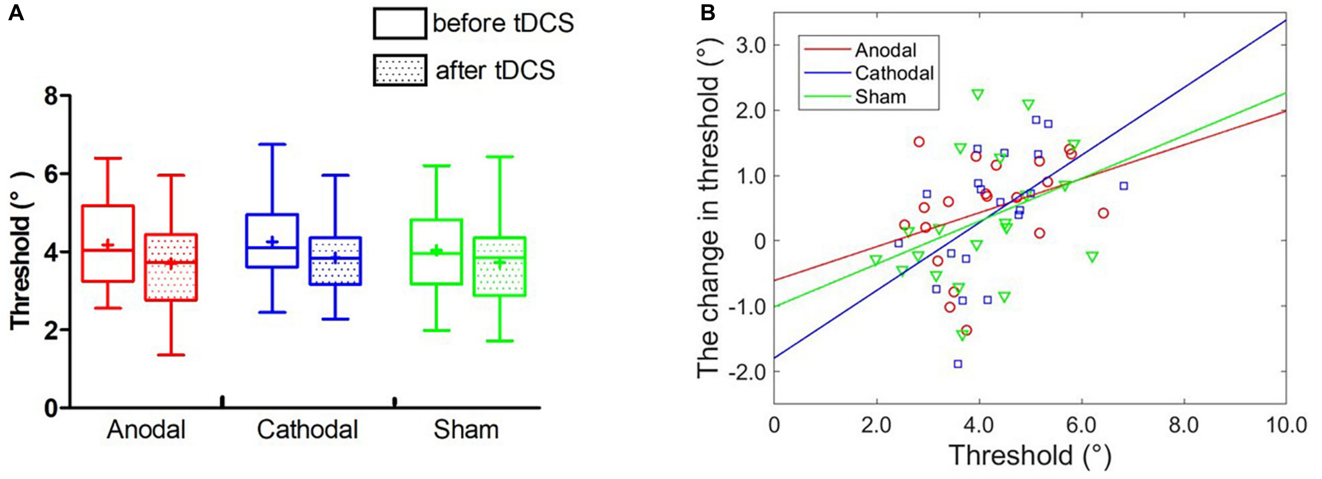


FIGURE 6 | Effect of tDCS over the right FFA on the threshold of face view discrimination. **(A)** The average thresholds are depicted before and immediately after tDCS. Error bars indicate standard errors (SE). **(B)** The baseline threshold as a function of the threshold changes for each condition.

The convergence effect may have two possible explanations. First, the current intensity and the baseline neural excitability work together to influence tDCS effects. In particular, initial performance may reflect cortical excitability related to the processing efficacy for incoming stimuli (Silvanto et al., 2018). A better initial performance indicates higher excitability; in contrast, a poorer performance signifies lower excitability. Here, the degree of cortical excitability caused by anodal tDCS may be located in the middle position between better and poorer performers. For better performers (high initial excitability), tDCS decreased excitability and then worsened performance. For poorer performers (low initial excitability), tDCS increased excitability and further enhanced performance. Second, the prestimulation cortical excitation/inhibition balance determines the stimulation effects on performance. Specifically, individuals possess various baseline balances between cortical excitation and inhibition within a certain brain area, which influence the stimulation effect based on whether the stimulation moves the balance toward or away from its optimum (London and Slagter, 2021). If a certain brain area already had optimal balance, tDCS would worsen efficiency since the optimal balance is broken. Conversely, if the area has been functioning suboptimally, tDCS would improve its efficiency. Thus, individuals with lower initial performance have suboptimal levels of cortical excitability, and their performance may be improved by tDCS, while individuals with higher initial performance have optimal or supraoptimal cortical excitability, and their performance may be impaired by tDCS. More details regarding the two likely explanations should be investigated in future research.

In an early neuroimaging study, Andrews and Ewbank (2004) found a role of the STS in face view discrimination. Specifically, they showed that a face with the same identity generated greater activation in the STS than faces with different identities when changing the head/gaze direction. Clearly, their study involved face identities using real face images, in addition to face views. In contrast, the current study focused only on the face view using artificial 3D face images. Thus, the above differences between their study and our study may be the reason for the inconsistent findings.

Our findings provide further evidence for hemispheric asymmetry in face processing. Many literatures have documented that the right lateralized responses to faces in the brain were much larger than those in the left hemisphere (Pinsk et al., 2005; Fang et al., 2007), but we still know little about the exact functional difference between these two hemispheres. Meng et al. (2012) proposed different functions in the bilateral cerebral hemispheres. Specifically, the left FFA performs the graded analyses of faces, while the right FFA performs the categorical analyses. Additionally, the left FFA is more susceptible to contextual information than the right FFA. Based on the findings of the current study, we argue that the left FFA is more susceptible to face view discrimination than the right FFA. Our view is consistent with two previous neuroimaging studies (Su et al., 2012; Bi et al., 2014). It's worth noting that although we confirmed the importance of the left FFA with HD-tDCS, we cannot deny potential contributions from other cortical areas (e.g., right FFA) since tDCS is limited by the low spatial resolution and weak intensity of current to cortex.

The duration of tDCS effects remains controversial. Some early research showed a short-lasting effect of tDCS, such as 7 min (Antal et al., 2001) or 10 min (Antal et al., 2004) after stimulation, which would limit its practical application. To investigate the duration, the threshold of face view discrimination was measured four times: before and immediately, 10 min and 20 min after tDCS. At the group level, neither anodal nor cathodal tDCS influenced the threshold regardless of the duration. Furthermore, the analyses of individual differences immediately after tDCS were significant: the correlation between the initial threshold and the change in threshold was significant after anodal tDCS rather than cathodal and sham tDCS; additionally, the difference in the slope of the linear models between anodal and sham tDCS reached marginal significance. In contrast, the analyses of individual differences at 10 min and 20 min after tDCS were not significant. However, we still found a similar tendency at these three time points. On the one hand, these results verified the reliability of our findings because a similar tendency was found in the three tests at different times. On the other hand, the non-significant results at 10 min and 20 min after tDCS may result from the gradual disappearance of the tDCS effect at these two times.

One potential limitation in this study is the non-specific effects of tDCS on left FFA. In an intact man, the brain is protected from electricity by the skull and by the scalp, both of which normally offer considerable resistance. Thus, the localization of the stimulus on the cortex will always be much less sharp, and the current decays very much. In other words, tDCS is better suited for superficial areas. However, the fusiform gyrus is ventral and medial and the location of the FFA may not be directly accessed. Additionally, it has been found that brain areas are not independent and are especially interconnected. Thus, it is possible that tDCS actually affects the whole network by modulating one part of the network, which may generate unexpected interactions between stimulation sites (Zheng et al., 2011; Krause and Cohen Kadosh, 2014). Given this, there is no direct stimulation of the FFA alone and any effects, if present, cannot be interpreted because of lack of specificity. The current study used HD-tDCS, which has been confirmed to generate more focal current on the target brain region than conventional tDCS with a 1×1 electrode configuration. HD-tDCS is more beneficial by improving the focality of the current and hence potentially limiting the interacting effects among different brain regions. Additionally, the results of this study are consistent with previous fMRI study in which left FFA is related with the face view discrimination (Bi et al., 2014). Thus, it is reasonable to speculate that the left FFA was stimulated in current research.

There were at least two contributions of this study. First, neuroimaging studies have identified the involvement of the left FFA and left STS in face view discrimination. tDCS has advantages in investigating the causal relevance of target brain regions for corresponding cognitive functions. To the best of our knowledge, this research is the first to confirm the role of the left FFA in face discrimination through HD-tDCS, contributing to a deeper understanding of the underlying neuromechanisms of face processing. Second, the previous literature has often used group-level results to validate tDCS effects. However, we did not find a significant influence of anodal tDCS over the left FFA at the group level. In contrast, we found a significantly negative correlation between the initial threshold and the change in threshold, still indicating the role of the left FFA in face view discrimination. These results suggest that the group average results may cover some important findings due to the great variability across individuals. Future research should take individual differences, such as baseline performance, into account.

The current study found that individuals with poorer initial performance showed more improvement following anodal tDCS over the left FFA but not the left STS and right FFA, and further verified the important role of the left FFA in face view discrimination. In future research, individual variability should be taken into account to decrease variability, uncover unclear mechanisms and develop individualized stimulation methods.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Air Force Medical University.

The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DW completed the experiment and wrote the manuscript. PZ and NL assisted with the experiment and analyzed the data. KS provided technical guidance and site support. WX conceived the idea and provided the financial support. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Natural Science Foundation of Hebei Province (C2021205005 to PZ).

REFERENCES

Andrews, T. J., and Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage* 23, 905–913. doi: 10.1016/j.neuroimage.2004.07.060

Antal, A., Nitsche, M. A., Kruse, W., Kincses, T. Z., Hoffmann, K. P., and Paulus, W. (2004). Direct current stimulation over V5 enhances visuomotor coordination by improving motion perception in humans. *J. Cogn. Neurosci.* 16, 521–527. doi: 10.1162/089892904323057263

Antal, A., Nitsche, M. A., and Paulus, W. (2001). External modulation of visual perception in humans. *Neuroreport* 12, 3553–3555. doi: 10.1097/00001756-200111160-00036

Axelrod, V., and Yovel, G. (2012). Hierarchical processing of face viewpoint in human visual cortex. *J. Neurosci.* 32, 2442–2452. doi: 10.1523/JNEUROSCI.4770-11.2012

Benwell, C. S., Learmonth, G., Miniussi, C., Harvey, M., and Thut, G. (2015). Non-linear effects of transcranial direct current stimulation as a function of individual baseline performance: evidence from biparietal tDCS influence on lateralized attention bias. *Cortex* 69, 152–165. doi: 10.1016/j.cortex.2015.05.007

Bi, T., Chen, J., Zhou, T., He, Y., and Fang, F. (2014). Function and structure of human left fusiform cortex are closely associated with perceptual learning of faces. *Curr. Biol.* 24, 222–227. doi: 10.1016/j.cub.2013.12.028

Bi, T., Chen, N., Weng, Q., He, D., and Fang, F. (2010). Learning to discriminate face views. *J. Neurophysiol.* 104, 3305–3311. doi: 10.1152/jn.00286.2010

Bolognini, N., Fregni, F., Casati, C., Olgiati, E., and Vallar, G. (2010). Brain polarization of parietal cortex augments training-induced improvement of visual exploratory and attentional skills. *Brain Res.* 1349, 76–89. doi: 10.1016/j.brainres.2010.06.053

De Souza, W. C., Eifuku, S., Tamura, R., Nishijo, H., and Ono, T. (2005). Differential characteristics of face neuron responses within the anterior superior temporal sulcus of macaques. *J. Neurophysiol.* 94, 1252–1266. doi: 10.1152/jn.00949.2004

Ding, Z., Li, J., Spiegel, D. P., Chen, Z., Chan, L., Luo, G., et al. (2016). The effect of transcranial direct current stimulation on contrast sensitivity and visual evoked potential amplitude in adults with amblyopia. *Sci. Rep.* 6:19280. doi: 10.1038/srep19280

Dmochowski, J. P., Datta, A., Bikson, M., Su, Y., and Parra, L. C. (2011). Optimized multi-electrode stimulation increases focality and intensity at target. *J. Neural Eng.* 8:046011. doi: 10.1088/1741-2560/8/4/046011

Fang, F., Murray, S. O., and He, S. (2007). Duration-dependent fMRI adaptation and distributed viewer-centered face representation in human visual cortex. *Cereb. Cortex* 17, 1402–1411. doi: 10.1093/cercor/bhl053

Fox, C. J., Iaria, G., and Barton, J. J. S. (2009). Defining the face processing network: optimization of the functional localizer in fMRI. *Hum. Brain Mapp.* 30, 1637–1651. doi: 10.1002/hbm.20630

Gobbini, M. I., and Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia* 45, 32–41. doi: 10.1016/j.neuropsychologia.2006.04.015

Hartmann, M., Singer, S., Savic, B., Müri, R. M., and Mast, F. W. (2019). Anodal high-definition transcranial direct current stimulation over the posterior parietal cortex modulates approximate mental arithmetic. *J. Cogn. Neurosci.* 32, 862–876. doi: 10.1162/jocn_a_01514

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233. doi: 10.1016/s1364-6613(00)01482-0

Horvath, J. C., Forte, J. D., and Carter, O. (2015). Evidence that transcranial direct current stimulation (tDCS) generates little-to-no reliable neurophysiologic effect beyond MEP amplitude modulation in healthy human subjects: a systematic review. *Neuropsychologia* 66, 213–236. doi: 10.1016/j.neuropsychologia.2014.11.021

Hsu, T. Y., Tseng, P., Liang, W. K., Cheng, S. K., and Juan, C. H. (2014). Transcranial direct current stimulation over right posterior parietal cortex changes prestimulus alpha oscillation in visual short-term memory task. *Neuroimage* 98, 306–313. doi: 10.1016/j.neuroimage.2014.04.069

Hussain, Z., Sekuler, A. B., and Bennett, P. J. (2009). Perceptual learning modifies inversion effects for faces and textures. *Vision Res.* 49, 2273–2284. doi: 10.1016/j.visres.2009.06.014

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311. doi: 10.3410/f.717989828.793472998

Kanwisher, N., and Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 2109–2128. doi: 10.1098/rstb.2006.1934

Krause, B., and Cohen Kadosh, R. (2014). Not all brains are created equal: the relevance of individual differences in responsiveness to transcranial electrical stimulation. *Front. Syst. Neurosci.* 8:25. doi: 10.3389/fnsys.2014.00025

Lang, S., Gan, L. S., Alrazi, T., and Monchi, O. (2019). Theta band high definition transcranial alternating current stimulation, but not transcranial direct current stimulation, improves associative memory performance. *Sci. Rep.* 9:8562. doi: 10.1038/s41598-019-44680-8

Li, L. M., Uehara, K., and Hanakawa, T. (2015). The contribution of interindividual factors to variability of response in transcranial direct current stimulation studies. *Front. Cell. Neurosci.* 9:181. doi: 10.3389/fncel.2015.00181

London, R. E., and Slagter, H. A. (2015). Effects of transcranial direct current stimulation over left dorsolateral pFC on the attentional blink depend on individual baseline performance. *J. Cogn. Neurosci.* 27, 2382–2393. doi: 10.1162/jocn_a_00867

London, R. E., and Slagter, H. A. (2021). No effect of transcranial direct current stimulation over left dorsolateral prefrontal cortex on temporal attention. *J. Cogn. Neurosci.* 33, 756–768. doi: 10.1162/jocn_a_01679

López-Alonso, V., Cheeran, B., Río-Rodríguez, D., and Fernández-del-Olmo, M. (2014). Inter-individual variability in response to non-invasive brain stimulation paradigms. *Brain Stimul.* 7, 372–380. doi: 10.1016/j.brs.2014.02.004

Lu, Z., Hua, T., Huang, C., Zhou, Y., and Dosher, B. A. (2011). Visual perceptual learning. *Neurobiol. Learn. Mem.* 95, 145–151. doi: 10.1016/j.nlm.2010.09.010

Lu, Z., Lin, Z., and Dosher, B. A. (2016). Translating perceptual learning from the laboratory to applications. *Trends Cogn. Sci.* 20, 561–563. doi: 10.1016/j.tics.2016.05.007

Mcmahon, D. B. T., and Leopold, D. A. (2012). Stimulus timing-dependent plasticity in high-level vision. *Curr. Biol.* 22, 332–337. doi: 10.1016/j.cub.2012.01.003

Meng, M., Cherian, T., Singal, G., and Sinha, P. (2012). Lateralization of face processing in the human brain. *Proc. Biol. Sci.* 279, 2052–2061. doi: 10.1098/rspb.2011.1784

Nieratschker, V., Kiefer, C., Giel, K., Krüger, R., and Plewnia, C. (2015). The COMT Val/Met polymorphism modulates effects of tDCS on response inhibition. *Brain Stimul.* 8, 283–288. doi: 10.1016/j.brs.2014.11.009

Nikolin, S., Loo, C. K., Bai, S., Dokos, S., and Martin, D. M. (2015). Focalised stimulation using high definition transcranial direct current stimulation (HD-tDCS) to investigate declarative verbal learning and memory functioning. *Neuroimage* 117, 11–19. doi: 10.1016/j.neuroimage.2015.05.019

Parkin, B. L., Ekhtiari, H., and Walsh, V. F. (2015). Non-invasive human brain stimulation in cognitive neuroscience: a primer. *Neuron* 87, 932–945. doi: 10.1016/j.neuron.2015.07.032

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., et al. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Exp. Brain Res.* 86, 159–173. doi: 10.1007/BF00231050

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., et al. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc. R. Soc. Lond. B. Biol.* 223, 293–317. doi: 10.1098/rspb.1985.0003

Pinsk, M. A., Desimone, K., Moore, T., Gross, C. G., and Kastner, S. (2005). Representations of faces and body parts in macaque temporal cortex: a functional MRI study. *Proc. Natl. Acad. Sci. U. S. A.* 102, 6996–7001.

Pitcher, D., Walsh, V., Yovel, G., and Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Curr. Biol.* 17, 1568–1573. doi: 10.1016/j.cub.2007.07.063

Plewnia, C., Zwissler, B., Längst, I., Maurer, B., Giel, K., and Krüger, R. (2013). Effects of transcranial direct current stimulation (tDCS) on executive functions: influence of COMT Val/Met polymorphism. *Cortex* 49, 1801–1807. doi: 10.1016/j.cortex.2012.11.002

Reinhart, R. M. G., Xiao, W., McClenahan, L. J., and Woodman, G. F. (2016). Electrical stimulation of visual cortex can immediately improve spatial vision. *Curr. Biol.* 26, 1867–1872. doi: 10.1016/j.cub.2016.05.019

Rotshtein, P., Henson, R. N. A., Treves, A., Driver, J., and Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat. Neurosci.* 8, 107–113. doi: 10.1038/nn1370

Sasaki, Y., Náñez, J. E., and Watanabe, T. (2012). Recent progress in perceptual learning research. *Wiley Interdiscip. Rev. Cogn. Sci.* 3, 293–299. doi: 10.1002/wcs.1175

Sczesny-Kaiser, M., Beckhaus, K., Dinse, H. R., Schwenkreis, P., Tegenthoff, M., and Höffken, O. (2016). Repetitive transcranial direct current stimulation induced excitability changes of primary visual cortex and visual learning effects—A pilot study. *Front. Behav. Neurosci.* 10:116. doi: 10.3389/fnbeh.2016.00116

Silvanto, J., Bona, S., Marelli, M., and Cattaneo, Z. (2018). On the mechanisms of Transcranial Magnetic Stimulation (TMS): how brain state and baseline performance level determine behavioral effects of TMS. *Front. Psychol.* 9:741. doi: 10.3389/fpsyg.2018.00741

Stagg, C. J., Jayaram, G., Pastor, D., Kincses, Z. T., Matthews, P. M., and Johansen-Berg, H. (2011). Polarity and timing-dependent effects of transcranial direct current stimulation in explicit motor learning. *Neuropsychologia* 49, 800–804. doi: 10.1016/j.neuropsychologia.2011.02.009

Stagg, C. J., and Nitsche, M. A. (2011). Physiological basis of transcranial direct current stimulation. *Neuroscientist* 17, 37–53. doi: 10.1177/1073858410386614

Su, J., Chen, C., He, D., and Fang, F. (2012). Effects of face view discrimination learning on N170 latency and amplitude. *Vision Res.* 61, 125–131. doi: 10.1016/j.visres.2011.08.024

Tseng, P., Hsu, T. Y., Chang, C. F., Tzeng, O. J. L., Hung, D. L., Muggleton, N. G., et al. (2012). Unleashing potential: transcranial direct current stimulation over the right posterior parietal cortex improves change detection in low-performing individuals. *J. Neurosci.* 32, 10554–10561. doi: 10.1523/JNEUROSCI.0362-12.2012

Turski, C. A., Kessler-Jones, A., Chow, C., Hermann, B., Hsu, D., Jones, J., et al. (2017). Extended multiple-field high-definition transcranial direct current stimulation (HD-tDCS) is well tolerated and safe in healthy adults. *Restor. Neurol. Neurosci.* 35, 631–642. doi: 10.3233/RNN-170757

Weidacker, K., Weidemann, C. T., Boy, F., and Johnston, S. J. (2016). Cathodal tDCS improves task performance in participants high in Coldheartedness. *Clin. Neurophysiol.* 127, 3102–3109. doi: 10.1016/j.clinph.2016.05.274

Woods, A. J., Antal, A., Bikson, M., Boggio, P. S., Brunoni, A. R., Celnik, P., et al. (2016). A technical guide to tDCS, and related non-invasive brain stimulation tools. *Clin. Neurophysiol.* 127, 1031–1048. doi: 10.1016/j.clinph.2015.11.012

Wu, D., Li, C., Liu, N., Xu, P., and Xiao, W. (2020). Visual motion perception improvements following direct current stimulation over V5 are dependent on initial performance. *Exp. Brain Res.* 238, 2409–2416. doi: 10.1007/s00221-020-05842-7

Wu, D., Zhou, Y., Lv, H., Liu, N., and Zhang, P. (2021). The initial visual performance modulates the effects of anodal transcranial direct current stimulation over the primary visual cortex on the contrast sensitivity function. *Neuropsychologia* 156:107854. doi: 10.1016/j.neuropsychologia.2021.107854

Zheng, X., Alsop, D. C., and Schlaug, G. (2011). Effects of transcranial direct current stimulation (tDCS) on human regional cerebral blood flow. *Neuroimage* 58, 26–33. doi: 10.1016/j.neuroimage.2011.06.018

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 30 November 2021
doi: 10.3389/fpsyg.2021.776867

Covert Attention to Gestures Is Sufficient for Information Uptake

Kendra Gimhani Kandana Arachchige, Wivine Blekic, Isabelle Simoes Loureiro and Laurent Lefebvre*

Department of Cognitive Psychology and Neuropsychology, University of Mons, Mons, Belgium

Numerous studies have explored the benefit of iconic gestures in speech comprehension. However, only few studies have investigated how visual attention was allocated to these gestures in the context of clear versus degraded speech and the way information is extracted for enhancing comprehension. This study aimed to explore the effect of iconic gestures on comprehension and whether fixating the gesture is required for information extraction. Four types of gestures (i.e., semantically and syntactically incongruent iconic gestures, meaningless configurations, and congruent iconic gestures) were presented in a sentence context in three different listening conditions (i.e., clear, partly degraded or fully degraded speech). Using eye tracking technology, participants' gaze was recorded, while they watched video clips after which they were invited to answer simple comprehension questions. Results first showed that different types of gestures differently attract attention and that the more speech was degraded, the less participants would pay attention to gestures. Furthermore, semantically incongruent gestures appeared to particularly impair comprehension although not being fixated while congruent gestures appeared to improve comprehension despite also not being fixated. These results suggest that covert attention is sufficient to convey information that will be processed by the listener.

Keywords: covert attention, iconic gestures, information uptake, eye tracking, incongruency effect

OPEN ACCESS

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Paul Adam,
University of the West of England,
United Kingdom
Costanza Navarretta,
University of Copenhagen, Denmark

****Correspondence:***
Kendra Gimhani Kandana Arachchige
kendra.kandanaarachchige@umons.ac.be

Specialty section:
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

Received: *14 September 2021*
Accepted: *29 October 2021*
Published: *30 November 2021*

Citation:
Kandana Arachchige KG, Blekic W, Simoes Loureiro I and Lefebvre L (2021) Covert Attention to Gestures Is Sufficient for Information Uptake.
Front. Psychol. 12:776867.
doi: 10.3389/fpsyg.2021.776867

INTRODUCTION

In daily conversational situations, our senses are continuously exposed to numerous types of information, not all of which are processed. Among the information that could benefit listeners' comprehension, iconic gestures are hand gestures that convey meaning semantically related to the speech they accompany (McNeill, 1992, 2008; Kendon, 2004). According to Kelly et al. (2004), these gestures could create a visuospatial context that would affect the subsequent processing of the message. Research in this field refers to the combination of gestural and verbal information to create a unified meaning as "gesture-speech integration" (Holle and Gunter, 2007).

Several studies have shown that listeners could indeed benefit from the presence of iconic gestures (Beattie and Shovelton, 2001, 2002; Holle and Gunter, 2007; Holler et al., 2009), particularly in the event of degraded speech (Drijvers and Özyürek, 2017, 2020; Drijvers et al., 2019). Drijvers and Özyürek (2017) observed a joint contribution of iconic gestures and visible speech (i.e., lip movements) to comprehension in a speech degraded context. According to these authors, the semantic information conveyed through iconic gestures adds to the phonological information present in visible speech. However, a minimum level of auditory input is required for an optimal enhancement by visible speech (Ross et al., 2007). A question that has been fairly less investigated concerns the allocation of visual attention to iconic gestures (Beattie et al., 2010; Drijvers et al., 2019).

Earlier studies on where visual attention is allocated in a conversational context showed that listeners mainly fixated the speaker's face (Gullberg and Holmqvist, 1999, 2001, 2006; Gullberg, 2003; Gullberg and Kita, 2009; Beattie et al., 2010; Drijvers et al., 2019) and only minimally fixated gestures (Gullberg and Holmqvist, 1999). In a face-to-face context, the increased attention allocated to a speaker's face is assumed to reflect interest and engagement (Gullberg, 2003). This pattern has also been observed using audio-visual stimuli (Gullberg and Holmqvist, 2001; Gullberg, 2003). According to Gullberg and Holmqvist (2006), speech and gestures compete for attention. Two mechanisms are of interest: a bottom-up selection, referring to attention being involuntarily captured by a physical characteristic of the stimulus (Theeuwes, 2010; Moore and Zirnsak, 2017; Wang and Theeuwes, 2020), and a top-down selection, where the individual voluntarily brings the stimulus into their focus of attention (Moore and Zirnsak, 2017; Wang and Theeuwes, 2020) thereby *fixating* it (Gullberg and Holmqvist, 1999). Considering that participants seem to spend more time focusing on a speaker's face, this theorizes the presence of the gesture in the peripheral visual field that could induce a bottom-up visual attention process in an attempt to retrieve task-relevant information (Gullberg and Holmqvist, 2006). Nevertheless, the amount of information retrieved from a fixated stimulus is higher than the amount retrieved from the stimulus attended at a peripheral location (Gullberg and Holmqvist, 1999).

Although the face area is mainly fixated, there are some instances where the gestures are more looked at. A first study by Rimé et al. (cited by Gullberg and Holmqvist, 1999) showed that when faced with speech in their non-native language, participants tended to fixate gestures more than when faced with speech in their own language. This distinction between native and non-native language has also been found in a more recent study (Drijvers et al., 2019) showing that while both groups of participants mostly fixated faces, the non-native group more oftenly gazed toward gestures than native listeners. Additionally, Gullberg and Holmqvist (2006) showed that when speakers looked at their own gestures, listeners were more likely to gaze toward the gestures. Several authors also highlighted that gestures containing holds (i.e., a temporary cessation of the gestural movement) attracted more the listener's visual attention than gestures without holds (Nobe et al., 1997; Gullberg and Holmqvist, 2006; Gullberg and Kita, 2009). Gullberg and Holmqvist (2006) associated the effect of a speaker's gaze to a top-down-related effect and the presence of a hold to a bottom-up-related effect.

To the best of the researchers' knowledge, the study conducted by Drijvers et al. (2019) was the first and only study to investigate how overt visual attention is allocated in the context of degraded speech. Native and non-native participants were presented with video clips of an actor enacting an iconic gesture combined with either clear or degraded speech. The participants were fitted with an eye tracking device and were asked to recognize which single verb was heard among four propositions. Their results demonstrated that an overt allocation of visual attention to gestures in the presence of degraded speech benefited native speakers. Drijvers et al. (2019) also showed that participants were more likely to gaze at the face and mouth areas both when presented with clear and degraded speech.

Exploring whether listeners actually attend and integrate gestural information, Gullberg and Kita (2009) found no evidence of an association between gesture fixation and information uptake. According to these authors, attention to gestures appears to be mostly covert (Gullberg and Kita, 2009), referring to the deployment of attention at a given space in the absence orienting eye movements (Carrasco and McElree, 2001). This attention can then be directed by the speaker either through speech or by using a deictic (i.e., pointing type) gesture (Wakefield et al., 2018). In contrast, Beattie et al. (2010) found that low-span character-viewpoint gestures (i.e., gestures that are produced from the viewpoint of the character and do not cross any boundaries in the gesture space; see Beattie et al., 2010 for details) were the most communicative gesture type and the most and longest fixated gestures. Their results suggest an association between attentional focus and the uptake of information from this particular type of iconic gesture (Beattie et al., 2010).

In view of the current scarcity of studies investigating the uptake of gestural information in the context of clear and/or degraded speech, the present study thus aims to explore this issue in a more ecological manner and using a different paradigm than in previous studies. Rather than being presented with isolated words (such as in Drijvers et al., 2019), participants would be presented with short sentences describing daily events (as in Beattie et al., 2010). In each sentence, one element is associated with an iconic gesture, representing either the action or a physical attribute of the object mentioned in speech (McNeill, 1992). In contrast to previous studies, participants will be presented with different types of gestures (i.e., semantically incongruent iconic gestures, syntactically incongruent iconic gestures, meaningless configurations, and semantically congruent iconic gestures). A gesture is considered semantically incongruent if its meaning fit its position in the sentence but does not match the sentence's meaning (e.g., the gesture "*small*" in the sentence "the car was on a *large* road"). In other words, a semantically incongruent gesture will, if a possible meaning is "translated" to words, retain a correct grammatical class for its position in the sentence. Another example could be the gesture "close" (which would be enacted by the arm, after moving away from the rest position on the lap, fist closed, in front of the actor in the gesture space, would come back toward the chest), enacted simultaneously to the verbal utterance "knocked on" in the sentence "she knocked on her neighbors' door." In this case, while a semantic congruency would occur in the event of audible speech, in the event of altered speech, the sentence would retain meaning as an action gesture takes place simultaneously to a verb in the sentence. A syntactically incongruent gesture is a gesture that conveys a meaning that would not fit that place in the sentence (e.g., the gesture "*rectangle*" in the sentence "he must know how to *drive*"). Finally, meaningless configurations were taken from Wu and

Coulson (2014). These "gestures" are actually *meaningless configurations* (rather than *gestures*) and have been found to be uninterpretable by Wu & Coulson (see Wu and Coulson, 2014 for more details).

In behavioral studies, the investigation of gesture-speech integration *via* the use of mismatching gesture-speech pairs is far from new and has consistently shown that iconic gestural and verbal information were integrated to form a unified representation (Kelly et al., 2010a,b; Margiotoudi et al., 2014; Zhao et al., 2018). The advantage of using mismatching gestures in investigating information uptake relies in the possibility of observing a negative effect of these gestures on comprehension. Moreover, it allows to explore whether fixating the gestures is a necessary requirement for information uptake or whether their presence alone affects comprehension. To the best of the researchers' knowledge, to date, no work has explored the visual allocation of attention to iconic gestures in a speech degraded context by contrasting the presentation of matching and mismatching gestures in a hope to shed light on information uptake.

Consistent with previous studies, we first expected participants to spend a longer time fixating the face compared to gestures, particularly in the event of degraded speech. No previous study having investigated the allocation of visual attention to different types of mismatching gestures, we could only speculate the presence of different visual allocation behaviors depending on the type of gestures presented. Regarding the comprehension task, we anticipated the higher scores in the semantic congruent condition compared to the other three conditions and the lowest scores in the semantically incongruent condition compared to the other conditions. Hence, we expected a difference in the processing of the three types of mismatching gestures.

MATERIALS AND METHODS

Participants

Hundred and thirty-six healthy French-speaking participants took part in the study. They were recruited through announcements on the University groups on social media. Exclusion criteria included neurological and current psychological disorders as well as visual and/or auditory impairments. Wearing glasses was also considered as an exclusion criterion due to the potential reflection that could disrupt the eye tracking recording. Six participants had to be excluded following technical failures. Two more participants were excluded for not having French as their mother tongue.[1] The final sample consisted of 128 French-speaking participants (35 men; $M_{\text{age}} = 21.34$; SD = 0.21; Min = 17; Max = 28). They each received 5€ for taking part in the experiment. This study was approved by the Ethics Committee of the University of Mons. All participants gave written informed consent before taking part in this study.

Material

Computerized Task

The task involved material that has never been used in previous studies. Hence, several steps were required to ensure the validity of our material. The creation of the task as well as the different validation steps are presented here below.

This task was performed using SMI Experiment Suite 360° software. During the computerized task, participants were asked to sit in front of a computer and keyboard. The experimental task comprised 200 trials (50 sentences × four types of gestures). The videos consisted of an actor uttering a sentence while performing a gesture. They were presented semi-randomly and were followed by a comprehension question. A graphical illustration of the trial structure is provided in **Figure 1**. The list of the sentences that were used can be found in **Appendix A**.[2] The Appendix also informs on which element was enacted through a gesture (in bold), whether congruent or incongruent (semantically and syntactically), which meaningless configuration was used (see Wu & Coulson for the references), and what statement was presented to assess participants' comprehension.

Stimuli and Equipment

Creation of Stimuli

Sentences

A) The final stimuli used for the experimental task consisted of video recordings of 50 short sentences relating daily topics. Each sentence contained one (and only one) element that was enacted through an iconic gesture (either describing an action, a physical attribute of an object or spatial relationship between two objects, McNeill, 1992). The element enacted through an iconic gesture is indicated in bold in **Appendix A**. All gestures were completed by an actor to whom no specific instructions were given besides the word that was to be described. Separate audio recordings were also collected, to ensure a good quality. For all audio recordings, the actor was asked to maintain a neutral prosody as to avoid conveying any emotional or complementary information through the voice. Because no previous studies investigating iconic gestures in a sentence context have been conducted in French, a validation of our stimuli was required beforehand. First, 60 short sentences relating daily topics were created. The study having taken place amidst the COVID-19 pandemic, a Google Form was generated to collect a first set of data on how emotionally loaded our sentences were. We were looking to only keep neutral sentences, emotional content potentially affecting eye gaze (Calvo and Lang, 2004; Rigoulot and Pell, 2012; Lanata et al., 2013). Forty-nine healthy French-speaking participants completed the online questionnaire after giving informed consent. After removing the outliers for age, the final sample consisted of 38 participants ($M_{\text{age}} = 25.55$; SD = 0.74; Min = 21;

---

[1] The inclusion criteria mentioned being fluent in French. However, during testings, the two participants who were not native French speakers, related their difficulties in following the task and the testings, were, therefore, stopped.

[2] Because of the differences between French and English sentences' construction, some incongruencies that work in French would not work in English. Therefore, the English translation is not proposed here but can be found in **Appendix B**.

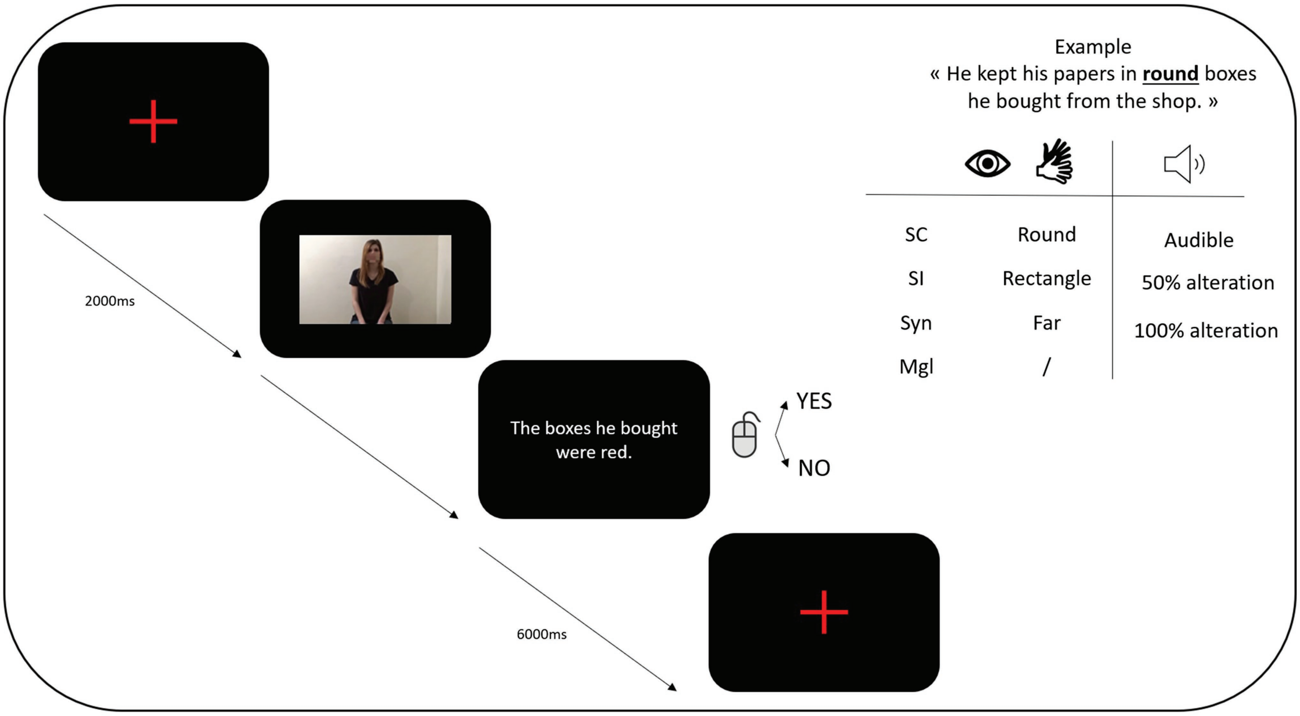


FIGURE 1 | Representation of the proceeding of a task (Sc = semantically congruent; SI = semantically incongruent; Syn = syntactically incongruent; Mgl = Meaningless).

Max = 39). They were asked to judge the emotional valence of 30 pairs of sentences on a 9-level Likert-type scale ranging from 1 (positive) to 9 (negative) with a neutral range located between levels 4 and 6. Participants were presented with Self-Assessment Manikin to help them visualize the graduation. No further instructions were given. After analysis, 29 pairs were scored as neutral (with a mean score between 4 and 6) and one pair was scored at 3.58. The discarded stimuli were still processed as they would be used in the training runs (explained here below). B) The comprehension sentences (relevant for the second part of the task) were formulated in such a way to test participants' comprehension relative to the enacted part of the sentence through a yes/no statement (e.g., if the video stated that "He kept his papers in round boxes he bought from the store.," with "round" being enacted; the comprehension sentence stated that "The boxes he bought were red").

Gestures

Second, 15 new healthy French-speaking participants (three men; $M_{age} = 26.8$; SD = 0.47; Min = 24; Max = 30) judged the semantic congruency and incongruency of 240 videos (60 gestures × 4 types of congruencies) on a 5-level Likert scale ranging from 1 (totally incongruent) to 5 (totally congruent). In other words, participants were asked to judge the level of correspondence between the gesture and the audio it accompanied. The congruent and incongruent gestures were inspired by Kandana Arachchige et al. (in press). The gestures for the meaningless configurations corpus were taken from Wu and Coulson's validated database (Wu and Coulson, 2014). The audio-visual stimuli were presented on Testable.[3] After analysis, the 60 congruent gestures were considered congruent at an average of 4.6/5 and the 180 incongruent gestures were considered incongruent at an average of 1.1/5. One supposedly congruent gesture being judged congruent at only 2.9/5, it has been removed from the stimuli set (along with the other 7 items associated with it). The revised congruent average for the other 58 gestures was of 4.7/5. Third, fifteen new healthy French-speaking participants (3 men; $M_{age} = 23.53$; SD = 0.99; Min = 21; Max = 36) judged the iconicity of each gesture presented with no sound. They were asked to name the gesture seen (interpretative task). The soundless videos were presented on Testable (www.testable.org), and a blank space was available to type their answer. If the gesture evoked no particular meaning, they were asked to type "N/A." Results showed a 57% recognition rate, replicating previous results (Zhao et al., 2018). However, three gestures supposed to show a similar rate were under-recognized (at an average of 13%). These three gestures were removed, and the revised recognition rate increased to 58%. Given the specific nature of iconic gestures that contain meaning *per se* but also depend on context to be understood (Holle and Gunter, 2007), we can assume

---

[3] www.testable.org

that these results support the iconicity of our gestures. At the end of these validation processes, our stimuli sample consisted of 52 sentences x 4 congruency levels: (1) congruent iconic gestures, (2) semantically incongruent iconic gestures, (3) syntactically incongruent iconic gestures, and (4) meaningless configurations. It is important to note that in all the items, the temporal alignment between the gesture/meaningless configuration was kept constant. Indeed, one characteristic of iconic gestures is their temporal alignment with the speech they relate to (McNeill, 1992, 2008; Obermeier and Gunter, 2014). In other words, while the *preparation* phase of the gesture (i.e., "the phase of movement leading up to the stroke," p.112, Kendon, 2004) precedes the verbal utterance, the *stroke* phrase of the iconic gesture (i.e., "the phase when the expression of the gesture (…) is accomplished," p.112, Kendon, 2004) occurs simultaneously to the verbal utterance (Kendon, 2004). This alignment was maintained in our stimuli set, in every condition.

Alteration

The audio sound files for the 50 and 100% alteration conditions were created on Audacity® version 2.3.0. [4]The target verbal utterances (i.e., those to which gestures were related) were separated from the sentence and processed alone. They were combined with a Multi-babble soundtrack (available on open access here: https://www.ee.columbia.edu/~dpwe/sounds/noise/; Richey et al., 2018) at different intensities. For the 50% alteration, both soundtracks were kept at the same volume intensity and superimposed. For the 100% alteration, the verbal utterance soundtrack was reduced of 24 dB, while the multibabble soundtrack was kept unchanged. The modified soundtrack segments were then mixed back with their original sentences. Following these manipulations, a fourth short validation test was conducted. Eleven new healthy French-speaking participants (one man; $M_{age} = 27.36$; SD = 0.96; Min = 25; Max = 36) took part in this pre-test. The sentences were presented on Testable (www.testable.org), and participants were asked whether they could, or not, hear the noised word. They were asked to respond 1 for any word they could understand and 2 for words they were unable to understand. After analysis, the mean score for the 100% alteration set was of 1.9/2. The mean score for the 50% alteration set was of 1.33/2. However, a pair of items were evaluated at an average of 1.88/2 and were therefore discarded from the stimuli set.

Mask

Finally, a mask was applied to the lip area of the actor using PowerDirector 365 software to avoid lip reading in the 50 and 100% alteration conditions. The use of a mask for blurring the mouth area to this end is common in this field of research (e.g., Holle and Gunter, 2007; Cocks et al., 2011; Wu and Coulson, 2014; Sekine et al., 2015; Momsen et al., 2020).

Equipment

An HP computer with SMI iView 250 RED and SMI Experiment Suite 360° software was used to generate the protocol and present the stimuli on a 34x19cm LCD screen running at a frame rate of 60 Hz. Stimuli were presented against a black background. Prior to testing, eye location was calibrated for each participant. The average viewing distance was 56 cm approximating the distance at which the eye tracker receives the best signal. This position was adjusted for each participant until the best possible eye tracking signal was acquired.

Procedure

Participants were seated approximately 56 cm in front of a computer and keyboard. They were first offered a training run to familiarize themselves with the task and eye tracking device as well as to find a comfortable sitting position in which they would be able to stay without moving. The instructions were given a first time verbally and were also presented on screen. Participants were asked to sit as still as possible and watch the videos that were presented. No information on the presence of gestures or the aim of the study was given, and no specific task was required from the participants during the presentation of the video. The sound of the video was either audible, half degraded, or completely degraded. After each video, participants were presented with a sentence and were asked to decide, by keypress (Q or M; counterbalanced between participants), whether the information corresponded or not to that conveyed in the video. In case of doubt, they were asked to answer anyway. The training run consisted of the exact same procedure as the experimental procedure (i.e., including a calibration process) and contained stimuli at 0, 50, and 100% alteration as well as congruent and mismatching gestures. Once the participants were comfortable and were clear with the instructions, they were invited to complete the experimental task. The researcher repeated the instructions verbally before the calibration process of the experimental task to avoid any head movement after the calibration. The instructions were then repeated on screen, and participants could then begin the task. The entire experimental protocol took approximately 11 min to complete.

Eye Tracking Data and Coding Areas of Fixation

Eye tracking data were analyzed using BeGaze software from Senso-Motoric Instruments (SMI). SMI software automatically performed data reduction and exported only usable eye movements. Each video for each participant was processed individually. Areas of interest (AOI) was (1) the face and (2) the hands. Each AOI was defined in a frame-by-frame manner following the course of the gesture and/or head movements.

ANALYSIS AND RESULTS

Analysis

Mean dwelling time (in ms) and number of fixations on the video clips and percentage of correct answers on the behavioral task were analyzed.

---

[4] https://www.audacityteam.org/

The statistical analyses were performed using the software SPSS (version 21). Paired t tests were conducted between the head and hand AOI to determine which zone was most and longer fixated. The full experimental design was a 4 (congruency; semantically congruent, semantically, and syntactically incongruent, meaningless configurations) x 3 (alteration; clear, partly, and completely degraded) factorial design, and a corresponding 4×3 repeated-measure ANOVA was used to analyze the data. Following the ANOVA, follow-up paired *t* tests were conducted where statistical significance was accepted at a Bonferroni-adjusted alpha level of 0.016 ($p = 0.005/3$) following the multiple (i.e., 3x) occurrence of the same variables in the t tests.

Finally, paired t tests were conducted to investigate how visual allocation to the hand AOI would vary with speech degradation.

Results

Dwelling Time on Hand Versus Head AOIs

Paired t tests were conducted to investigate whether the face area would attract more attention in general than the hand area. Results showed significant differences in dwelling time in all conditions, with more time spent on the face area compared to the hand area.

Number of Fixations on Hand Versus Head AOIs

Paired t tests were conducted to investigate the number of fixations in the face area compared to the hand area. Results showed significant differences in number of fixations in all conditions, with more fixations made on the face area compared to the hand area.

Dwelling Time on Face AOI

Paired *t* tests were conducted to investigate whether the face area would attract more attention when the auditory information was degraded. Results yielded no significant differences between any of the alteration levels, for any types of gestures.

Dwelling Time on Hand AOI

The repeated measures ANOVA yielded a main effect of Congruency [$F_{(3,381)} = 32.96$; $p < 0.001$], with more time spent on average on the syntactically incongruent gestures ($M = 320.81$; SD = 18.75), compared to meaningless configurations ($M = 237.08$; SD = 18.02), semantically incongruent gestures ($M = 210.30$; SD = 17.86), and semantically congruent gestures ($M = 201.43$; SD = 15.94). A main effect of alteration was also found [$F_{(2,254)} = 13.71$; $p < 0.001$] with more time spent on average on the hand AOI when the sound was clear ($M = 274.74$; SD = 18.11), compared to in a 50% alteration condition ($M = 235.06$; SD = 17.03) and 100% alteration condition ($M = 217.41$; SD = 15.57). Furthermore, the results also yielded a Congruency x Alteration effect [$F_{(6,762)} = 3.11$; $p < 0.01$, Greenhouse-Geisser correction] reflecting an interaction between the types of gestures presented and the level of hearing alteration (**Figure 2**). A summary of the ANOVA results can be found in **Table 1**.

Follow-up paired t tests were conducted to further clarify the nature of Alteration by Congruency interaction. In the absence of audio alteration, significant differences were observed between meaningless configurations and syntactically incongruent gestures [$t_{(127)} = -3.47$; $p = 0.001$], with longer dwelling times on the latter ($M = 366.39$; SD = 24.01) compared to the former ($M = 277.86$; SD = 24.81). Differences were also found between meaningless configurations and semantically congruent gestures [$t_{(127)} = 3.03$; $p = 0.003$], with more time spent on the former ($M = 277.86$; SD = 24.81) compared to the latter ($M = 199.65$; SD = 22.04). Syntactically incongruent gestures ($M = 366.39$; SD = 24.01) were also longer fixated than semantically congruent [$M = 199.65$; SD = 22.04; $t_{(127)} = 7.14$; $p < 0.001$] and incongruent [$M = 255.06$; SD = 25.66; $t_{(127)} = 3.85$; $p < 0.001$] gestures. The difference between semantically congruent and incongruent gestures just failed to reach significance [$t_{(127)} = -2.24$; $p = 0.02$], No significant differences were found between meaningless configurations and semantically incongruent gestures.

In the case of 50% alteration, significant differences were observed between (1) meaningless configurations and syntactically incongruent [$t_{(127)} = -2.48$; $p = 0.01$] with more time spent on the latter ($M = 311.45$; SD = 22.15) compared to the former ($M = 245.74$; SD = 24.48). A significant difference was also found between meaningless configurations and semantically incongruent gestures [$t_{(127)} = 3.50$; $p = 0.001$], with more time spent on meaningless configurations ($M = 245.74$; SD = 24.48) compared to semantically incongruent gestures ($M = 166.58$; SD = 18.88). More time was also spent on syntactically incongruent gestures ($M = 311.45$; SD = 22.15) compared to semantically congruent [$M = 216.46$; SD = 21.79; $t_{(127)} = 4.13$; $p < 0.001$] and incongruent [$M = 166.58$; SD = 18.88; $t_{(127)} = 8.39$; $p < 0.001$] gestures. The difference between semantically congruent and incongruent gestures just failed to reach significance [$t_{(127)} = 2.25$; $p = 0.02$]. No significant differences were found between the meaningless configurations and semantically congruent gestures.

For the conditions where the audio was 100% altered, significant differences were observed between (1) meaningless configurations and syntactically incongruent gestures [$t_{(127)} = -5.13$; $p < 0.001$], (2) syntactically incongruent gestures and semantically congruent [$t_{(127)} = 4.51$; $p < 0.001$] and incongruent [$t_{(127)} = 3.61$; $p < 0.001$] gestures. More time was spent on the syntactically incongruent gestures ($M = 284.58$; SD = 20.36) compared to the meaningless configurations ($M = 187.62$; SD = 18.50), the semantically congruent gestures ($M = 188.19$; SD = 19.62), and semantically incongruent gestures ($M = 209.25$; SD = 21.98). No significant differences were observed between meaningless configurations and semantically congruent gestures, meaningless configurations, and semantically incongruent gestures and between semantically congruent and semantically incongruent gestures.

Paired t tests were conducted to investigate how visual allocation to the hand AOI would vary with speech degradation. Results showed significant differences for the meaningless configurations with less time spent on them in the 100% alteration condition ($M = 187.62$; SD = 18.50) compared to the clear condition [$t_{(127)} = 3.9$; $p < 0.001$; $M = 277.86$; SD = 24.81] and the 50% alteration condition [$t_{(127)} = 2.57$; $p = 0.01$; $M = 245.74$; SD = 24.48]. A significant difference was also highlighted for the syntactically incongruent gestures, with more time spent on them in the clear condition ($M = 366.39$; SD = 24.01) compared

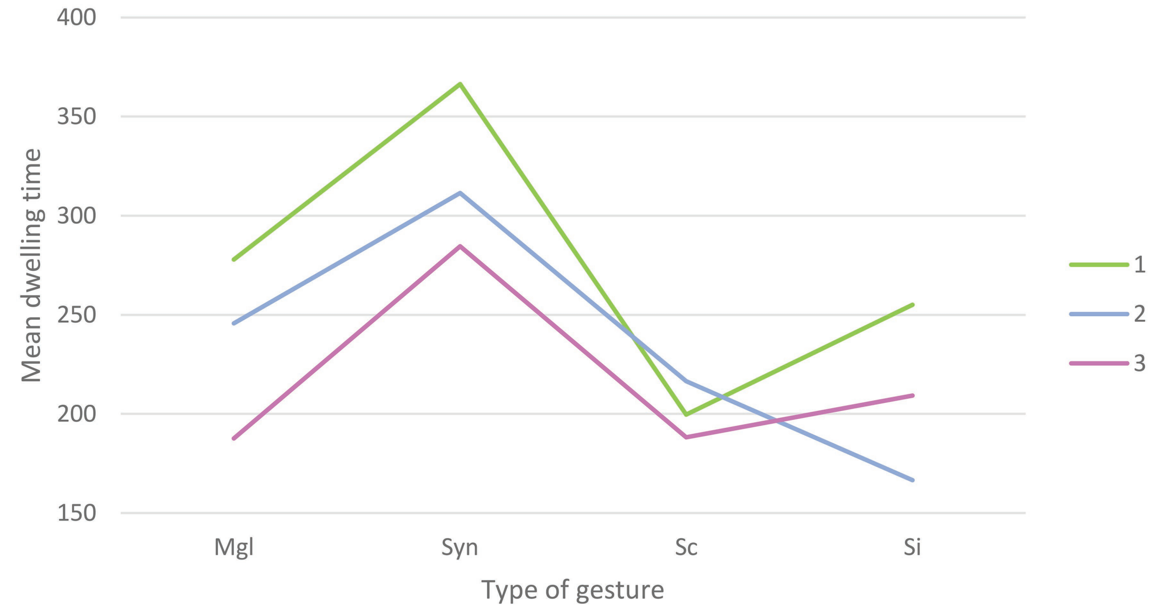


FIGURE 2 | Graphical representation of the interaction for dwelling time between the types of gestures presented and the level of hearing alteration (Mgl = Meaningless; Syn = syntactically incongruent; Sc = semantically congruent; SI = semantically incongruent).

TABLE 1 | ANOVA results for hand AOI dwelling time.

Variable	DoF	F	Sig.
Alteration	2	13.71	0.000
Congruency	3	32.96	0.000
Alteration*Congruency	5.08	3.11	0.008

to the 50% alteration condition [$t_{(127)} = 2.44$; $p = 0.01$; M = 311.45; SD = 22.15] and the 100% alteration condition [$t_{(127)} = 4.21$; p < 0.001; $M = 284.48$; SD = 20.36]. Finally, dwelling time for semantically incongruent gestures was significantly higher [$t_{(127)} = 3.48$; $p = 0.001$] in the clear condition ($M = 255.06$; SD = 25.66) compared to the 50% alteration condition ($M = 166.58$; SD = 18.88) and significantly lower [$t_{(127)} = -2.57$; $p = 0.01$] in the 50% alteration condition compared to the 100% alteration condition ($M = 209.25$; SD = 21.98). No significant differences were observed for the semantically congruent gestures.

Number of Fixations on Hand AOI

The repeated measures ANOVA yielded a main effect of Congruency [$F_{(3,381)} = 17.58$; $p < 0.001$], with more fixations for syntactically incongruent gestures ($M = 0.63$; SD = 0.03) compared to meaningless configurations ($M = 0.50$; SD = 0.03), semantically incongruent ($M = 0.047$; SD = 0.03), and semantically congruent ($M = 0.46$; SD = 0.03) gestures. A main effect of Alteration was also found [$F_{(2,254)} = 13.80$; $p < 0.001$], with more fixations in the clear sound condition ($M = 0.59$; SD = 0.03) compared to in the 50% alteration ($M = 0.50$; SD = 0.03) and 100% alteration ($M = 0.46$; SD = 0.03). Furthermore, the results also yielded a significant Congruency x Alteration effect [$F_{(6,762)} = 2.03$; $p = 0.03$, Greenhouse-Geisser correction] reflecting an interaction between the types of gestures presented and the level of hearing alteration (**Figure 3**). A summary of the ANOVA results can be found in **Table 2**.

Follow-up paired t tests were conducted to further clarify the nature of Alteration by Congruency interaction. In the absence of audio alteration, significant differences were observed between (1) meaningless configurations and semantically congruent [$t_{(127)} = 2.74$; $p < 0.01$] gestures, with more fixations on the meaningless configurations ($M = 0.61$; SD = 0.05) compared to the semantically congruent gestures ($M = 0.47$; SD = 0.04). Significant differences were also observed between syntactically incongruent gestures and semantically congruent [$t_{(127)} = 5.21$; $p < 0.001$] and incongruent [$t_{(127)} = 4.01$; $p < 0.001$] gestures, with more fixations on the syntactically incongruent gestures ($M = 0.73$; SD = 0.04) compared to the semantically congruent ($M = 0.47$; SD = 0.04) and semantically incongruent ($M = 0.53$; SD = 0.04) gestures. The difference between meaningless configurations and syntactically incongruent failed to reach significance at the Bonferroni-adjusted level [$t_{(127)} = -2.183$; $p = 0.03$]. No significant differences were observed between meaningless configurations and semantically incongruent gestures and between semantically congruent and incongruent gestures.

In the case of 50% alteration, significant differences were observed between (1) meaningless configurations and syntactically incongruent [$t_{(127)} = -2.56$; $p = 0.01$] with a higher number of fixations in the latter ($M = 0.06$; SD = 0.04) compared to the former ($M = 0.47$; SD = 0.04) and (2) syntactically incongruent gestures and semantically incongruent gestures [$t_{(127)} = 5.02$; $p < 0.001$] with a higher number of fixations in the former ($M = 0.06$; SD = 0.04) compared to the latter ($M = 0.42$; SD = 0.04). No other significant differences were observed.

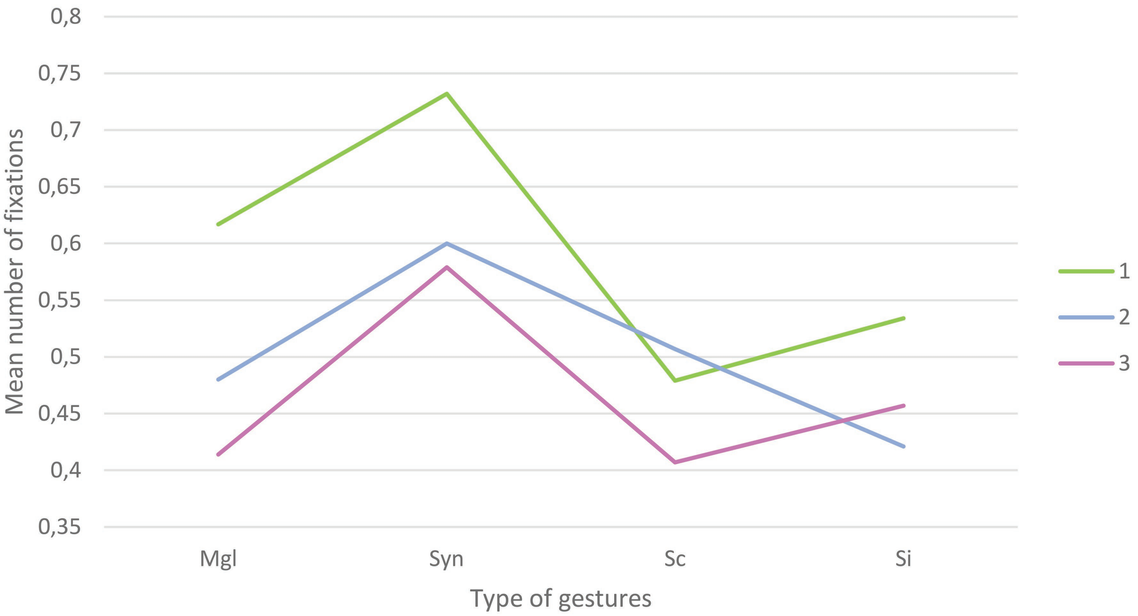


FIGURE 3 | Graphical representation of the interaction for number of fixations between the types of gestures presented and the level of hearing alteration (Mgl = Meaningless; Syn = syntactically incongruent; Sc = semantically congruent; SI = semantically incongruent; 1 = clear speech; 2 = 50% degradation; 3 = 100% degradation).

TABLE 2 | ANOVA results for number of fixations on hand AOI.

Variable	DoF	F	Sig.
Alteration	2	13.80	0.000
Congruency	3	17.58	0.000
Alteration*Congruency	5.59	2.30	0.03

For the conditions where the audio was 100% altered, significant differences were observed between (1) meaningless configurations and syntactically incongruent gestures [$t_{(127)} = -4.18$; $p < 0.001$], with more fixations in the latter ($M = 0.57$; SD = 0.04) compared to the former ($M = 0.41$; SD = 0.04), and (2) syntactically incongruent gestures and semantically congruent [$t_{(127)} = 4.46$; $p < 0.001$] and incongruent [$t_{(127)} = 2.67$; $p < 0.01$] gestures where syntactically incongruent gestures are more fixated ($M = 0.57$; SD = 0.04) than semantically congruent ($M = 0.40$; SD = 0.03) and incongruent ($M = 0.45$; SD = 0.04) gestures.

Percentage of Correct Answer

The repeated measures ANOVA yielded a main effect of Congruency [$F_{(3,381)} = 24.69$; $p < 0.001$], with a higher percentage of correct responses in the semantically congruent condition ($M = 82.32$; SD = 1.18) compared to the meaningless configurations ($M = 74.78$; SD = 1.00), syntactically incongruent ($M = 73.49$; SD = 1.18), and semantically incongruent gestures ($M = 68.18$; SD = 1.16). A main effect of Alteration was also found [$F_{(2,254)} = 149.06$; $p < 0.001$], with a higher percentage of correct responses in the clear sound condition ($M = 86.22$; SD = 0.77) compared to the 50% ($M = 73.73$; SD = 0.93) and 100% ($M = 64.12$; SD = 0.97) alteration conditions. Furthermore, the results also yielded a significant Congruency x Alteration effect [$F_{(6,762)} = 7.16$; $p < 0.001$] reflecting an interaction between the types of gestures presented and the level of hearing alteration (**Figure 4**). A summary of the ANOVA results can be found in **Table 3**.

In the clear sound condition, follow-up paired t tests showed a significant difference between syntactically incongruent gestures and semantically congruent gestures [$t_{(127)} = -2.76$; $p < 0.01$], with a higher percentage of correct answer for the latter ($M = 88.89$; SD = 1.39) compared to the former ($M = 83.2$; SD = 1.60). No other significant differences were highlighted in the clear sound condition.

In the 50% alteration condition, follow-up paired t tests showed a significant difference between meaningless configurations and semantically congruent gestures [$t_{(127)} = -4.28$; $p < 0.001$] with a better correct response percentage in the latter ($M = 83.91$; SD = 1.82) compared to the former ($M = 72.66$; SD = 1.82). A significant difference was also observed between the meaningless configurations condition and the semantically incongruent gestures [$t_{(127)} = 2.93$; $p < 0.01$], with a higher correct answer percentage when in presence of meaningless configurations ($M = 72.66$; SD = 0.82) compared to semantically incongruent ($M = 65.38$; SD = 1.88). The percentage of correct answer was also higher [$t_{(127)} = 2.56$; $p = 0.01$] for syntactically incongruent gestures ($M = 72.98$; SD = 2.12) compared to semantically incongruent gestures ($M = 65.38$; SD = 1.88). A significant difference was also found between the syntactically incongruent gestures and semantically congruent gestures [$t_{(127)} = -3.80$; $p < 0.001$], with better results for the latter

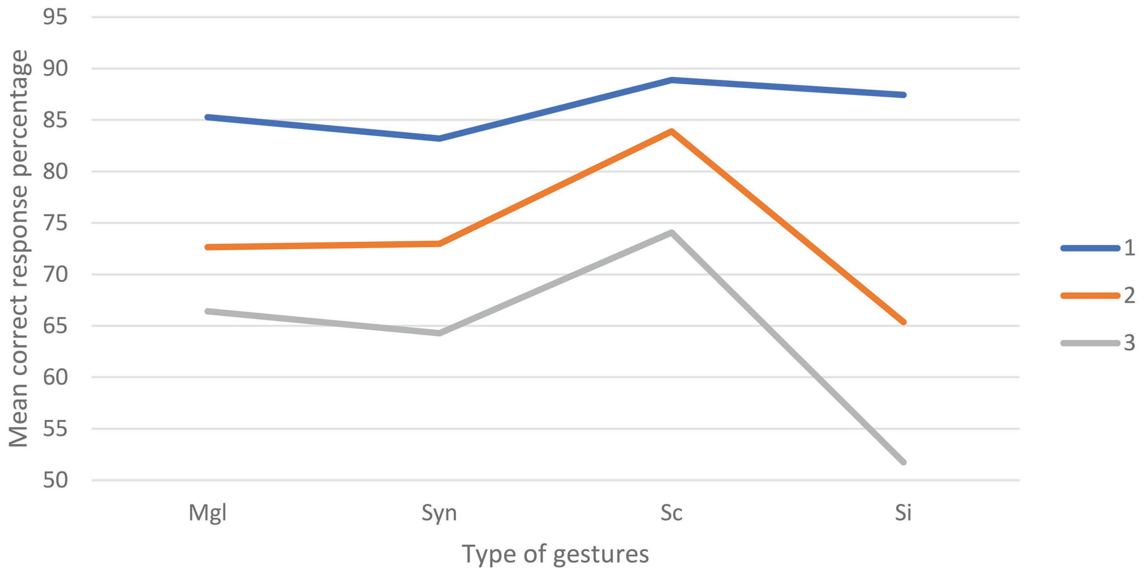


FIGURE 4 | Graphical representation of the percentage of correct answer according to the types of gestures presented and the level of hearing alteration (Mgl = Meaningless; Syn = syntactically incongruent; Sc = semantically congruent; SI = semantically incongruent; 1 = clear speech; 2 = 50% degradation; 3 = 100% degradation).

TABLE 3 | ANOVA results for percentage of correct answers.

Variable	DoF	F	Sig.
Alteration	2	149.06	0.000
Congruency	3	24.69	0.000
Alteration*Congruency	5.35	7.16	0.000

($M = 83.91$; SD = 1.82) compared to the former ($M = 72.98$; SD = 2.12). A higher percentage of correct responses [$t_{(127)} = 7.12$; $p < 0.001$] was also found in the presence of semantically congruent ($M = 83.91$; SD = 1.82) compared to semantically incongruent ($M = 65.38$; SD = 1.88) gestures. No significant differences were found between the meaningless configurations and syntactically incongruent gestures.

In the 100% alteration condition, a significant difference was found between the meaningless configurations and the semantically congruent gestures [$t_{(127)} = -2.94$; $p < 0.01$] with a higher percentage of correct responses in the latter ($M = 74.08$; SD = 2.11) compared to the former ($M = 66.41$; SD = 1.72). Significant differences were also found between meaningless configurations and semantically incongruent gestures [$t_{(127)} = 4.71$; $p < 0.001$], with better performances in the presence of meaningless configurations ($M = 66.41$; SD = 1.72) compared to semantically incongruent gestures ($M = 51.74$; SD = 2.45). Better performances were also highlighted [$t_{(127)} = -3.14$; $p < 0.01$] for semantically congruent gestures ($M = 74.08$; SD = 2.11) compared to syntactically incongruent gestures ($M = 64.28$; SD = 2.1). Semantically incongruent gestures ($M = 51.74$; SD = 2.45) induced a lower percentage of correct responses compared to syntactically incongruent gestures [$t_{(127)} = 3.75$; $p < 0.001$] and semantically congruent gestures [$t_{(127)} = 6.67$; $p < 0.001$].

DISCUSSION

The aim of the study was to investigate participants' visual behavior when confronted with different types of gestures in an (un)favorable listening context and how different types of gestures would influence information uptake. To date, this is the first study exploring these questions together. The main findings show a difference in visual attention allocation depending on type of gesture and on the clarity of the verbal message as well as evidence of information uptake during covert attention. These results suggest (1) that although visual attention is not explicitly focused on the gesture, its presence can affect comprehension (e.g., negatively if semantically incongruent, or positively if semantically congruent) and (2) that not all mismatching gestures are processed equally.

First, this study replicates previous results (Gullberg and Holmqvist, 1999, 2006; Gullberg, 2003; Gullberg and Kita, 2009; Beattie et al., 2010) showing longer and more fixations on the face area compared to hand gesture area, and this, in the presence of any type of gestures. The preferential fixation of the face area is not surprising given the importance of this body part in social interactions (Kanwisher and Yovel, 2006). From an early age, humans are naturally attracted to faces (Johnson et al., 1991) and neuroimaging studies have highlighted brain areas either broadly involved in their processing, such as the medial temporal lobe (Haxby et al., 1996), or specifically dedicated to their processing, such as a small region located in the right lateral fusiform gyrus (McCarthy et al., 1997). The absence of an increase in fixation time to face areas with speech degradation could appear contradictory to previous studies (Saryazdi and Chambers, 2017; Drijvers et al., 2019). However, it is likely that this absence of effect was consequent to the blurring of the mouth area in the present study. In Drijvers and Özyürek's (2020) study, participants appeared to have benefited from the presence of visible speech, particularly when the auditory message was degraded. The alteration in the auditory information led participants to focus more on the lip/mouth region likely to gain phonological information (Drijvers and Özyürek, 2017). Due to the material used, this effect could have not been present in the current study and was indeed not observed.

While no differences in visual attention allocation for the face area were highlighted across the different levels of speech alteration, gestures showed a more complex pattern. First, overall, more attention was paid to gestures in a clear auditory context compared to both degraded speech conditions, replicating previous results (Drijvers et al., 2019). When taking a closer look, participants' gaze toward semantically congruent gestures appeared to be constant throughout the different levels of auditory degradation. For mismatching gestures, this varied depending on the type of incongruency (i.e., semantic and syntactic). The presence of different patterns of visual attention allocation depending on the type of mismatch is consistent with previous electrophysiological research (Friederici et al., 1993) showing different event-related brain potentials in the presence of different types of mismatching information. In the case of meaningless configurations, participants spent more time fixating them when in the presence of clear speech or a 50% alteration compared to when the speech was inaudible. For syntactically incongruent gestures, more time was spent fixating them in a clear auditory context compared to any degradation of speech. Finally, semantically incongruent gestures were more fixated in clear or completely degraded speech compared to the 50% alteration condition. One explanation for this pattern of results resides in the amount of attention required to understand the conveyed message (Wild et al., 2012). In the presence of simple sentences and clear speech, information comprehension does not impose a particular demand on attentional processes (Hartsuiker and Moors, 2017). In the case of meaningless configurations, we can assume the deployment of these processes to attempt extracting meaning. Because these gestures do not, per definition, convey any meaning, their presence should not be disrupting participant's comprehension. Indeed, no differences in correct response percentage were observed between the meaningless configurations condition and the semantically congruent gestures. In a clear speech context, the same observation can be made for syntactically incongruent gestures. In the presence of speech degradation, attentional processes become required to comprehend the conveyed message (Wild et al., 2012). In this situation, the decrease in time spent on syntactically incongruent gestures can be explained by the increased necessity to focus on the auditory input to attempt gaining information. Participants therefore turn their attention away from not only irrelevant but also disrupting information to focus on speech. For processing meaningless configurations, a partial speech degradation appears to be acceptable. However, in the presence of a total degradation of the auditory information, attention is taken away from the gestures and probably captured by attempting to understand the conveyed message. For semantically incongruent gestures, the pattern varies. For these gestures, more time is spent fixating them in the clear speech and in the completely degraded speech conditions compared to the partially degraded speech condition. In the clear condition, participants could have enough attentional resources to attempt resolving the incongruency compared to the partial degradation condition. In the latter, the semantically incongruent gestures are highly disrupting the comprehension process since they convey information that is directly contradictory to the one presented in the faded speech. When the verbal utterance is completely degraded the semantically incongruent gesture, fitting the sentence construction, could be conveying relevant information.

When considering the time spent on fixating gestures alone, the current study highlights an interaction between the degree of auditory alteration and type of gesture presented. Across the different levels of alteration, syntactically incongruent gestures were consistently fixated for a longer time than any other types of gestures. Syntactically incongruent gestures essentially conveyed meaning that presented a syntactic violation in the sentence (i.e., they did not fit into the sentence construction). Exploring language comprehension, a previous electrophysiological study (Hahne and Friederici, 1999), demonstrated the existence of a two-step processing in sentence comprehension, with a first highly automatic parsing process dealing with word category (Friederici, 1995). The gestures presented in the syntactically incongruent condition having been specifically selected to convey a meaning that would not fit their position in the sentence, it is possible that they particularly attracted attention and disturbed processing. In fact, in the clear speech condition, the percentage of correct responses was significantly lower when in presence of these gestures. Furthermore, a previous eye tracking study investigating the differences in the perception and time course of syntactic and semantic violations showed that syntactic violations were detected earlier than semantic violations (De Vincenzi et al., 2003). Although the current study did not explore the time course of fixations, the longer dwelling times on syntactically incongruent gestures could suggest an earlier entry and local attempt to resolve the incongruency (Braze et al., 2002). Interestingly, unlike syntactically incongruent gestures, the presence of meaningless configurations or semantically incongruent gestures did not impair comprehension. When the speech is clear and easily understandable, a syntactic violation thus appears to disturb comprehension at a higher level than semantic violation, even when presented through a gesture. This is associated with an increased amount of time spent fixating these gestures.

The results in the presence of a verbal alteration are more complex. First, although semantically congruent gestures were not particularly more or for longer looked at than other types of gestures, comprehension scores were significantly higher in their presence. As mentioned above, different gaze patterns were observed for the different types of mismatching gestures, along with different levels of comprehension. In both alteration conditions, while more time was spent on fixating syntactically incongruent gestures compared to meaningless configurations, no significant difference in comprehension was highlighted. However, although semantically incongruent gestures were the least fixated of all mismatching gestures, they induced the most incorrect responses in the comprehension task. These results suggest that the presence of gestures did in fact affect comprehension, and an overt allocation of attention them was not required for information uptake. This is inconsistent with the general suggestion of a higher quality of retrieved information in the case of gesture fixation (Gullberg and Holmqvist, 1999) as the presence of semantically incongruent gestures clearly impaired comprehension. It is, however, consistent with previous claims (Gullberg and Kita, 2009) suggesting that attention to gestures is mostly covert and that the uptake of information appears to be independent of fixation.

While this study offers a number of interesting results, several adjustments could be made for future research. First, the current study did not consider complementary iconic gestures. Indeed, the iconic gestures used (i.e., in the congruent condition) were all redundant (i.e., the information contained in the iconic gesture repeats that contained in speech). Future studies could therefore investigate whether different results would be observed for complementary and redundant gestures. Second, we did not differentiate between the types of iconic gestures (i.e., action, shape, position). Because of the potential difference of importance for comprehension between these types of iconic gestures (see Kandana Arachchige et al., 2021), it would be interesting to see whether and how visual attention is distinctively allocated to all of them. Finally, in the current study, although some sentences (16/50 items) retained some meaning in the event of a total degradation (e.g., "He kept his papers in the red boxes he bought from the shop" remains understandable without "red") and others did not (e.g., "He mixed his cement mixture," makes little sense without "mixed"), we do not believe this had an effect on the observed results. Indeed, on the one hand, the majority of the sentence lost meaning in the event of an alteration, and on the other hand, the comprehension statements were specifically designed to investigate the comprehension of the bold item (see **Appendix A**). Nevertheless, future research could differentiate these conditions and verify whether distinguished results would arise from them.

CONCLUSION

To conclude, the current study is the first to show that different types of mismatching gestures differently attract visual attention and differently affect comprehension. Furthermore, as suggested by previous authors, overt visual attention to gestures is not required for information uptake as semantically incongruent gestures significantly impaired comprehension while being the least looked at. In contrast, the presence of semantically congruent gestures significantly aided comprehension although they were among the least fixated.

DATA AVAILABILITY STATEMENT

The data that supports the finding of this study are available from the corresponding author, KK, upon reasonable request.

ETHICS STATEMENT

This study was reviewed and approved by the Ethics committee of the University of Mons. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KK: conceptualization, original draft preparation, and writing and editing. WB: conceptualization and editing. ISL and LL: supervision – reviewing and editing. All authors contributed to the article and approved the submitted version.

FUNDING

KK was funded by a doctoral award from the University of Mons.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.776867/full#supplementary-material

REFERENCES

Beattie, G., and Shovelton, H. (2001). An experimental investigation of the role of different types of iconic gesture in communication: a semantic feature approach. *Gesture* 1, 129–149. doi: 10.1075/gest.1.2.03bea

Beattie, G., and Shovelton, H. (2002). What properties of talk are associated with the generation of spontaneous iconic hand gestures? *Br. J. Soc. Psychol.* 41, 403–417. doi: 10.1348/014466602760344287

Beattie, G., Webster, K., and Ross, J. (2010). The fixation and processing of the iconic gestures that accompany talk. *J. Lang. Soc. Psychol.* 29, 194–213. doi: 10.1177/0261927X09359589

Braze, D., Shankweiler, D., Ni, W., and Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *J. Psycholinguist. Res.* 31, 25–44. doi: 10.1023/A:1014324220455

Calvo, M. G., and Lang, P. J. (2004). Gaze patterns when looking at emotional pictures: motivationally biased attention. *Motiv. Emot.* 28, 221–243. doi: 10.1023/B:MOEM.0000040153.26156.ed

Carrasco, M., and McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proc. Natl. Acad. Sci.* 98, 5363–5367. doi: 10.1073/pnas.081074098

Cocks, N., Morgan, G., and Kita, S. (2011). Iconic gesture and speech integration in younger and older adults. *Gesture* 11, 24–39. doi: 10.1075/gest.11.1.02coc

De Vincenzi, M., Job, R., Di Matteo, R., Angrilli, A., Penolazzi, B., Ciccarelli, L., et al. (2003). Differences in the perception and time course of syntactic and semantic violations. *Brain Lang.* 85, 280–296. doi: 10.1016/S0093-934X(03)00055-5

Drijvers, L., and Özyürek, A. (2017). Visual context enhanced: The joint contribution of iconic gestures and visible speech to degraded speech comprehension. *J. Speech Lang. Hear. Res.* 60, 212–222. doi: 10.1044/2016_JSLHR-H-16-0101

Drijvers, L., and Özyürek, A. (2020). Non-native listeners benefit less from gestures and visible speech than native listeners during degraded speech comprehension. *Lang. Speech* 63, 209–220. doi: 10.1177/0023830919831311

Drijvers, L., Vaitonytė, J., and Özyürek, A. (2019). Degree of language experience modulates visual attention to visible speech and iconic gestures during clear and degraded speech comprehension. *Cogn. Sci.* 43:e12789. doi: 10.1111/cogs.12789

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain Lang.* 50, 259–281. doi: 10.1006/brln.1995.1048

Friederici, A. D., Pfeifer, E., and Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cogn. Brain Res.* 1, 183–192. doi: 10.1016/0926-6410(93)90026-2

Gullberg, M. (2003). "Eye movements and gestures in human face-to-face interaction" in *The Mind's Eye*. eds. J. Hyöna, R. Radach, and Deubel, H (Amsterdam, North-Holland: Elsevier), 685–703.

Gullberg, M., and Holmqvist, K. (1999). Keeping an eye on gestures: visual perception of gestures in face-to-face communication. *Pragmat. Cogn.* 7, 35–63. doi: 10.1075/pc.7.1.04gul

Gullberg, M., and Holmqvist, K. (2001). "Visual attention towards gestures in face-to-face interaction vs. on screen." in *International Gesture Workshop*.

Gullberg, M., and Holmqvist, K. (2006). What speakers do and what addressees look at: visual attention to gestures in human interaction live and on video. *Pragmat. Cogn.* 14, 53–82. doi: 10.1075/pc.14.1.05gul

Gullberg, M., and Kita, S. (2009). Attention to speech-accompanying gestures: eye movements and information uptake. *J. Nonverbal Behav.* 33, 251–277. doi: 10.1007/s10919-009-0073-2

Hahne, A., and Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. *J. Cogn. Neurosci.* 11, 194–205. doi: 10.1162/089892999563328

Hartsuiker, R. J., and Moors, A. (2017). "On the automaticity of language processing," in *Entrenchment and the psychology of language: How we reorganize and adapt linguistic knowledge*. ed. H.-J. Schmid (American Psychological Association, De Gruyter Mouton), 201–225.

Haxby, J. V., Ungerleider, L. G., Horwitz, B., Maisog, J. M., Rapoport, S. I., and Grady, C. L. (1996). Face encoding and recognition in the human brain. *Proc. Natl. Acad. Sci.* 93, 922–927. doi: 10.1073/pnas.93.2.922

Holle, H., and Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *J. Cognit. Neurosci.* 19, 1175–1192. doi: 10.1162/jocn.2007.19.7.1175

Holler, J., Shovelton, H., and Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *J. Nonverbal Behav.* 33, 73–88. doi: 10.1007/s10919-008-0063-9

Johnson, M. H., Dziurawiec, S., Ellis, H., and Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* 40, 1–19. doi: 10.1016/0010-0277(91)90045-6

Kandana Arachchige, K. G., Holle, H., Rossignol, M., Simoes Loureiro, I., and Lefebvre, L. (in press). High verbal working memory load impairs gesture-speech integration: Evidence from a dual-task paradigm. *Gesture*

Kandana Arachchige, K. G., Simoes Loureiro, I., Blekic, W., Rossignol, M., and Lefebvre, L. (2021). The role of iconic gestures in speech comprehension: An overview of various methodologies. *Front. Psychol.* 12. doi: 10.3389/fpsyg.2021.634074

Kanwisher, N., and Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 2109–2128. doi: 10.1098/rstb.2006.1934

Kelly, S. D., Creigh, P., and Bartolotti, J. (2010a). Integrating speech and iconic gestures in a Stroop-like task: evidence for automatic processing. *J. Cogn. Neurosci.* 22, 683–694. doi: 10.1162/jocn.2009.21254

Kelly, S. D., Kravitz, C., and Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain Lang.* 89, 253–260. doi: 10.1016/S0093-934X(03)00335-3

Kelly, S. D., Özyürek, A., and Maris, E. (2010b). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychol. Sci.* 21, 260–267. doi: 10.1177/0956797609357327

Kendon, A. (2004). *Gesture: Visible Action as Utterance.* Cambridge: Cambridge University Press.

Lanata, A., Valenza, G., and Scilingo, E. P. (2013). Eye gaze patterns in emotional pictures. *J. Ambient. Intell. Humaniz. Comput.* 4, 705–715. doi: 10.1007/s12652-012-0147-6

Margiotoudi, K., Kelly, S., and Vatakis, A. (2014). Audiovisual temporal integration of speech and gesture. *Procedia Soc. Behav. Sci.* 126, 154–155. doi: 10.1016/j.sbspro.2014.02.351

McCarthy, G., Puce, A., Gore, J. C., and Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *J. Cogn. Neurosci.* 9, 605–610. doi: 10.1162/jocn.1997.9.5.605

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought.* Chicago: University of Chicago press.

McNeill, D. (2008). *Gesture and Thought.* Chicago: University of Chicago press.

Momsen, J., Gordon, J., Wu, Y. C., and Coulson, S. (2020). Verbal working memory and co-speech gesture processing. *Brain Cogn.* 146:105640. doi: 10.1016/j.bandc.2020.105640

Moore, T., and Zirnsak, M. (2017). Neural mechanisms of selective visual attention. *Annu. Rev. Psychol.* 68, 47–72. doi: 10.1146/annurev-psych-122414-033400

Nobe, S., Hayamizu, S., Hasegawa, O., and Takahashi, H. (1997). "Are listeners paying attention to the hand gestures of an anthropomorphic agent? An evaluation using a gaze tracking method." in *International Gesture Workshop*.

Obermeier, C., and Gunter, T. C. (2014). Multisensory integration: the case of a time window of gesture-speech integration. *J. Cogn. Neurosci.* 27, 292–307. doi: 10.1162/jocn_a_00688

Richey, C., Barrios, M. A., Armstrong, Z., Bartels, C., Franco, H., Graciarena, M., et al. (2018). Voices obscured in complex environmental settings (voices) corpus. arXiv preprint arXiv: 1804.05053.

Rigoulot, S., and Pell, M. D. (2012). Seeing emotion with your ears: emotional prosody implicitly guides visual attention to faces. *PLoS One* 7:e30740. doi: 10.1371/journal.pone.0030740

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., and Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb. Cortex* 17, 1147–1153. doi: 10.1093/cercor/bhl024

Saryazdi, R., and Chambers, C. G. (2017). "Attentional factors in listeners' uptake of gesture cues during speech processing." in Paper presented at the INTERSPEECH.

Sekine, K., Sowden, H., and Kita, S. (2015). The development of the ability to semantically integrate information in speech and iconic gesture in comprehension. *Cogn. Sci.* 39, 1855–1880. doi: 10.1111/cogs.12221

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta Psychol.* 135, 77–99. doi: 10.1016/j.actpsy.2010.02.006

Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S., and Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding their visual attention. *Dev. Sci.* 21:e12664. doi: 10.1111/desc.12664

Wang, B., and Theeuwes, J. (2020). Salience determines attentional orienting in visual selection. *J. Exp. Psychol. Hum. Percept. Perform.* 46, 1051–1057. doi: 10.1037/xhp0000796

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., and Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *J. Neurosci.* 32, 14010–14021. doi: 10.1523/JNEUROSCI.1528-12.2012

Wu, Y. C., and Coulson, S. (2014). A psychometric measure of working memory capacity for configured body movement. *PLoS One* 9:e84834. doi: 10.1371/journal.pone.0084834

Zhao, W., Riggs, K., Schindler, I., and Holle, H. (2018). Transcranial magnetic stimulation over left inferior frontal and posterior temporal cortex disrupts gesture-speech integration. *J. Neurosci.* 38, 1891–1900. doi: 10.1523/JNEUROSCI.1748-17.2017

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 24 January 2022
doi: 10.3389/fpsyg.2021.701281

Effect of Encoding on Prospective Memory

Youzhen Chen[1]†, Manman Zhang[1]*†, Cong Xin[2], Yunfei Guo[3], Qin Lin[1], Zhujun Ma[1], Jinhui Hu[1], Weiting Huang[1] and Qianfang Liao[1]

[1] Department of Psychology, Fujian Normal University, Fuzhou, China, [2] School of Psychology, Nanjing Normal University, Nanjing, China, [3] Department of Education, Henan University, Kaifeng, China

Event-based prospective memory (ProM) refers to remembering to execute planned actions in response to a target ProM cues. Encoding modality influences ProM performance; visual encoding has been studied more than auditory encoding. Further, it has not yet been examined whether different encoding may influence ProM performance in different encoding modalities. This study examines the effects of encoding modality (visual vs. auditory), cue-encoding specificity (specific cue vs. non-specific cue), and encoding modes (standard vs. implementation intention) on event-based ProM tasks. In Experiment 1, cue specificity and encoding modality were manipulated as a within-groups encoding of visual cues is more commonly and between-groups variable. Results revealed the facilitative effect of cue specificity on ProM performance. Also, with respect to encoding modality, participants showed better performance when receiving auditory instructions compared with the visual encoding condition. In Experiment 2, as in Experiment 1, cue specificity and encoding modality were manipulated. Encoding mode was added as a new between-group variable. Result revealed that there was a significant interaction between encoding modality and encoding modes. Visual implementation intention encoding was a more effective method for improving ProM performance compared with visual standard encoding. Furthermore, there was a significant interaction between cue-encoding specificity and encoding modes. Implementation intention encoding enhances ProM performance in non-specific cue-encoding conditions. Overall, the present study found that (1) auditory encoding modality showed superior ProM performance compared with visual encoding, although implementation intention had facilitative on ProM performance regardless of the encoding modalities, and (2) there was better ProM performance under specific encoding compared with non-specific encoding, and implementation intention had a facilitative effect on ProM performance in the non-specific condition.

Keywords: prospective memory, implementation intention encoding, specific encoding, non-specific encoding, visual encoding, auditory encoding

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Naveen Kashyap,
Indian Institute of Technology
Guwahati, India
Sanjram Premjit Khanganba,
Human Factors and Applied
Cognition Lab, Indian Institute
of Technology Indore, India

****Correspondence:***
Manman Zhang
18359140166@163.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: *27 April 2021*
Accepted: *03 December 2021*
Published: *24 January 2022*

Citation:
Chen YZ, Zhang MM, Xin C, Guo YF, Lin Q, Ma ZJ, Hu JH, Huang WT and Liao QF (2022) Effect of Encoding on Prospective Memory.
Front. Psychol. 12:701281.
doi: 10.3389/fpsyg.2021.701281

INTRODUCTION

Prospective memory (ProM) is a memory of an action that refers to executing a delayed intended action in the appropriate context or at a planned time (Sanjram and Khan, 2011). ProM, such as sending a letter when you pass a mailbox, is fundamental to our lives. ProM includes four processes, namely, encoding, maintaining, retrieving intentions, and executing an intended action in the future. The previous ProM studies focused primarily on the involvement in maintenance and retrieval processes in performing a delayed intended action. Only recently have researchers started to explore encoding strategies but mostly focused on visual encoding, and the auditory encoding modality process remains somewhat neglected. The present study focuses on the encoding process, as encoding is a necessary precondition for successfully executing a ProM task.

In laboratory studies, ProM manipulation of encoding is conventionally achieved by providing instructions to participants. Participants were informed of intentions through a verbal or written introduction (Cai and Dong, 2012; Li et al., 2016). Regarding information processing, auditory and visual stimuli are the two most common and independent ways through which humans receive information. The visual pathway leads from the occipital cortex to the inferior temporal cortex, and the auditory pathway leads from the superior temporal cortex to the ventral prefrontal regions (Maeder et al., 2001). The separation of process streams may affect ProM encoding differently. Vedhara et al. (2004) randomly allocated participants to one of four habitual ProM task conditions, namely, no cues, visual cues, auditory cues, or dual cues (auditory and visual cues), of which the last three conditions were operated by visual, auditory, or dualhabitual ProM cue introductions reminder. The results showed that habitual ProM of participants was optimal in the dual-cue condition, suggesting this condition was the most beneficial to habitual ProM performance. Additionally, participants performed better under auditory than visual cue conditions. However, Yang et al. (2008) confirmed that there was no difference between auditory and visual encoding on activity-based ProM for undergraduates, regardless of whether ProM intention was important. Both studies on the effect of the ProM encoding modality have produced inconsistent results. It may be caused by the heterogeneity of subjects in the two studies and the different control of the experiment. Vedhara et al. (2004) adopted a natural experiment with old adults with type 2 diabetes as participants; Yang et al. (2008) adopted a laboratory experiment with young adults as participants. Therefore, it is necessary to explore whether auditory encoding is different from visual encoding. Chen et al. (2014) found that the frontal lobe was activated and engaged in monitoring ProM targets. The auditory pathway and ProM task involved the same brain region—the frontal lobe, so we hypothesized that ProM performance would be better under the auditory encoding modal.

People commonly give specific instructions, but they can also give unclear instructions. Both instructions consist of cue-encoding specificity. Previous research showed that participants perform better in ProM tasks under conditions where they could encode visual specific cues (e.g., "jaguar," "lion," and "tiger") rather than visual general cues (e.g., "animal"; Einstein et al., 1995). Kuhlmann and Rummel (2014) confirmed that intentions encoded by visual specific cues form a tight encoding trace, allowing participants to flexibly allocate cognitive resources, thereby enhancing ProM performance (Kuhlmann and Rummel, 2014). For encoding, these previous studies (e.g., Ball and Bugg, 2018) showed a "specific advantage", indicating that the memory content comprised mostly specific events with a high proportion of specific memories; to recall this content, there was an advantage to remembering specific details (Chen, 2013). Furthermore, Scullin et al. (2018) systematically investigated the encoding process for ProM and confirmed that specific was better than non-specific cue encoding, for which 22.5% of participants gave little thought to the ProM tasks and tended to translate categories to specific exemplars (Pereira et al., 2018). It may be that participants using non-specific cue encoding had to pay closer attention than those using specific cue encoding to correctly determine ProM targets.

However, all of the above studies compared visually specific and non-specific conditions and did not provide direct evidence of auditory specificity encoding. In the current study, we tested the effects of encoding specificity and encoding modality with undergraduate students, building on these different findings. In the first experiment, we investigated whether there was a specific advantage both in the visual and auditory encoding process and whether non-specific auditory encoding was better than non-specific visual encoding in ProM.

We hypothesized that a manipulation, which reduces cognitive resource requirements by enhancing both visual and auditory target cue specificity, would improve ProM, and the performance in the auditory-specific encoding is better than in the visual-specific encoding. Additionally, we hypothesized that participants would perform better in the auditory non-specific content than in the visual non-specific content.

EXPERIMENT 1

Methods

Participants

The sample size was based on an *a priori* power analysis using the GPOWER 3 software. The effect size *f* was based on previous research (Pereira et al., 2018). The alpha level was 0.05, power was 0.95, and an effect size of 0.5 was considered. To find a statistically significant effect in the model, 54 participants would be necessary. Thus, the goal sample size was 60, to account for dropouts.

The initial sample included 60 undergraduate students ($M_{\text{age}} = 20.55$ years, range 18–25; 27.12% males). Each participant had normal or correct-to-normal vision and audition, and none had previously taken part in a similar experiment. Four participants were excluded from the analysis because their parameter estimates of the ongoing task or their ProM task were more than three standard deviations (SDs) from the mean of their respective group. Thus, there were 56 participants in total ($N_{\text{auditory}} = 29$, $N_{\text{visual}} = 27$).

Participants provided signed informed consent before the experiment. The only demographic information collected from the participants was their age and gender; no names or personal information were recorded. Participants were given a small gift as compensation.

Experimental Design

The experiment had a two encoding modality (visual vs. auditory) × two cue-encoding specificity (specific cue vs. non-specific cue) mixed factorial design with the second factor as a within-groups variable. Participants received verbal instructions through earphones and written texts for auditory modality and visual modality, respectively. Instructions were provided until participants fully understood them. The within-subjects variable was counter-balanced among participants to avoid a practice effect. Half of the participants started with the specific cue-encoding block, and the other half started with the non-specific cue-encoding block.

Materials

The experimental stimuli were 20 capital English letters (excluding "A," "E," "I," "O," "F," and "J"). The letters "A," "E," "I," and "O" were excluded to balance-specific and non-specific cue conditions, and "F" and "J" were omitted because they were reaction keys for the ongoing task. The experimental trials were presented visually in a random order. The ProM target was presented six times. To match the number of presentations of specific and non-specific cues, the ProM cues were the same in all specific conditions. Whether the cue was specific or not, we set "U" as the target. Both visual and auditory ProM targets were used for non-salient cues, to avoid the interference effects of cue salience (see Kliegel et al., 2013). All items were presented in white, Courier New, 60-point font on a black background.

Experimental Task

Throughout the experimental task, participants performed two tasks simultaneously, namely, the ongoing task and the ProM task. The ongoing task was a one-back task, in which participants needed to compare the present letter with the previous one and then press "J" if they were the same and "F" if they were different. In the ProM task, participants were asked to press the spacebar for a ProM target.

Procedure

This experiment was programmed using the E-Prime 1.1 software. All participants were tested individually in a quiet environment. Participants were initially told that the research goal was to study people's performances on various computer-based tasks and gauge their thoughts during those tasks.

In the first step, participants performed a practice block (see **Figure 1**). They received instructions for the one-back task (i.e., ongoing task). For each trial, a fixation cross (+) was presented for 500 ms, followed by a letter for 3,000 ms. The letter would disappear once the participant responded or 3,000 ms had elapsed. Participants received feedback on their accuracy. The "yes" and "no" response keys decisions were counterbalanced, with each decision taking up an equal proportion of the trials.

After participants completed 50 practice trials, they received instructions for simultaneously completing the ProM and ongoing tasks in the normal experimental blocks, i.e., the intention encoding phase. In the visual-specific encoding condition, instructions were presented in the form of text, and encoding content was as follows: "The vowel U may appear during the block. When you see it, you do not need to compare whether the letter is the same as the previous one. Press the spacebar directly." In the visual non-specific encoding condition, instructions were also presented in the form of text, and encoding content was as follows: "Vowels may appear during the block. When you see a vowel, you do not need to compare whether the letter is the same as the previous one. Press the spacebar directly." For the audio-specific encoding condition, instructions were played in the form of voice (sound files were embedded in E-prime), and encoding content was the same as the visual-specific encoding condition. When assigned to audio non-specific encoding condition, participants were also played instructions in the form of voice, and encoding content was the same as visual non-specific encoding condition.

Then, we asked participants to repeat the instructions to the experimenter to demonstrate if they understood them. To prevent participants from monitoring the ProM target letters, participants were told that the ProM and ongoing tasks were equally important.

Following the ProM encoding process (i.e., ProM instructions in the different conditions), participants completed an interference task of simple digital arithmetic and then began the formal experiment, which included 94 trials of the ongoing task and six presentations of the ProM target ("U") for per block. The non-specific and specific cue-encoding blocks were the same, with the exception that the ProM instructions required participants to press the spacebar for vowels. Between the specific and non-specific encoding blocks, there were an interference task and a 2-min rest period.

Results

For all analyses, the alpha level was set at 0.05. Unless otherwise noted, the dependent variables were reaction time (RT) and accuracy in the ProM target and ongoing trials.

Prospective Memory Task Performance

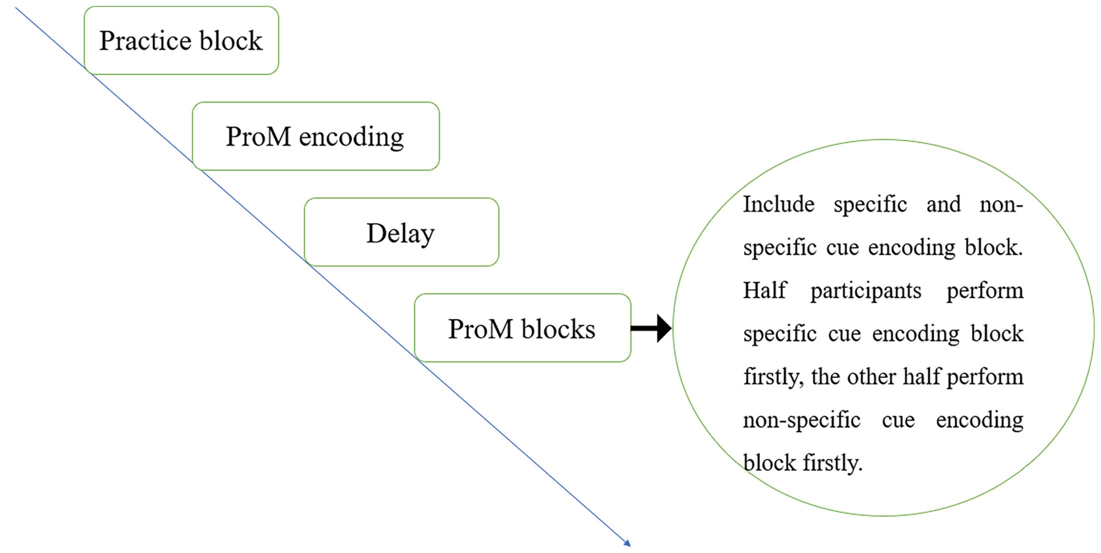
Prospective memory hits occurred when a participant pressed the spacebar on the ProM target trials. ProM performance was defined as the number of hits divided by the number of target events. The means and SDs are presented in **Table 1**.

To examine the ProM hit, we conducted a two encoding modality (visual vs. auditory) × two (specific vs. non-specific) mixed analysis of variance (ANOVA) with the last variable as a within-groups variable. There was a significant main effect of encoding modality, $F(1,54) = 5.67$, $p = 0.021 < 0.05$, $\eta_p{}^2 = 0.10$; the accuracy in the auditory encoding modality was higher than visual encoding modality. Results also showed a significant main effect of cue-encoding specificity, $F(1,54) = 6.65$, $p = 0.013 < 0.05$, $\eta_p{}^2 = 0.11$. As **Table 1** indicates, the accuracy in specific cue-encoding condition was significantly higher than the non-specific cue-encoding condition. There was no significant interaction effect for encoding modality and cue specificity, $F(1,54) = 0.03$, $p = 0.87$.

A 2 × 2 mixed ANOVA was conducted with RTs for correct responses to ProM targets as the dependent variable. There were no significant main effects or interactions, $Fs < 0.44$, $ps > 0.24$.

Ongoing Task Performance

The results of both accuracy and RT data are displayed in **Table 1**. To analyze the performance of the ongoing task, we conducted a two (encoding modality: visual vs. auditory) × two



Include specific and non-specific cue encoding block. Half participants perform specific cue encoding block firstly, the other half perform non-specific cue encoding block firstly.

FIGURE 1 | The procedure for Experiment 1.

(cue-encoding specificity: specific cue vs. non-specific cue) mixed factorial ANOVA with the last variable as the within-subjects variable and with the ongoing task accuracy and RTs as dependent variables separately.

The analysis of ongoing task accuracy showed that there were no significant main effects of cue encoding modality and cue-encoding specificity, and no interaction effect between cue encoding modality and cue-encoding specificity, and no interaction effect between cue specificity and encoding modality, $F(1,54) < 2.92$, $ps > 0.09$.

The analysis of ongoing task RTs revealed a main effect of cue-encoding specificity, $F(1,54) = 7.29$, $p = 0.009 < 0.05$, $\eta_p{}^2 = 0.12$, suggesting that participants made yes-or-no decisions more faster during the specific cues encoding blocks compared with non-specific cues encoding blocks in both the auditory and visual conditions. There were no other significant differences in RT between the auditory and visual encoding modalities, $F(1,54) = 0.09$, $p = 0.77 > 0.05$, and no interaction effect between cue-encoding specificity and encoding modality, $F(1,54) = 1.83$, $p = 0.18 > 0.05$.

Discussion

Notably, there was a significant difference in ProM performance of visual and auditory encoding in Experiment 1. Participants had a higher accuracy in the auditory encoding condition than in the visual encoding condition, which was consistent with our hypothesis and confirmed our suspicion that auditory encoding was more convenient. Auditory encoding information is transmitted to the ventral prefrontal regions (Maeder et al., 2001), which happens to be the lobe activated by the ProM intention. The common lobe may make auditory encoding more advantageous. However, the discovery was inconsistent with the conclusion of Yang et al. (2008), who concluded that participants could successfully complete ProM tasks regardless of whether they used visual encoding or auditory encoding, as long as they formed the correct ProM intention. The reason for the different results may be that Yang et al. (2008) explored the effect of encoding on an activity based on ProM, which did not require pausing ongoing tasks and was relatively simple. However, our study explored the effect of encoding on event-based ProM, which required participants to pause ongoing tasks and translate recognitive cost to the ProM task. The difficulty of ProM task may be an important factor.

In Experiment 1, we found advantages of specific cue encoding for ProM, which was consistent with our hypothesis and in line with Einstein et al. (1995) and Hicks et al. (2005), that participants can detect more ProM targets in the specific condition than in non-specific intentions. Participants performed significantly faster in ongoing tasks in the specific condition relative to the non-specific condition, which was consistent with the conclusion of Hicks et al. (2005) that non-specific intentions caused more task interference than specific intentions to the ongoing task. However, RTs of ProM tasks have no significant differences, suggesting possessing two intentions (i.e., ProM intention and ongoing task intention) does not exert the same

TABLE 1 | The performance of dual tasks in different encoding modalities and encoding specificity *M (SD)*.

Task conditions	**ProM tasks**		**Ongoing tasks**	
	ACC	**RT (ms)**	**ACC**	**RT (ms)**
Visual specific	0.78 (0.17)	770 (148)	0.92 (0.06)	670 (138)
Visual non-specific	0.64 (0.35)	828 (215)	0.91 (0.07)	759 (166)
Auditory specific	0.87 (0.16)	820 (155)	0.92 (0.06)	710 (132)
Auditory non-specific	0.75 (0.29)	831 (206)	0.90 (0.08)	740 (201)

ACC, accuracy; ProM, prospective memory; RT, reaction time.

costs as each would exert individually; the slow RTs of the ongoing task do not necessarily preserve ProM performance (Hicks et al., 2005).

Our finding indicated that participants in the specific cue-encoding condition performed better than those in the non-specific cue-encoding condition, and the RTs of the ongoing task were shorter in the specific cue-encoding condition than in the non-specific cue-encoding condition, which somewhat differs from those of Scullin et al. (2018), who found that some ProM encoding components can be done "in passing" (i.e., a perfunctory and transient manner). We can infer that the encoding process requires cognitive resources, and specific cue encoding is more effective than non-specific cue encoding. Although Scullin et al. (2018) were more focused on thought probes during ProM encoding instead of RTs for retrieval and manipulated only non-specific encoding, their results were consistent with those of the present study in suggesting that specific encoding is more convenient and efficient than non-specific encoding, regardless of cue focality. Since non-specific cue encoding is a common and unavoidable phenomenon, we conducted Experiment 2 to investigate how to improve ProM performance using non-specific cue encoding.

Taken together, the results of Experiment 1 showed that auditory encoding was superior to visual encoding, and specific cue encoding was superior to non-specific cue encoding in terms of ProM intention.

EXPERIMENT 2

In light of the findings of Experiment 1 concerning inferior performance, Experiment 2 further investigated how implementation intention could improve ProM performance under visual and non-specific encoding conditions. Meeks and Marsh (2010) reported that the intention encoding of visual implementation was an efficient encoding method. Thus, we performed Experiment 2 to explore whether implementation intention could improve performance on ProM tasks in the visual and non-specific encoding conditions.

Implementation intention encoding is a conscious formation of a specific intention and a response to a specific stimulus (Gollwitzer and Peter, 1999; Scott, 2016) that determines when, where, and how to put a goal into action based on the goal intention. Implementation intention encoding consisted of a typical statement in the form of "if situation x occurs, then I will perform intended action y." The statement was often accompanied by asking participants to mentally visualize (usually for 30 s) the intended action and repeating the implementation intention instructions to the researcher. McFarland and Glisky (2012) argued that the verbal articulation of implementation intention is sufficient to improve ProM performance, and imagery instruction is unnecessary. Sculling et al. (2017) reported that a verbal statement and an imagery procedure for implementation intention significantly increased the generation of high typicality exemplars, suggesting that verbally repeating instructions and imagery procedures for implementation intention produced the same results. However, imagining the context of implementation intention is more difficult to control and may result in additional variables (Brewer et al., 2011). Therefore, the present study asks participants to repeat ProM introductions two times (Guo et al., 2016). Conversely, in standard encoding condition (hereafter referred to as standard encoding in this study), participants are told what they should do but without asking them to form an implementation intention or other intention encoding (Mcdaniel et al., 2008; Meeks and Marsh, 2010; Rummel et al., 2012), and the researchers check with participants to make sure they know what to do without an articulated strategy or approach to the task. Implementation intention differs from standard encoding in that it asks participants to consciously encode ProM targets and ProM intended behaviors, which forges connections between ProM targets and actions and does not utilize the cognitive resources of ongoing tasks.

By enhancing the automated connection between ProM cues and actions, implementation intention has been shown to improve ProM performance (Chasteen et al., 2010; Zimmermann and Meier, 2010; Rummel et al., 2012; Chen et al., 2015; Guo et al., 2016). For example, Lin (2016) found that compared with standard encoding, implementation intention improved the accuracy of ProM but not the RTs, regardless of cognitive load. Lv (2010) compared the differences between implementation intention and standard encoding, concluding that implementation intention strengthened the relationship between ProM cues and responses to ProM targeted actions, contributing to participants performing ProM tasks automatically without cognitive resources.

Above all, existing research has focused on intention encoding of visual implementation in the specific cue-encoding condition and justified that the intention of visual implementation was useful for enhancing ProM; however, it remains unclear whether auditory implementation intention encoding is as effective as visual implementation intention encoding and whether the advantage of implementation intention can be generalized to the non-specific cue-encoding condition. Experiment 2 hypothesize that implementation intention can improve ProM performance in the visual and non-specific encoding conditions, and auditory implementation intention is also an efficient encoding method.

Methods

Participants

The sample size was based on an *a priori* power analysis using the GPOWER 3 software. The effect size *f* was based on previous research (Pereira et al., 2018). The alpha level was 0.05, power was 0.95, and a size effect of 0.5 was considered. To find a statistically significant effect in the model, 76 participants would be necessary. A goal sample size of 120 was set to account for dropouts.

Experiment 2 was conducted with 120 undergraduate students (M_{age} = 20.64, range: 18–26; 32.20% males) who received a gift for participating. Two participants were excluded because their RT data were ± 3 SDs from mean of their group. There were 62 participants in the implementation intention conditions (i.e., 30 in the visual condition and 32 in the audio condition) and 56 participants in the standard encoding conditions (i.e., 27 in the visual condition and 29 in the audio condition). The other conditions were the same as Experiment 1.

Design

The experiment had a two encoding modality (visual vs. auditory) × two encoding modes (standard vs. implementation intention) × two cue-encoding specificity (specific cue vs. non-specific cue) mixed factorial design with the last factor as the within-groups variable. The cue specificity order was balanced among participants: half executed the specific cue condition first, while the other half executed the non-specific cue condition first. RTs and accuracy rates served as the dependent measures.

Materials

We used an Acer computer and a 14.5-inch CRT display. The experiment was programmed using the E-Prime 1.1 software. All details regarding the stimuli were consistent with Experiment 1.

Procedure

The instructions given to participants describing the one-back task were almost identical to those in Experiment 1, with the only difference that participants were encoded by implementation intention in the implementation intention condition. For both the auditory and visual implementation intention conditions, participants were asked to repeat the implementation intention instructions two times. The implementation intention instructions were, "If you see any of the vowels, press the spacebar directly" or "If you see the letter U, press the spacebar directly." The standard condition instructions were, "The letter U may appear during the experiment. When you encounter the letter U, it is not necessary to compare whether the letters are the same or not, just press the spacebar." Participants in the standard encoding condition were not required to repeat the instructions about the ProM task. The remaining instructions were the same as in Experiment 1.

Results

Prospective Memory Task Performance

Table 2 shows the means and SDs for ProM tasks. An ANOVA was carried out for (encoding modality: visual vs. auditory) × two encoding modes (standard vs. implementation intention) × two (cue-encoding specificity: specific cue vs. non-specific cue) mixed factorial design with the last factor as within-groups variable on ProM accuracy. Results yielded a significant main effect of encoding specificity, $F(1,114) = 5.02$, $p = 0.027 < 0.05$, $\eta_p{}^2 = 0.04$, indicating that the accuracy rate was significantly higher in the specific cue-encoding condition ($M = 0.82$, $SD = 0.02$) than in the non-specific cue-encoding condition ($M = 0.75$, $SD = 0.30$).

We found a marginal significant interaction between encoding mode and encoding specificity, $F(1,114) = 3.40$, $p = 0.068$, $\eta_p{}^2 = 0.03$. Further simple effects analyses revealed that there was a significant difference between implementation intention and standard encoding in the non-specific cue-encoding condition, $F(1,114) = 4.08$, $p = 0.046 < 0.05$, $\eta_p{}^2 = 0.04$, suggesting that implementation intention ($M = 0.80$, $SD = 0.04$) had a higher accuracy than standard encoding ($M = 0.81$, $SD = 0.02$). There was no significant difference between implementation intention and standard encoding in the specific cue-encoding condition, $p = 0.69 > 0.05$. In the standard encoding condition, there was a significant difference between the specific cue-encoding and non-specific cue-encoding condition, $F(1,114) = 7.94$, $p = 0.006 < 0.05$, $\eta_p{}^2 = 0.07$, suggesting that specific cue-encoding condition ($M = 0.83$, $SD = 0.03$) had higher accuracy than non-specific cue encoding ($M = 0.69$, $SD = 0.04$). There was no significant difference between the specific cue encoding and non-specific cue encoding condition in the implementation intention encoding condition, $p = 0.78 > 0.05$.

We also found a significant interaction between encoding mode and encoding modality, $F(1,114) = 7.56$, $p = 0.007 < 0.05$, $\eta_p{}^2 = 0.07$. The further simple effects analysis found there was a significant difference between standard encoding and implementation intention encoding in the visual encoding condition, $F(1,114) = 8.85$, $p = 0.004 < 0.05$, $\eta_p{}^2 = 0.07$, suggesting that participants had a higher ProM accuracy in the implementation intention encoding condition ($M = 0.81$, $SD = 0.03$) than the standard encoding condition ($M = 0.71$, $SD = 0.03$). In the auditory encoding condition, there was no significant difference between standard encoding and implementation intention encoding, $p = 0.38 > 0.05$. We found a significant difference between the visual and auditory encoding condition in the standard encoding condition, $F(1,114) = 5.38$, $p = 0.02 < 0.05$, $\eta_p{}^2 = 0.05$, suggesting that participants had higher ProM accuracy in the auditory encoding condition ($M = 0.81$, $SD = 0.03$) than the visual encoding condition ($M = 0.71$, $SD = 0.03$). In the implementation intention encoding condition, there was no significant difference between the visual and auditory encoding condition, $p = 0.12 > 0.05$. There were no other significant main effects or interaction effects, $Fs < 2.34$, $ps > 0.13$.

Prospective memory RT of correct responses (i.e., trials where participants correctly responded to the PM targets) were analyzed by performing an ANOVA for two encoding modality (visual vs. auditory) × two encoding modes (standard vs. implementation intention) × two cue-encoding specificity (specific cue vs. non-specific cue) mixed factorial design with last factor as within-groups variable. There was a significant main effect of cue specificity, $F(1,114) = 11.88$, $p = 0.001 < 0.05$, $\eta_p{}^2 = 0.09$, indicating RTs were much faster in the specific cue condition ($M = 787$, $SD = 13.92$) versus the non-specific cue condition ($M = 851$, $SD = 18.43$). There were no other significant main effects or interaction effects, $Fs < 2.77$, $ps > 0.09$.

Ongoing Task Performance

To analyze performance in the ongoing task, we conducted a two encoding modality (visual vs. auditory) × two encoding modes (standard vs. implementation intention) × two cue-encoding specificity (specific cue vs. non-specific cue) mixed factorial design with the last factor as within-groups variable on accuracy and RT of ongoing task, respectively. **Table 3** shows the means and SDs of ongoing task trials.

The analysis of accuracy of ongoing tasks found that the only significant finding was a main effect of cue-encoding specificity, $F(1,114) = 10.33$, $p = 0.002 < 0.05$, $\eta_p{}^2 = 0.08$, showing a higher accuracy on specific cues ($M = 0.92$, $SD = 0.01$) compared with

TABLE 2 | The performance of ProM tasks in different conditions *M (SD)*.

Encoding models	Encoding modality	Encoding specificity			
		Specific encoding		Non-specific encoding	
		ACC	RT (ms)	ACC	RT (ms)
Standard encoding	Audio	0.87 (0.16)	820 (155)	0.75 (0.29)	831 (207)
	Visual	0.78 (0.17)	771 (148)	0.64 (0.35)	828 (216)
Implementation intention	Audio	0.78 (0.23)	756 (166)	0.77 (0.29)	832 (192)
	Visual	0.84 (0.17)	780 (131)	0.83 (0.22)	913 (186)

TABLE 3 | The performance of ongoing tasks in different conditions *M (SD)*.

Encoding models	Encoding modality	Encoding specificity			
		Specific encoding		Non-specific encoding	
		ACC	RT (ms)	ACC	RT (ms)
Standard encoding	Audio	0.92 (0.06)	710 (132)	0.90 (0.08)	740 (201)
	Visual	0.92 (0.06)	670 (139)	0.91 (0.07)	759 (166)
Implementation intention	Audio	0.92 (0.05)	671 (142)	0.90 (0.06)	732 (169)
	Visual	0.94 (0.04)	703 (138)	0.92 (0.05)	805 (162)

non-specific cues (M = 0.91, SD = 0.01). There were no other significant main or interaction effects, $Fs < 1.85$, $ps > 0.17$.

Results of RT of ongoing tasks yielded a significant main effect of cue specificity, $F(1,114) = 21.32$, $p = 0.000 < 0.001$, $\eta_p^2 = 0.16$, with participants demonstrating faster RTs in the specific (M = 689, SD = 12.70) than the non-specific condition (M = 759, SD = 16.15). There were no other significant main or interaction effects, $Fs < 2.67$, $ps > 0.11$.

Discussion

Consistent with previous studies (e.g., Scullin et al., 2017) and our hypothesis, Experiment 2 confirmed that visual implementation intention was effective for improving ProM. The result of Experiment 2 showed there was a significant difference in ProM accuracy between visual implementation intention encoding and visual standard encoding, indicating that visual implementation intention can improve ProM accuracy relative to visual standard encoding.

We did not find specificity advantage in the implementation intention encoding condition, but we found participants had a higher ProM accuracy in the implementation intention encoding condition than the standard encoding condition in the non-specific cue-encoding condition, which was consistent with our hypothesis and suggested that implementation intention can improve the performance in the non-specific encoding condition and extent the finding that implementation intention was suitable to non-specific cue encoding. We found no significant difference between the standard encoding and implementation intention encoding condition in specific cue encoding, which was consistent with Meeks and Marsh (2010), who found that specific cues (e.g., "deer" and "cow") led to a ceiling effect in both implementation intention (imagery; imagery + when-then) and conventional event-based specific cue ProM conditions; thus, there was no opportunity to investigate benefits of implementation intention. Both Meeks and Marsh (2010) and the present study demonstrated that non-specific implementation intention can improve ProM performance.

The current finding of no significant difference of ongoing tasks in different encoding modes was somewhat in line with Lv (2010), who found that under implementation intention conditions, reduced ongoing task performance did not improve ProM task performance. The finding that implementation intention did not reduce ongoing task performance of participants relative to standard instructions contrasted Chen et al. (2015), who reported that individuals had longer RTs for ongoing tasks when encoded by implementation intention. We speculated that this disparity could be explained by the relatively simple ongoing task used in the current study; the further study can manipulate the two-back task as the ongoing task.

GENERAL DISCUSSION

The result of Experiment 2 showed that there was a significant difference between the visual and auditory modalities in the standard encoding condition, which was consistent with Experiment 1 that proved that auditory standard encoding was more useful than visual standard encoding. The finding confirmed our hypothesis that auditory encoding and ProM process mechanism share the common lobe, which is beneficial for auditory encoding. However, our results are inconsistent with those of Yang et al. (2008), who demonstrated that encoding modality had no effect on activity-based ProM. Yang et al. (2008) explored whether sensory modality (auditory and visual) influences ProM encoding, by providing different sensory encoding modalities.

The results showed that irrespective of whether audio or visual information was encoded, intentions of participants relative to the ProM task were the same. When formed correct intention, participants were able to successfully complete the task in the appropriate amount of time. The current research denied the conclusion and proved the benefit of auditory encoding.

The findings in the two experiments produced a consistent pattern, showing the advantages of specific cues encoding in the standard encoding condition. We purposely used a vowel ("U") as a ProM target, whether it was used in a specific encoding condition or non-specific encoding condition. This decision was based on situations one might face in everyday life, such as going to the supermarket to buy food for lunch and wondering which purchase would be better. For example, an individual is more likely to purchase fruit when given the name of a specific fruit than when only told to purchase fruit. Our results are consistent with Lourenço et al. (2013), who demonstrated that target specification can reduce costs in non-focal ProM and trial-by-trial changes in task interference, as the result of top-down attention control processes. Previous research suggested that non-specific cue encoding may activate more connections than specific cue encoding; therefore, participants need more cognitive to perform intended tasks and resulting in slower RTs (Hicks et al., 2005; Loft et al., 2008). If target cues are specific and unambiguous, then the connection between target and behavior becomes more specific, thereby facilitating ProM retrieval. The current findings were consistent with Pan (2010), who found that when ProM encoding was specific, participants adopted automatic processing, the attention required relatively few resources, and ProM performance was high. However, when ProM encoding was non-specific, the monitoring processing method was used more often, resulting in a relatively higher resource consumption, affecting ProM performance, and significantly impacting RT and accuracy in the ongoing task.

Compared with non-specific encoding, specific encoding requires fewer cognitive resources. This finding helps to explain why it can be more difficult to perform an intended action without a specific task description. For example, it may be difficult to choose flowers when only told to get "fresh flowers" and not told to select a specific type of flower. There were also significant effects of accuracy and RT for the ongoing task. Participants who were assigned to the specific ProM target condition had more cognitive resources with which to perform the ongoing task than those in the non-specific ProM target condition.

Implementation intention encoding can not only make up for the disadvantages of visual encoding but also improve ProM performance for non-specific encoding conditions. The present study confirmed that implementation intention enhanced goal attainment by facilitating the initiation of planned responses upon encountering critical situations (Gollwitzer and Brandstätter, 1997). The mental representation of the ProM tasks specified in the if-part becomes a highly activated and easily accessible cue. By using implementation intention encoding, individuals establish a vivid psychological image between ProM intention and behavior, which has a superposition effect with words encoded by visual, resulting in that visual implementation intention encoding can significantly improve ProM performance. As a consequence, the ProM task receives attentional and perceptual priority (Achtziger et al., 2011; Janczyk et al., 2015) and is easily detected in the environment. A strong link is forged between ProM cue and ProM response specified in then-part. This renders ProM response automatically (Gan et al., 2017, 2020), and no other cognitive resources are needed to perform theProM task correctly. This study shows that the advantages of implementation intention encoding are not affected by audiovisual encoding modality; moreover, implementation intention encoding can compensate for the shortcomings of visual encoding and promote the execution of ProM tasks, without the cost of ongoing tasks.

Implementation intention creates a strong mental representation of the situation and a strong linkage between situation and response that makes it easy to execute the behaviors. It helps people to overcome the gaps between the intentions and the actual behaviors, and improves ProM performance (Chen et al., 2015). The empirical data of the present study confirmed that the effect of implementation intention can be generalized to non-specific situations (Bieleke et al., 2018; Huang et al., 2020). Through the connection of implementation intention, the ProM performance under non-specific conditions has been significantly improved, reaching a level equivalent to that under specific conditions. Combined with the fact that there was no significant difference between the implementation intention and standard encoding condition, we inferred that implementation intention can specify intention, which is similar to specify encoding.

The Limitations of the Present Study

One potential limitation of the present study is that the one-back task was our ongoing task, and almost all participants performed it perfectly, with accuracy rates above 0.90, which demonstrated high ceiling effects. Maybe the one-back task was too easy for undergraduates. Thus, future studies should use more difficult tasks and manipulate task load. Also, we manipulated encoding and neglected the consistency of encoding and extraction modality. It is also worth further research to determine whether there is an audio-visual modality effect on ProM performance, by comparing the consistency of encoding and extraction modality of visual and auditory encoding.

CONCLUSION

The results of our study suggest there are specific advantages in the field of ProM, and one such advantage is that specific encoding can contribute to ProM performance relative to non-specific encoding. Also, implementation intention is an effective method for enhancing ProM performance in a visual encoding condition and in a non-specific encoding condition. Regardless of implementation intention or standard encoding, individuals could successfully complete tasks as long as they were encoded by specific intention.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Fujian Normal University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONSS

MZ performed the analysis and was responsible for the design and planning of the study. YC, MZ, CX, YG, QL, ZM, JH, WH, and QFL performed the study. MZ and YC wrote the final report. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Social Science Foundation of China, “the influencing factors, mechanism, and promotion strategy of PRoM in low-achieving students” (BBA180082).

ACKNOWLEDGMENTS

We would like to thank Editage (www.editage.cn) for English language editing.

REFERENCES

Achtziger, A., Bayer, U. C., and Gollwitzer, P. M. (2011). Committing to implementation intentions: Attention and memory effects for selected situational cues. *Motivat. Emot.* 36, 287–300.

Ball, B. H., and Bugg, J. M. (2018). Aging and the strategic use of context to control prospective memory monitoring. *Psychol. Aging* 33:527. doi: 10.1037/pag0000247

Bieleke, M., Legrand, E., Mignon, A., and Gollwitzer, P. M. (2018). More than planned: Implementation intention effects in non-planned situations. *Acta Psychol.* 184, 64–74. doi: 10.1016/j.actpsy.2017.06.003

Brewer, G. A., Knight, J., Meeks, J. T., and Marsh, R. L. (2011). On the role of imagery in event-based prospective memory. *Consciousness Cognit.* 20, 901–907. doi: 10.1016/j.concog.2011.02.015

Cai, R. D., and Dong, Y. P. (2012). The influence of information types, coding channels and coding languages on the breadth of working memory: evidence supporting the hierarchical view. *Foreign Lang. Teaching Res.* 044, 376–388.

Chasteen, A. L., Park, D. C., and Schwarz, N. (2010). Implementation intentions and facilitation of prospective memory. *Psychol. Sci.* 12, 457–461.

Chen, G. D. (2013). *A research about how representation style and external cues affect the high-specific ratio of Involuntary Autobiographical Memory*. Ph. D. thesis. Jilin: Northeast Normal University.

Chen, X. J., Wang, Y., Liu, L. L., Cui, J. F., Gan, M. Y., Shum, D. H., et al. (2015). The effect of implementation intention on prospective memory: A systematic and meta-analytic review. *Psychiatry Res.* 226, 14–22. doi: 10.1016/j.psychres.2015.01.011

Chen, Y. Z., Meng, Y. F., Yuan, H., and Yang, L. X. (2014). Neural correlates of the interference effect of a type of mixed prospective memory. *J. Psychol. Sci.* 37, 1073–1078.

Einstein, G. O., Mcdaniel, M. A., Richardson, S. L., Guynn, M. J., and Cunfer, A. R. (1995). Aging and prospective memory: examining the influences of self-initiated retrieval processes. *J. Exp. Psychol. Learn. Mem. Cognit.* 21:996. doi: 10.1037//0278-7393.21.4.996

Gan, J., Guo, Y., Guo, Y., Zhang, Z., and Chen, Y. (2017). The Effect of Implementation Intentions on Prospective Memory Under Different Cognitive Load. *Stud. Psychol. Behav.* 15, 743–749. doi: 10.1016/j.jecp.2014.01.011

Gan, J., Zhang, M. M., Hu, J. H., Guo, Y. F., and Chen, Y. Z. (2020). The underlying mechanism of implementation intention on event-based prospective memory. *Stud. Psychol. Behav.* 18, 53–59.

Gollwitzer, and Peter, M. (1999). Implementation intentions: strong effects of simple plans. *Am. Psychol.* 54, 493–503. doi: 10.1037/0003-066x.54.7.493

Gollwitzer, P. M., and Brandstätter, V. (1997). Implementation intentions and effective goal pursuit. *J. Personal. Soc. Psychol.* 73, 186–199. doi: 10.1037/0022-3514.73.1.186

Guo, Y. F., Gan, J. Q., Zhang, Z., Huang, T. H., and Chen, Y. Z. (2016). The effects of encoding types, cognitive loading, and number of cues on prospective memory. *J. Psychol.* 39, 1058–1063.

Hicks, J. L., Marsh, R. L., and Cook, G. I. (2005). Task interference in time-based, event-based, and dual intention prospective memory conditions. *J. Mem. Lang.* 53, 430–444. doi: 10.1016/j.jml.2005.04.001

Huang, X., Chen, S., Gao, W., Yang, J., and Yuan, J. (2020). Emotion regulation by implementation intention is generalizable to unspecified situations: The nature of the underlying goal matters. *Acta Psychol.* 210:103144. doi: 10.1016/j.actpsy.2020.103144

Janczyk, M., Dambacher, M., Bieleke, M., and Gollwitzer, P. M. (2015). The benefit of no choice: goal-directed plans enhance perceptual processing. *Psychol. Res.* 79, 206–220. doi: 10.1007/s00426-014-0549-5

Kliegel, M., Mahy, C., Voigt, B., Henry, J. D., and Aberle, I. (2013). The development of prospective memory in young schoolchildren: The impact of ongoing task absorption, cue salience, and cue centrality. *J. Exp. Child Psychol.* 116, 792–810. doi: 10.1016/j.jecp.2013.07.012

Kuhlmann, B. G., and Rummel, J. (2014). Context-specific prospective-memory processing: evidence for flexible attention allocation adjustments after intention encoding. *Memory Cognit.* 42, 943–949. doi: 10.3758/s13421-014-0405-2

Li, Q., Zhang, X. W., and Huang, X. T. (2016). The Audio-visual Modality Effect in Ego Depletion. *J. Psychol. Sci.* 39, 514–519.

Lin, W. (2016). *The impact of implementation intentions on prospective memory under different cognitive load*. Ph. D. thesis. Wulumuqi: XinJiang Normal University.

Loft, S., Kearney, R., and Remington, R. (2008). Is task interference in event-based prospective memory dependent on cue presentation? *Psychol. Med.* 36, 139–148. doi: 10.3758/mc.36.1.139

Lourenço, J. S., White, K., and Maylor, E. A. (2013). Target context specification can reduce costs in nonfocal prospective memory. *J. Exp. Psychol. Learn. Mem. Cognit.* 39, 1757–1764. doi: 10.1037/a0033702

Lv, Y. B. (2010). *An experiment of implementation intentions within prospective memory*. Ph. D. thesis. Jilin: Jilin University.

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., et al. (2001). Distinct pathways involved in sound recognition and localization: a human fmri study. *Neuroimage* 14, 802–816. doi: 10.1006/nimg.2001.0888

Mcdaniel, M. A., Howard, D. C., and Butler, K. M. (2008). Implementation intentions facilitate prospective memory under high attention demands. *Mem. Cognit.* 36, 716–724. doi: 10.3758/mc.36.4.716

McFarland, C., and Glisky, E. (2012). Implementation intentions and imagery: individual and combined effects on prospective memory among young adults. *Mem. Cognit.* 40, 62–69. doi: 10.3758/s13421-011-0126-8

Meeks, J. T., and Marsh, R. L. (2010). Implementation intentions about nonfocal event-based prospective memory tasks. *Psychol. Res.* 74:82. doi: 10.1007/s00426-008-0223-x

Pan, L. (2010). *The development of prospective memory based on task-characteristics*. Ph. D. thesis. Tianjin: Tianjin Normal Univercity.

Pereira, A., Altgassen, M., Atchison, L., de Mendonça, A., and Ellis, J. (2018). Sustaining Prospective Memory Functioning in Amnestic Mild Cognitive Impairment: A Lifespan Approach to the Critical Role of Encoding. *Neuropsychology* 32, 634–644. doi: 10.1037/neu0000441

Rummel, J., Einstein, G. O., and Rampey, H. (2012). Implementation-intention encoding in a prospective memory task enhances spontaneous retrieval of intentions. *Memory* 20:803. doi: 10.1080/09658211.2012.707214

Sanjram, P. K., and Khan, A. (2011). Attention, polychronicity, and expertise in prospective memory performance: Programmers' vulnerability to habit intrusion error in multitasking. *Int. J. Hum. Comput. Stud.* 69, 428–439. doi: 10.1016/j.ijhcs.2011.01.006

Scott, I. A. (2016). *The role of context and memory strategies in prospective memory*. Ph. D. thesis. Chattanooga: Tennessee University.

Scullin, M. K., Kurinec, C. A., and Nguyen, K. (2017). The effects of implementation intention strategies on prospective memory cue encoding. *J. Cognit. Psychol.* 29, 929–938. doi: 10.1080/20445911.2017.1329205

Scullin, M. K., McDaniel, M. A., Dasse, M. N., Lee, J. H., Kurinec, C. A., Tami, C., et al. (2018). Thought probes during prospective memory encoding: Evidence for perfunctory processes. *PLoS One* 13:e0198646. doi: 10.1371/journal.pone.0198646

Vedhara, K., Wadsworth, E., Norman, P., Searle, A., Mitchell, J., Macrae, N., et al. (2004). Habitual prospective memory in elderly patients with type 2 diabetes: implications for medication adherence. *Psychol. Health Med.* 9, 17–27. doi: 10.1080/13548500310001637724

Yang, J., Guo, X. Y., and Ma, F. (2008). A research on features of activity-based prospective memory. *J. Psychol. Sci.* 31, 1348–1352. doi: 10.16719/j.cnki.1671-6981.2008.06.023

Zimmermann, T. D., and Meier, B. (2010). The effect of implementation intentions on prospective memory performance across the lifespan. *Appl. Cognit. Psychol.* 24, 645–658.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

METHODS
published: 13 April 2022
doi: 10.3389/fpsyg.2022.838878

Effects of Leisure Activities on the Cognitive Ability of Older Adults: A Latent Variable Growth Model Analysis

Chang-e Zhu, Lulin Zhou and Xinjie Zhang*

Department of Management, Jiangsu University, Zhenjiang City, China

Based on the data of four periods of CLHLS (2008, 2011, 2014, 2018), the latent variable growth model (LGCM) was applied to 2344 older adults who completed four follow-up surveys, to study the trajectory of leisure activities and cognitive ability and explore the relationship between leisure activities and cognitive ability of older adults. The results showed that: (1) leisure activities and cognitive ability of older adults showed a non-linear downward trend; (2) leisure activities significantly and positively predicted the cognitive ability of older adults at every time point; (3) the initial level of leisure activity positively predicted the initial level of cognitive ability but negatively predicted the rate of cognitive decline; In addition, cognitive activities had a greater effect on cognitive ability than non-exercise physical activities; (4) the rate of decline of leisure activities also significantly and positively predicted the rate of decline of cognitive ability; (5) cross-lagged regression analysis further suggested the overall positive predictive effect of leisure activity on cognitive ability; (6) overall, education level had a significant contribution to cognitive ability, and the higher the education level, the slower the decline of cognitive ability; and (7) smoking could promote cognitive ability in older adults and no significant effect was found between alcohol drinking and cognitive ability. Accordingly, the government should encourage older adults to do more leisure activities, especially the cognitive activity, to effectively prevent cognitive decline.

Keywords: cognitive ability, leisure activities, development trajectory, latent variable growth model (LGCM), cross-lagged regression analysis

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Jeanine Parisi,
Johns Hopkins University,
United States
Ludmila Zajac-Lamparska,
Kazimierz Wielki University, Poland

****Correspondence:***
Lulin Zhou
dorischange@126.com

Specialty section:
This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: *18 December 2021*
Accepted: *28 February 2022*
Published: *13 April 2022*

Citation:
Zhu C, Zhou L and Zhang X (2022) Effects of Leisure Activities on the Cognitive Ability of Older Adults: A Latent Variable Growth Model Analysis. Front. Psychol. 13:838878. doi: 10.3389/fpsyg.2022.838878

BACKGROUND

With the deepening process of population aging, how to promote healthy aging and maintain the physical and mental health of older adults had become the focus of attention from all walks of life. The results of China's seventh population census revealed that in 2021, China's older population had reached 264.02 million, accounting for 18.7% of the total population (National Bureau of Statistics of China, 2021). According to the United Nations estimate, by 2050, China's older population would reach 488 million, accounting for 35.6% of the total population (United Nations, 2012), which meant that China had stepped into a deep aging society, and suffered the problem brought from aging.

Although life expectancy continued to increase, it was necessary to consider adverse outcomes in advance given the increasing likelihood that cognitive ability will decline with age. A previous study showed that deterioration of cognitive ability could lead to physical dysfunction and certain chronic diseases, and severe deterioration of cognitive ability was highly associated with mortality (Yue et al., 2021). Furthermore, the decline of subjective (slight) cognitive ability could be regarded as a precursor of dementia, which seriously impaired the health-related quality of life of older adults (Pusswald et al., 2015).

It was estimated that the number of dementia patients aged 60 and above in China was 10 million to 11 million, the number of dementia patients aged 65 and above was 9 to 10 million, and more than 60% of dementia patients had Alzheimer's disease (Jia et al., 2020). Accordingly, studying the change trajectory of cognitive ability and exploring the determinants of cognitive ability in older adults, and formulating appropriate intervention policies should be one of the most important tasks for China to address the challenges of population aging. Indeed, previous studies suggested that individuals with more years of education showed less cognitive decline (Liu and Lachman, 2020). But the years of education that older adults received at a young age could not be changed, so it made more sense to study variables that could be modifiable. Studies indicated that leisure activities, as one of the main lifestyles of the elderly in their later years, had an important impact on their cognitive ability. However, the current research conclusions were inconsistent. To be specific, intellectual leisure activities such as reading books and newspapers had a positive impact on cognitive ability (Silverstein and Parker, 2002). However, watching TV was examined to exert an adverse effect on cognitive ability (Rundek and Bennett, 2006; Hamer and Stamatakis, 2014). Accordingly, it was necessary to study the relationship between leisure activities and classify leisure activities, and explain the differences in the impact of different types of leisure activities on cognitive ability.

Development of Cognitive Ability of Older Adults

Cognitive ability referred to a series of intelligent processing processes in which the human brain received external information and acquires knowledge, which included perception, attention, memory, thinking and other abilities (Ruining and Yinge, 2015). Maintaining the cognitive ability of the older adults was of great significance for the old population in their later life. Firstly, degradation of cognitive ability was strongly associated with morbidity and mortality from a variety of diseases. Secondly, consequences of cognitive decline in older adults not only damaged physical health, but also significantly increased the probability of Alzheimer's disease (Prince et al., 2013). Last but not least, deterioration of cognitive ability could lead to physical dysfunction and certain chronic diseases, and severe deterioration of cognitive ability was highly associated with mortality (Yue et al., 2021). Therefore, it was important and urgent to study the changing trajectory of cognitive ability of older adults. The theory of human cognitive aging proposed that individual cognition ability would change with age. To be specific, with the increase of age, the neural substrate of people will change, and cognitive resources such as processing speed, working memory, attention span, and inhibitory ability may decline (Wingfield and Grossman, 2006). In other words, the cognitive ability of older adults would decline overtime.

Based on the existing study, the following assumption was proposed:

> **H1:** With the increase of age, the cognitive ability of older adults showed a downward trend;

Development of Leisure Activities for Older Adults

Leisure activity was a kind of meaningful activity which was not related to earning a living or getting a reward. It is an inclusive intellectual activity, multi-dimensional concepts of social, recreational and sports activities (Wuzhen, 2008). The older adults retired from work, so leisure activities became an important arrangement for the older people in their later years. The young elderly (62–69) gradually steps into the ranks of the middle-aged (70–76) and even older adults (77–94) over time (Williams et al., 2012). During this period, the physical function and psychological course of older adults will change, which will further lead to changes in the way of leisure activities. For example, when they were in the low level of aging, outdoor physical exercise and other leisure activities that required much physical exertion were more common than those for middle and older aged people, while older aged people mainly read newspapers and watched TV indoors due to the limitation of their physical conditions and the change of their mental state (Robinson et al., 2004). In addition, leisure activities related to social and intellectual activities also tended to decline due to the lack of energy of older adults. In general, older adults experienced a decline in physical and mental functioning over time, resulting in lower levels of leisure activity. In other words, the level of leisure activities of older people may decrease over time.

Based on the existing literature, the following assumption was proposed:

> **H2:** With the increase of age, the leisure activities of older adults showed a downward trend;

Relationship Between Leisure Activities and Cognitive Ability

At present, there were a large number of studies to analyze the relationship between leisure activities and cognitive ability and leisure activities were considered to be a protective factor for cognitive abilities in older adult (Mao et al., 2020), which would reduce the risk of cognitive decline or dementia (Bennett et al., 2014). For example, a prospective cohort study of community-dwelling adults over 70 years of age found that leisure hobbies significantly reduced the risk of cognitive decline in older adults (Iwasa et al., 2012). In addition, regular outdoor exercise was considered to be an important protection strategy to effectively prevent the decline of cognitive ability (Ding et al., 2021). The theory of social communication held that social activities are an important way for individual human beings to survive, live and practice, a prerequisite for material production and a direct link with individual mental health (Weixiong, 2004). Social interaction can not only stimulate the individual's thinking, promote older people's emotional communication, but also improve and promote the older population's understanding, judgment, memory and expression and other cognitive functions. Previous studies revealed that social activities such as keeping close contact with friends could improve cognitive ability to a certain extent, which was a key factor in effectively preventing cognitive decline (Hui-Xin et al., 2013; Noice et al., 2014). In addition, participation in recreational activities was suggested to be a protective factor against cognitive decline in older adults (Hui-Xin et al., 2002). Finally, numerous studies suggested watching TV was associated with cognitive decline (Hamer and Stamatakis, 2014; Fancourt and Steptoe, 2019). However, watching TV, as a very common form of leisure in old age, was an important source of information for older people and contributed to their overall cognitive function (Ostlund, 2010). Accordingly, leisure activities could be considered to be positively associated with cognitive ability on the whole.

Based on the above analysis, the following assumptions were put forward:

> **H3:** Leisure activities of older adults positively predicted their cognitive ability during the same period;

> **H4**: The initial level of leisure activities of older population positively predicted the initial level of cognitive ability;

In addition, leisure activities was confirmed to be a protective factor for the cognitive ability of older adults. Leisure activities may reduce the rate of cognitive decline, which indicated that the higher the level of leisure activities, the slower the decline of their cognitive ability. Therefore, the following assumptions were put forward:

> **H5:** The initial level of leisure activities of older adults negatively predicted the change speed of their cognitive ability;

The level of leisure activities and cognitive ability of older adults showed signs of decline to some extent over time. According to the life cycle theory (Haizhong, 2014), from the early stage to the middle age stage, the physical and psychological functions of older adults often showed a slow decline. From the middle age to the advanced age stage, the older people reached the last stage of the life cycle, both physical function and cognitive function declined precipitously. In other words, the change speed of physical function and psychological cognition of older population was consistent. Therefore, the following assumptions were proposed:

> **H6:** The change rate of leisure activities of older adults positively predicted the change rate of their cognitive ability;

In addition, many studies suggested the positive relationship between leisure activities and cognition ability, which, however, were mainly based on the cross-sectional data or short follow-up period. Cognitive decline is a continuous and insidious process, so it is necessary to prevent and explore its protective factors in advance (Kåreholt et al., 2011). Additionally, cognitive reserve theory suggested that intellectual activities in midlife will created a stock for cognitive capacity, which will influenced their cognitive ability in later life by improving the ability to compensate for age-related neuronal damage (Costa et al., 2007). In other words, baseline leisure time activities were examined to be a protective factor for late-life cognition (Kåreholt et al., 2011). Most importantly, previous studies have proved the positive relationship between leisure activities on cognitive ability, but endogenous problem could not be ruled out (Andel et al., 2015; Lee et al., 2019). Therefore, cross-lagged regression analysis was used to further determine the temporal order of the relationship between leisure activity level and cognitive ability. To verify that there was a causal relationship between two variables, the key step was to ensure that there was a temporal difference between the two variables. In other words, one variable must occur before the other. Hence, it was necessary to study the influence of current leisure activities on cognitive ability in the later period. Therefore, the assumption was put forward:

> **H7:** The leisure activities of older adults in the current period positively predicted cognitive ability in the later period.

EDUCATION DIFFERENCES IN COGNITIVE ABILITY

The concept of cognitive reserve theory suggested that innate intelligence or life experience (such as educational or professional achievement) may provide reserves in the form of a set of skills to better cope with the decline of cognitive ability (Stern, 2002). Education contributed to the development of brain structure and neural networks, and to the further development of concept formation, vocabulary expression and cognitive functions such as perception and memory (Konttinen et al., 2016). Education was considered a protective factor for cognitive ability, and studies indicated that older people with higher levels of education had higher cognitive abilities (Hazzouri et al., 2011; Ruining and Yinge, 2015). And older adults with higher education had a slower rate of cognitive decline (Sunmin et al., 2003). Therefore, the following assumption was proposed:

> **H8:** The educational level of older adults positively predicted their cognitive ability;

To sum up, it was suggested that studies on cognitive abilities of older people were mainly based on cross-sectional data (Saji et al., 2020; Kashibayashi et al., 2021). Furthermore, the existing research on the effect of leisure activities on cognitive ability was mainly based on one leisure activity and did not distinguish the differences in the impact of different types of leisure activities on cognitive ability. In addition, the existing research did not explain the mechanism of the influence of leisure activities on cognitive ability, but simply verified the positive relationship between the levels of two variables, and lacked discussion on the relationship between the initial level of two variables and change rate. Accordingly, longitudinal multi-time data was used to further enrich the connotation of leisure activities and track the change trajectory of cognitive ability and leisure activities of older adults over time. Additionally, leisure activities were categorized into two parts, which included two major aspects: cognitive activities and non-exercise physical activities to study the effects of different types of leisure activities on cognitive ability. Last but not least, this paper studied the influence mechanism of leisure activity on their cognitive ability and discussed the relationship between the initial level of two variables and change rate.

METHOD

Source of Data

The data were obtained from the Chinese Longitudinal Healthy Longevity Survey (CLHLS), which were conducted by the National Development Institute and the Center for Research on Healthy Aging and development of Peking University. The present study used data from the last four waves of datasets, including the 2008, 2011, 2014, and 2018 waves and the participations in this study were all aged over 60 at the baseline. This paper mainly studied the changing trend of cognitive ability of older adults and the effect of leisure activities on cognitive ability. In order to effectively track the change trajectory of cognitive ability of the older adults, the individuals who simultaneously participated in the survey in 2008, 2011, 2014, and 2018 were included in this study. The sample data acquisition process was shown in **Figure 1**. The individuals with missing values on any variable were excluded and then a total of 2344 valid individuals were obtained. In addition, for the attrition, the missing completely at random (MCAR) test was conducted to clarify the trend of the missing data and the result was not significant ($\chi 2/\mathrm{df} = 6.8$, $P > 0.05$), indicating the missing participants were at random.

Variable Definition and Measurement

The cognitive ability was assessed by Chinese version of the Mini-Mental State Examination (MMSE), which was suggested to be validated in Chinese older population (Zeng et al., 2017; Lv et al., 2018). To be specific, the cognitive ability was measured by general ability (3 items), responsiveness (3 items), attention and calculation ability (6 items), recall (3 items), and language comprehension and self-coordination (6 items), which totally consists of 24 items. Responses were given on dichotomous scale (1 = correct, 0 = wrong) with the exception of the item"number of kinds of food named in 1 min," which allows a maximum score of 7, so the range for cognitive ability was 0–30. An example item is "What time of day is it right now (morning, afternoon, evening)? The coefficients alpha at every point was 0.863, 0.855, 0.886, and 0.928.

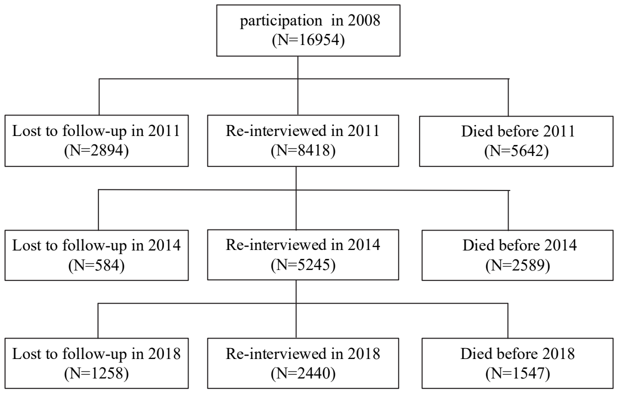


FIGURE 1 | Source of sample structure.

Leisure activities were measured by eight questions, namely (1) housework (cooking and taking care of children), (2) outdoor activities (such as Taijiquan, square dance, crossing the door and communicating with friends), (3) planting flowers and keeping pets, (4) reading books and newspapers, (5) raising poultry and livestock, (6) playing cards or mahjong, (7) listening to the radio, and (8) participating in organized social activities. Responses were given on a five-point Likert scale (1 = almost every day to 5 = rarely or never). An example item is "Do you now perform the following activities regularly (Outdoor activities)?" In this paper, the negative score of the original scale was transformed, so the higher the score was, the higher the level of leisure activities was and the final score range of leisure activities for older adults was 8–40. The coefficients alpha at every point was 0.768, 0.807, 0.856, and 0.908.

A total of 23,44 (1,097 males and 1,247 females) older adults were included in the study (*female* = 0, *male* = 1). Education level was measured by the number of years of education older adults had received (0–22). In addition, lifestyle included whether they had smoked/drunk alcohol was tested (*yes* = 1, *no* = 0). Specifically, 840 (35.8%) people smoked in 2008,837 (35.7%) in 2011, 761 (32.4%) in 2014 and 758 (32.3%) in 2018, 815 (34.8%) drank alcohol in 2008, 821 (35.0%) in 2011, 679 (29.0%) in 2014, and 637 (27.2%) in 2018. The definition of all variables was shown in **Table 1**.

Data Analysis Strategy

In this paper, Mplus8.0 was used to construct a latent variable growth model to test the trend of cognitive ability of older

TABLE 1 | Descriptive analysis of samples.

	Definition	Mean	SD
CA1	The total score of cognitive ability (0–30)	26.77	4.53
LA 1	The total score of leisure activity (8–40)	20.95	5.43
SM1	Smoked = 1; Never smoked = 0	0.36	0.48
AD1	Drunk = 1; never drunk = 0	0.35	0.48
AGE1	The true age of older adults (61–108)	75.16	8.29
EDU	The years of schooling (0–20)	2.84	3.7
SEX	Male = 1; female = 0	0.47	0.5

CA denotes cognitive ability; CA1 = CA (2008); LA denotes leisure activities; LA1 = LA (2008); SM denotes smoking; SM1 = SM (2008); AD denotes alcohol drinking; AD 1 = AD (2008); EDU denotes the years of schooling; SD denotes standard deviation.

TABLE 2 | Correlation coefficient matrix.

	CA1	***CA 2***	***CA 3***	***CA 4***	***LA 1***	***LA 2***	***LA 3***	***LA 4***	***SM1***	***SM2***	***SM3***	***SM4***	***AD1***	***AD2***	***AD3***	***AD4***
CA 1	1															
CA 2	0.310***	1														
CA 3	0.272***	0.412***	1													
CA 4	0.311***	0.383***	0.452***	1												
LA1	0.338***	0.225***	0.191***	0.201***	1											
LA 2	0.223***	0.308***	0.233***	0.277***	0.310***	1										
LA 3	0.246***	0.260***	0.315***	0.354***	0.323***	0.440***	1									
LA 4	0.236***	0.269***	0.279***	0.492***	0.294***	0.393***	0.459***	1								
SM1	0.101***	0.116***	0.127***	0.144***	0.102***	0.144***	0.130***	0.138***	1							
SM2	0.097***	0.118***	0.111***	0.136***	0.105***	0.155***	0.135***	0.137***	0.910***	1						
SM3	0.102***	0.108***	0.128***	0.131***	0.074***	0.163***	0.142***	0.130***	0.694***	0.714***	1					
SM4	0.103***	0.110***	0.130***	0.127***	0.075***	0.160***	0.151***	0.139***	0.693***	0.695***	0.687***	1				
AD1	0.075***	0.070***	0.077***	0.081***	0.098***	0.092***	0.094***	0.087***	0.458***	0.441***	0.370***	0.365***	1			
AD2	0.080***	0.071***	0.078***	0.069***	0.086***	0.107***	0.100***	0.080***	0.425***	0.448***	0.379***	0.360***	0.831***	1		
AD3	0.087***	0.087***	0.106***	0.116***	0.067***	0.129***	0.153***	0.118***	0.376***	0.394***	0.477***	0.373***	0.569***	0.606***	1	
AD4	0.099***	0.091***	0.116***	0.124***	0.072***	0.127***	0.137***	0.133***	0.369***	0.375***	0.377***	0.461***	0.543***	0.559***	0.561***	1

**p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001.*

people. Latent variable growth model (LGCM) was a variation of structural equation model, which could describe the variation type between repeated measurements by the potential trajectory of the hypothesis. Unlike traditional statistical methods (such as ANOVA with repeated measurements) that focused only on group mean, LGCM could estimate both group and individual variation during development (McArdle, 2008). The LGCM first defined two latent variable structures, i.e., the starting level and the slope. These two latent variable structures were then estimated in the model using the actual measurements of a variable at different time points.

RESULTS

Common Method Bias

In this study, the single-factor Harman test was carried out on the data of the four surveys. The results showed that the variation explained by the first factor was less than 40% in the four measurement periods, which were 24.617, 22.647, 26.776, and 34.778%, respectively. This indicated that there was no common method bias in this study (Podsakoff et al., 2003).

Descriptive Statistical Results

Descriptive statistics and bivariate correlations among all observed variables were shown in **Tables 1**, **2**, respectively. It can be concluded that the average scores for cognitive ability and leisure activity were within 23–27 and 16–22, respectively. Overall, they declined over time. The scores of leisure activities and cognitive abilities of older adults during the four measurement periods were intuitively shown in **Figure 2**. Additionally, it was suggested that cognitive ability was positively correlated with leisure activities from 2008 to 2018 (T1–T4).

Development Trajectory of Cognitive Ability of Older Adults (Model 1/2)

In order to examine the trend of cognitive ability of older people, a linear growth model and quadratic growth model were constructed, which were shown in **Figures 3**, **4** and denoted as Model 1 (M1) and Model 2 (M2), respectively.

For cognitive ability, both the linear growth model and the quadratic growth model fit the data very well (for the linear growth model 1, χ^2(df) = 20.7 (5), χ^2/df = 4.14, CFI = 0.864, RMSEA = 0.131, and SRMR = 0.084, P = 0.000; for the quadratic growth model 2, χ^2(df) = 3.381(1), χ^2/df = 3.381, CFI = 0.994, RMSEA = 0.063, and SRMR = 0.013, P = 0.000). The linear growth model was nested under the quadratic growth model (Chan, 1998), so we compared leisure activities' linear growth

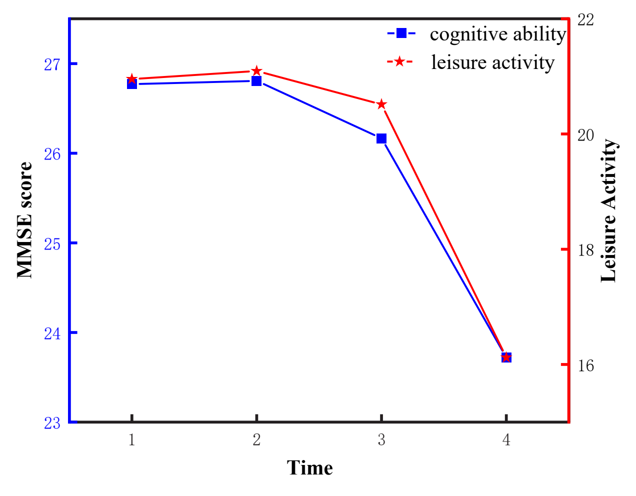


FIGURE 2 | Change trajectory of cognitive ability and leisure activities.

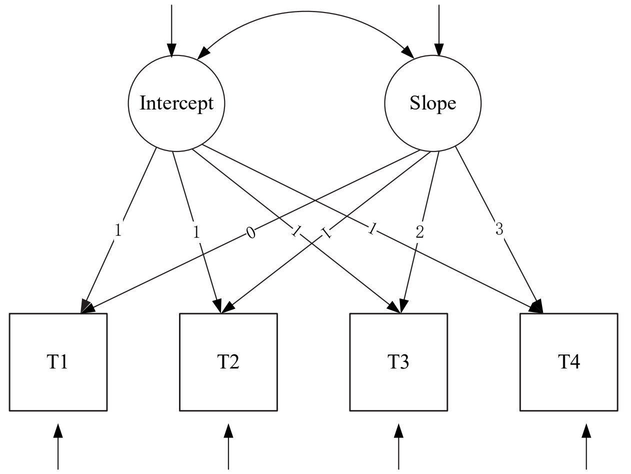


FIGURE 3 | Linear unconditional latent variable growth model.

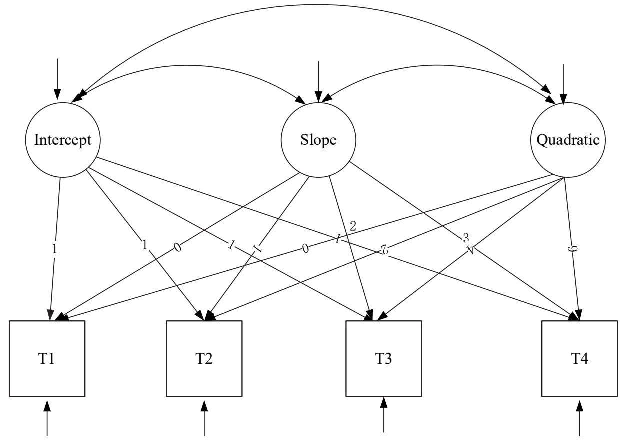


FIGURE 4 | Non-linear unconditional latent variable growth model.

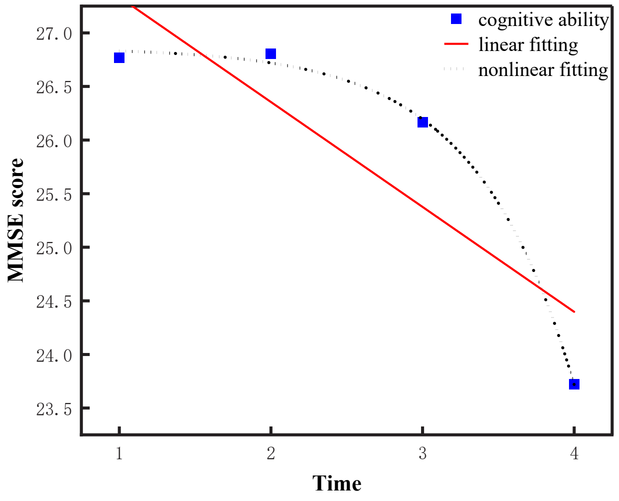


FIGURE 5 | Cognitive ability fitting.

TABLE 3 | Model comparisons for leisure activity and cognitive ability.

Variables	Model	χ^2(df)	*P*	CFI	RMSEA	SRMR
Cognitive ability	Model 1	20.7(5)	0.000	0.864	0.131	0.084
	Model 2	3.381(1)	0.000	0.994	0.063	0.013,
Leisure activities	Model 3	66.3(5)	0.000	0.584	0.237	0.107
	Model 4	5.69(1)	0.000	0.965	0.054	0.032

model and quadratic growth model by using the chi-square test. we compared cognitive ability's linear growth model and quadratic growth model by using the chi-square test. The result was significant ($\Delta\chi^2 = 6.24$, $\Delta df = 2$, $P < 0.05$). Furthermore, the fitting effect of quadratic growth model was better than that of linear growth model, which was shown in **Figure 5**, so we used the quadratic growth model of cognitive ability. To be specific, the initial level of cognitive ability status (the intercept) was 26.729 ($P < 0.001$). Cognitive ability decreased during the four tests (slope = 0.819, $P < 0.001$), and the rate of decline increased year by year (curve slope = −0.593, $P < 0.01$), suggesting a non-linear downward trend in cognitive ability over the four test periods. In addition, the variation of intercept ($\sigma^2 = 10.108$, $P < 0.001$) and slope ($\sigma^2 = 0.626$, $P < 0.05$) were both significantly different among the older population, which indicated that there was significant difference in the initial level and the rate of change of cognitive ability among older adults. Therefore, Hypothesis 1 was supported. **Table 3** summarizes the model fit indices of latent growth models for leisure activities and cognitive ability. **Table 4** shows parameter estimates of latent growth models for leisure activities and cognitive ability.

Development Trajectory of Leisure Activities of Older Adults (Model 3/4)

Similarly, in order to examine the trend of leisure activities of the older population, a linear growth model and quadratic growth model were constructed, which were shown in **Figures 3**, **4** and denoted as Model 3 (M3) and Model 4 (M4), respectively.

For leisure activities, both the linear growth model and the quadratic growth model fit the data very well (for the linear growth model 3, χ^2 (df) = 66.3 (5), $\chi^2/df = 13.26$, CFI = 0.584, RMSEA = 0.237, and SRMR = 0.107, $P = 0.000$; for the quadratic growth model 4, $\chi^2/(df) = 5.69(1)$, $\chi^2/df = 5.69$, CFI = 0.965, RMSEA = 0.054, and SRMR = 0.032, $P = 0.000$). we compared leisure activities' linear growth model and quadratic growth model by using the chi-square test. The result was significant ($\Delta\chi^2 = 8.23$, $\Delta df = 3$, $P < 0.05$).

In addition, the fitting effect of quadratic growth model was better than that of linear growth model, which could be obviously seen from **Figure 6**, so we used the quadratic growth model of leisure activities. To be specific, the initial level of leisure activity status (the intercept) was 20.773 ($P < 0.001$). Leisure activities decreased during the four tests (slope = 1.936, $P < 0.001$), and the rate of decline increased year by year (curve slope = −1.149, $P < 0.001$), suggesting a non-linear downward trend in leisure

TABLE 4 | Parameter Estimates of latent growth models for leisure activities and cognitive ability.

Variables	Model	Means of growth factors			Variance of growth factors		
		Intercept	Slope	Quadratic	Intercept	Slope	Quadratic
Cognitive ability	Model 1	27.254***	-0.786***		3.799***	0.816***	
	Model 2	26.729***	0.819***	-0.593***	10.108***	8.496***	0.626*
Leisure activities	Model 3	21.976***	-1.587***		8.578***	0.116	
	Model 4	20.773***	1.936***	-1.149**	6.126*	1.836	0.277

activity over the four test periods. In addition, the variation of intercept was significant ($\sigma^2 = 6.126$, $P < 0.05$) and the variation of curve slope wasn't significant ($\sigma^2 = 0.277$, $P > 0.05$), which indicated that there were differences in the initial level of leisure activity and no differences in the decline rate of leisure activity level among the older population. Therefore, hypothesis 2 was supported.

Effects of Leisure Activities on Cognitive Abilities (Model 5)

This study constructed a model with a time-invariant variable (education) and time-variant variables (smoking, alcohol drinking, leisure activities) and treated education, smoking and alcohol drinking as control variables, which were shown in **Figure 7**. Therefore, this paper mainly studied the effect of leisure activities on cognitive ability.

For model 5, $\chi^2(38) = 3.2$, $P = 0.000$, CFI = 0.929, RMSEA = 0.046, SRMR = 0.055. Education level has a significant effect on initial cognitive ability ($\gamma_0 = 0.241$, $P < 0.001$), which indicated that older adults with high education level have higher cognitive ability. In addition, the value of slope ($\gamma_1 = -0.071$, $P > 0.05$) and curve slope ($\gamma_2 = 0.027$, $P < 0.05$) indicated that the higher the level of education, the slower decline of cognitive ability of the older population. Therefore, Hypothesis 8 was supported.

Importantly, the results showed that at any time point, the more the leisure activity, the higher the cognitive ability (2008: $\beta = 0.218$, $P < 0.001$; 2011: $\beta = 0.125$, $P < 0.001$; 2014: $\beta = 0.195$, $P < 0.001$; 2018: $\beta = 0.499$, $P < 0.001$), which meant that leisure activities did promote the cognitive ability of the older people at every time point. Therefore, hypothesis 3 was supported.

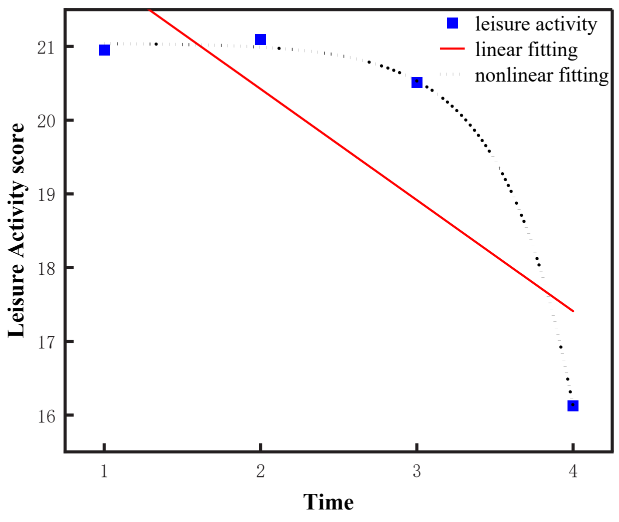


FIGURE 6 | Leisure activity fitting.

In order to examine the effect of different types of leisure activities on cognitive ability, leisure activities were categorized into two types according to the predominant element of each activity (Karp et al., 2006). Specifically, leisure activities were categorized into cognitive activities and non-exercise physical activities (Zhu et al., 2017). Cognitive activities required a cognitive component participation, such as reading books, listening to the radio, playing cards and participation in organized activity.

Non-exercise physical activity was not intended to develop and maintain fitness, which included doing housework, outdoor activity, keeping domestic animals or pets and gardening. Therefore, the total scores for both cognitive activities and non- exercise physical activities ranged from 0 to 20. If the cognitive activity score was greater than the leisure activity score, the individual would be considered participating in a cognitive activity; otherwise, non- exercise physical activity. *T* test was performed to compare the score of cognitive ability between different types of leisure activities. And the result was $P < 0.001$, which indicated the significant difference between cognitive activities and non- exercise physical activities.

And the effect of different types of activities on cognitive ability was shown in **Figure 8**. It can be suggested from the **Figure 8** that the cognitive ability of individuals who participated in cognitive activities was significantly higher than that of non-exercise physical activities participants in the first three periods. However, in the last period, it turned out the opposite.

Finally, it can be concluded that the impact of smoking on the cognitive ability was not significant in 2008, 2011, and 2018. In 2014, smoking had a significant impact on improving the cognitive ability of the older population ($\beta = 0.537$, SE = 0.236, $P < 0.05$). In addition, it was found that the effect of alcohol drinking on cognitive ability was positive in this study but it was not significant during the periods of four measurements.

Parallel Development Model (Model 6)

In order to avoid measurement errors to more accurately examine the relationship between leisure activities and cognitive ability, a

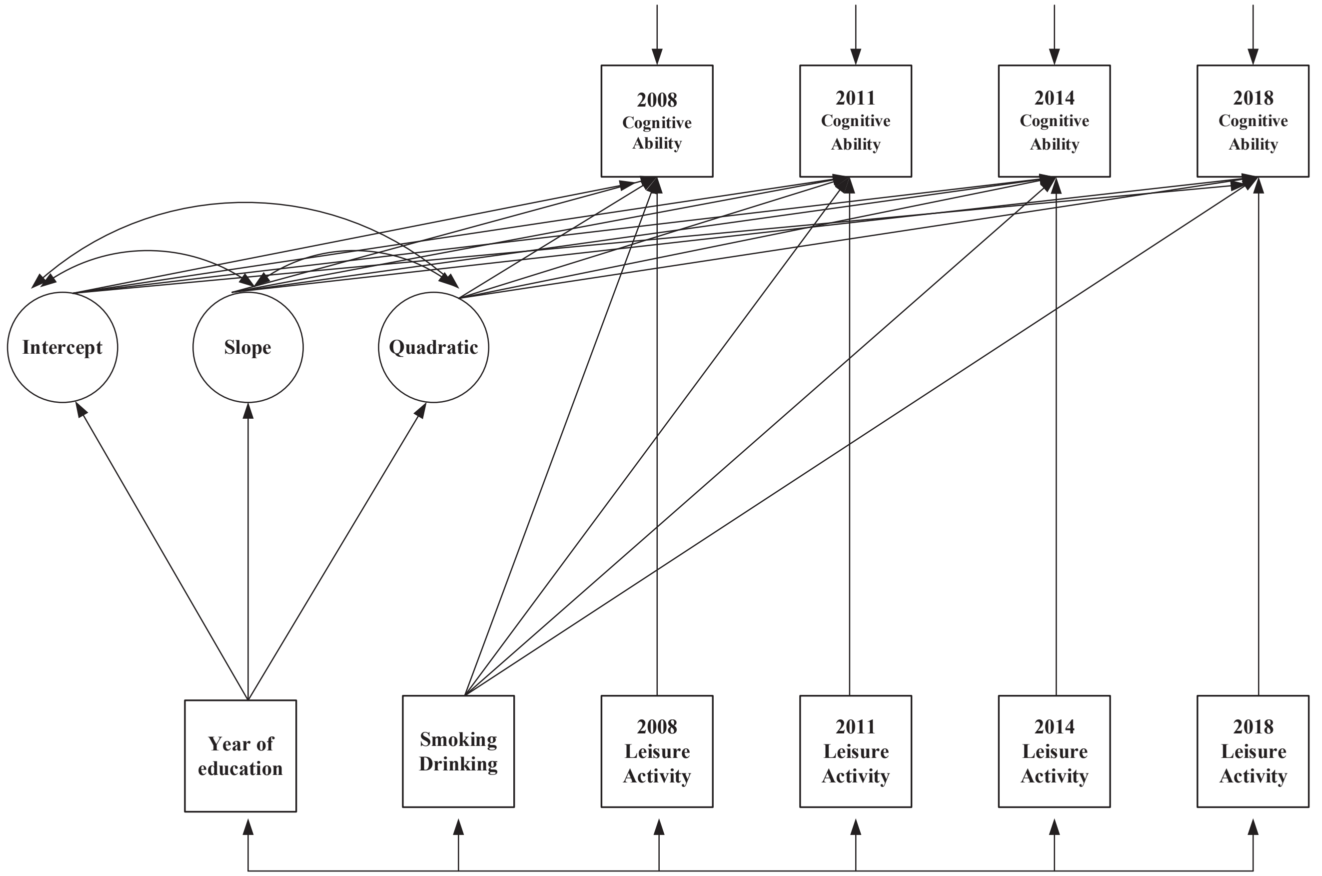


FIGURE 7 | Impact of leisure activities on the change track of cognitive ability.

parallel growth model was developed to examine the influence process between leisure activities and cognitive ability by setting up a regression equation between growth factors, which was denoted as Model 6 (M6). The intercept and slope of leisure activities were used to predict the increase of cognitive ability. The conceptual model was shown in **Figure 9**.

The model fitting metrics were as follows: $\chi^2(df) = 5.6$, $P = 0.000$, CFI = 0.949, SRMR = 0.09; RMSEA = 0.042. In the latent variable parallel growth model of leisure activities and cognitive ability of older adults, the regression coefficient of the intercept of leisure activities of older people on the intercept of cognitive ability was significant ($\beta = 0.43$, $P = 0.031$), indicating that the higher the initial level of leisure activities, the higher the initial level of cognitive ability among the older adults. Hypothesis 4 was supported. The intercept of leisure activities of the older population influenced the slope of their cognitive ability ($\beta = -0.40$, $P = 0.012$), indicating that the higher the initial level of leisure activities, the slower the decline of cognitive ability. Hypothesis 5 was supported. The slope of leisure activities had no significant effect on the intercept of cognitive ability ($\beta = 0.52$, $P = 0.09$), suggesting that the initial level of cognitive ability would not be influenced by the change rate of leisure activities. The slope of leisure activities had a significant impact on the slope of cognitive ability ($\beta = 0.82$, $P = 0.013$), indicating that the faster the level of leisure activities of the older population decreased, the

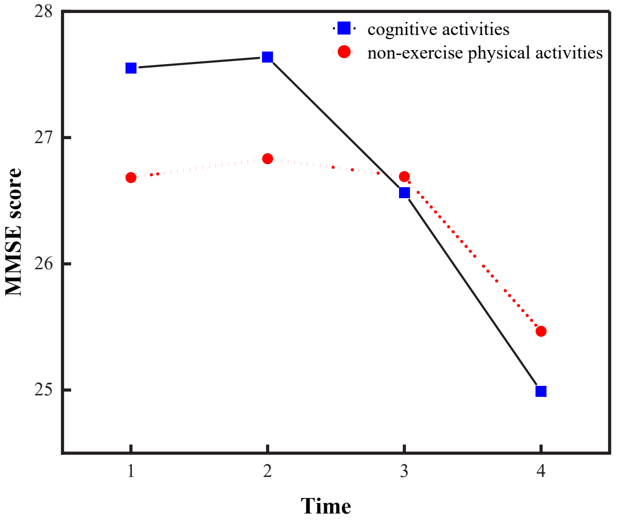


FIGURE 8 | Cognitive ability scores for different types of leisure activities.

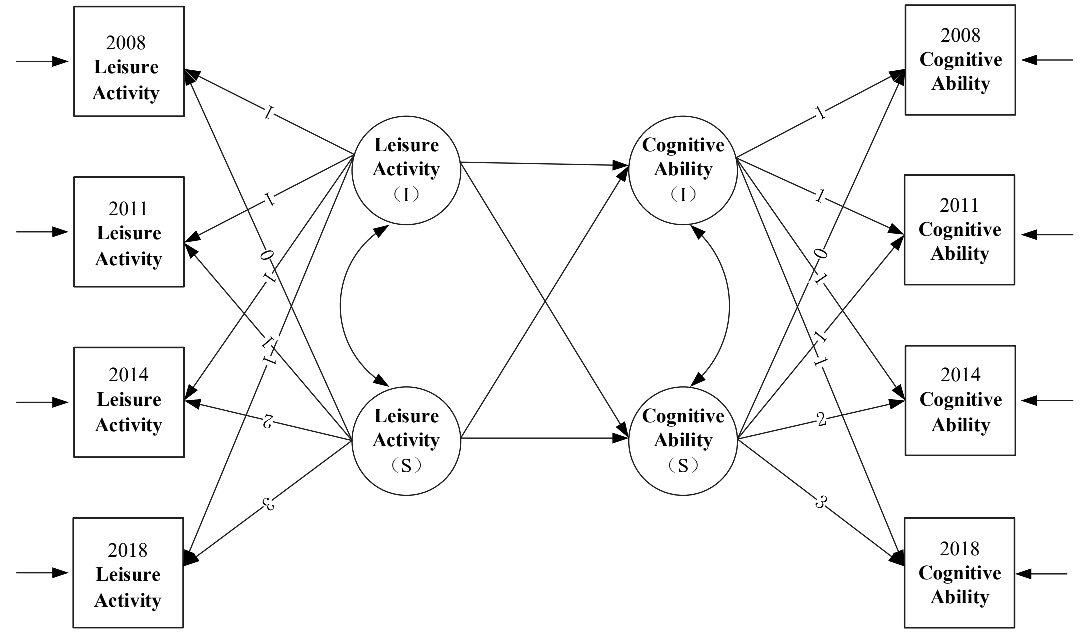


FIGURE 9 | Parallel development model. I denote intercept; S denote slope.

faster the level of cognitive ability decreased. Hypothesis 6 was supported. The specific fitting results of the model were shown in **Table 5**.

Cross-Lagged Regression Analysis (Model 7)

The latent variable growth model was used to study the dynamic characteristics of the variable. To further examine the leading lag relationship between leisure activities and cognitive abilities of the older population over time and to strengthen the demonstration for causal direction, the cross-lagged regression analysis was carried out in four measurements. Cross-lagged regression analysis can reveal complex relationships between two variables. The autoregressive effect of each variable was controlled by setting the stability coefficient, which was the best way to test the "Pure"effect among variables (Preacher, 2015) and used to understand how well one variable predicted another variable in general. A growing number of researchers believed that a combination of methods should be considered in order to obtain more robust conclusions in causal inference, thus allowing for sensitivity analysis in a broader sense (Curran and Bollen, 2001; De Stavola et al., 2006; Pakpahan et al., 2017). Therefore, the following cross-lagged regression model was constructed in this paper, as shown in **Figure 10**.

The results showed that the level of leisure activities of older adults in 2008 significantly positively predicted their cognitive ability in 2011 (β = 0.112, SE = 0.017, P = 0.000). The level of leisure activities of the older population in 2011 significantly positively predicted their cognitive ability in 2014 (β = 0.106, SE = 0.018, P = 0.000). The level of leisure activities of older adults in 2014 significantly positively predicted their cognitive ability in 2018 (β = 0.300, SE = 0.024,P = 0.000). Therefore, Hypothesis 7 was supported. Conversely, the level of cognitive ability of older adults in 2008 had no significant impact on their leisure activities in 2011 (β = 0.171, SE = 0.027, P = 0.061). The level of cognitive ability of older adults in 2011 had no significant impact on their leisure activities in 2014 (β = 0.188, SE = 0.026, P = 0.072). The level of cognitive ability of older adults in 2014 had no significant impact on their leisure activities in 2018 (β = 0.160, SE = 0.021, P = 0.085). The results also provided further evidence that there was no endogeneity problem in this study and the findings were scientifically valid.

DISCUSSION

To our knowledge, the current study was the first longitudinal study to investigate the change trajectory of leisure activity and long-term effects of leisure activity on cognitive ability. In addition, endogenous problems were considered in our study. What's more, the study was based on a large, representative Chinese sample, which ensured the generalizability of the findings. Finally, both time-variant and time-invariant covariates on older population's cognitive ability was assessed.

Changing Trajectory of Cognitive Ability

According to the results of the unconditional linear model of the cognitive ability of the older population, the cognitive ability of older adults showed a significant downward trend from 2008 to 2018, which was consistent with the existing research conclusions (Konttinen et al., 2016). With the increase of age, the cognitive function of older people gradually weakened, and

TABLE 5 | Model fitting metrics.

Model	χ^2(df)	CFI	SRMR	RMSEA	*P*
Model 5	3.2	0.929	0.055	0.046	0.000
Model 6	5.6	0.949	0.09	0.042	0.000

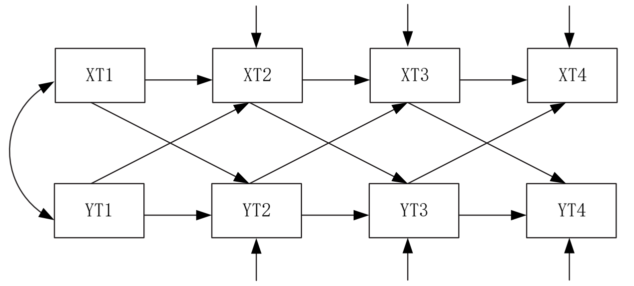


FIGURE 10 | Cross-lagged regression model.

their general ability, reaction ability, attention and calculation ability, memory, language understanding and self-coordination ability also further declined, which were determined mainly by the decline of biological changes in the brain (Bäckman et al., 2000). In addition, Unconditional non-linear model indicated that the cognitive ability of the older people showed a trend of slow decline at the beginning and then accelerated decline, which was consistent with the theory of human cognitive aging (Wingfield and Grossman, 2006). Over time, cognitive function decline in the older population accelerated. Especially from 2014 to 2018, the cognitive ability of older adults showed a precipitous decline, which may be because during the four measurement periods, the older population gradually transferred from the young middle-aged elderly in 2008 to the middle-aged or older elderly in 2018. This was the period when the cognitive and physical functions of the older population degraded the most. Accordingly, the cognitive ability of older adults deteriorated drastically over time, especially in the late-life.

Changing Trajectory of Leisure Activities

It was indicated that the level of leisure activities of older adults decreased significantly in the four measurement periods, which was consistent with previous studies (Feng et al., 2020). Since the subjects of this study were over 60 years old in the base period, with the increase of age, the older adults began to step into the recession period in the life cycle, and the physical function of older adults also decreased. Consequently, the elderly gradually suffered from diseases, such as arthritis, which severely restricted physical and recreational activities of the elderly (Zimmer et al., 1997). Even if leisure activities such as watching TV may increase over the late-life to some extent (Robinson et al., 2004), watching TV was a small part of all leisure activities which could not adverse the trajectory of leisure activities decline on the whole. In addition to changes in physical function, the reason for the decline in leisure activities among older adults may be due to changes in today's social interaction patterns. Nowadays, it was not uncommon for older adults to be unfamiliar with their next-door neighbors and thus social interactions-rated activities decreased (Feng et al., 2020). Furthermore, unconditional non-linear model indicated that the leisure activities of the older people showed a trend of slow decline at the beginning and then accelerated decline, which was line up with the development of physical function (Xu et al., 2018). After stepping into the oldest-old stage, it was uncommon and hard for older adults to maintain regular leisure activities. As a result, leisure activity levels in older adults declined at an accelerated rate later in the life cycle. Accordingly, the overall level of leisure activities of the older population gradually showed a downward trend over time.

Effects of Time-Variant and Time-Invariant Covariates on Cognitive Ability

Through the study of model 5, it was found that the cognitive ability was influenced by the level of education. Specifically, the higher the level of education, the higher the cognitive ability at the same period, which could be explained by that the older people with high education level are more likely to engage in work related to cognitive tasks, such as reasoning and memory and their cognitive ability will be improved through processing these tasks (Lindenberger et al., 1993). In addition, it was found that the higher level of education, the slower decline of cognitive ability, which was in line with the previous study and could be explained by that the likelihood of cognitive ability decline associated with lacunar infarcts was lower among individuals with high education compared to those with low education (Farfel et al., 2013). Education was a kind of solidified intelligence obtained at a young age, which can maintain a better condition for older adults and slower the degradation of cognitive ability (Mazzonna and Peracchi, 2012).

In addition, Although most studies indicated that smoking can significantly reduce cognitive ability (Ott et al., 2004; Nooyens et al., 2008), the positive effects of smoking on the cognitive ability of the older adults was found in 2014, which may be explained by the neuroprotective effects of nicotine in cigarettes from a biological point of view (Kihara et al., 1998; Mihailescu and Drucker-Colín, 2000). Additionally, the frequency of smoking was the key factor affecting cognitive ability. Due to the smoking habits and social habits of the older population, the smoking frequency of older adults was generally lower than that of the young. Therefore, it was found that smoking can improve the cognitive ability of the older population to a certain extent.

Similarly, it was revealed that alcohol drinking was a risk factor for the cognitive ability of the older population (Katja et al., 2014), however, it was found that the effect of alcohol drinking on cognitive ability was protective in this study even if it was not significant, which may be due to differences in the amount of alcohol drinking. A small amount of alcohol drinking was beneficial to physical and mental health, while excessive drinking will damage brain cells, thus damaging the cognitive function of the older adults. In addition, it may be related to the frequency of alcohol drinking. Regular and small consumption was beneficial to cognitive function, while occasional drinking will reduce the cognitive ability of older people (Horvat et al., 2015; Reas et al., 2016).

Last but not least, the positive impact of leisure activities on the cognitive ability of the older population was suggested. Specifically, it was indicated that the level of leisure activities of older adults significantly positively predicted their cognitive ability during the same period. In other words, the higher the level of leisure activities in 2008, the higher the level of cognitive ability in 2008. Similarly, the positive prediction relationship was still significant in 2011, 2014 and 2018, which can be explained by the cognitive reserve theory that the older population's participation in leisure activities may produce a more effective cognitive network, so as to provide a cognitive reserve and delay the decline of their physiological cognitive ability (Scarmeas and Stern, 2003; Cristina et al., 2010).

In addition, from the perspective of different types of activities, cognitive activities had a greater effect on cognitive ability than non- exercise physical activities during the first three measurement periods, Cognitive activities maintained the cognitive ability of older adults by stimulating brain activity. Therefore, during the first three measurement periods, the cognitive ability of the older population who participated in cognitive activities was significantly higher than the counterpart who participated in non- exercise physical activities.

Influence Process of Leisure Activities on the Change of Cognitive Ability

It was suggested that initial level of leisure activities in 2008 not only predicted the initial level of cognitive ability in 2008 but also the level of cognitive ability in 2011, 2014, and 2018. In addition, through the parallel development model, it was revealed that the higher the initial level of leisure activities, the slower the decline of cognitive ability, which was consistent with the previous analysis, that was, leisure activities were the protective factor of the cognitive ability of older adults (Park et al., 2019). The higher the level of leisure activities of the older population can effectively inhibit the decline of their cognitive ability. Cause it has been suggested that participation in leisure activities may provide cognitive reserve as an attitude toward an active lifestyle in older adults, which may delay the clinical manifestations of cognitive decline (Cristina et al., 2010). Furthermore, it was demonstrated that the rate of decline in leisure activities of older adults can predict the rate of decline in their cognitive ability, which meant that the faster the level of leisure activities decreased, the faster the level of cognitive ability decreased. On the contrary, the cognitive decline was also slower among older adults whose leisure activity levels declined more slowly, which indicated that the level of leisure activities and the level of cognitive ability had the consistent change trend, so the slope of leisure activities was significantly related to the slope of cognitive ability, and the change direction was consistent.

Finally, a cross-lagged regression analysis was conducted on the older population's leisure activities and cognitive ability, in order to explore the time sequence between leisure activities and cognitive ability. As a result, it was examined that the level of leisure activities of the older population could positively predict the subsequent cognitive ability, which further verified the causal inference between the current leisure activities and cognitive ability among older adults.

STUDY LIMITATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

The study was subject to a few limitations. First, this paper only considered the influence of educational factors on cognitive ability, and other time-invariant factors, such as gender, marriage, and region, had not been considered. Therefore, future studies should take more demographic variables into consideration. In addition, Since CLHLS was a face-to-face questionnaire, the older population's answers to leisure activity-related questions were only a subjective perception. Furthermore, the measurements of leisure activities were based on frequency in this study, but the quality and duration of leisure activity was more important. Therefore, in the future, objective data and the quality of leisure activities could be better to measure the leisure activities of the older population.

CONCLUSION

Based on the above discussion and analysis, the following research conclusions were drawn from this paper:

1. On the whole, the level of cognitive ability of older adults showed a non-linear decreasing trend, and the decreasing trend gradually increased over time. In addition, there was significant difference in the initial level and the rate of change of cognitive ability among older adults.
2. Overall, the level of leisure activities of the older population showed a non-linear decreasing trend, and the decreasing trend gradually increased over time. What's more, there were differences in the initial level of leisure activity and no differences in the decline rate of leisure activity level among the older population.
3. At every time point, the level of leisure activities had a significant positive impact on cognitive ability among older people, that was, the higher the level of leisure activities, the higher the level of cognitive ability. In addition, cognitive activities had a greater effect on cognitive ability than non-exercise physical activities.
4. The older population with high level of initial leisure activities had higher initial level of cognitive ability, and the decline of cognitive ability was slow.
5. The faster the level of leisure activities decreased, the faster the level of cognitive ability decreased.
6. The level of leisure activities in the previous period can positively predict the cognitive ability of the older population in the later period.
7. Education can significantly promote the initial cognitive ability of older adults, and the higher the level of education, the slower the decline of cognitive ability.

(8) For the older population, smoking shed a significant positive effect on cognitive ability to some extent and no significant effect was found between alcohol drinking and cognitive ability.

Overall, the correlation between cognitive ability and leisure activity suggested that more targeted interventions should be undertaken to promote existing leisure activities among older adults, especially cognitive activities.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Peking University (IRB00001052–13074). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

CZ: data curation, formal analysis, methodology, and resources. XZ and LZ: funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This study has been supported by the National Natural Science Foundation of China (Grant nos. 71804061 and 71974064), People's Republic of China.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.838878/full#supplementary-material

REFERENCES

Andel, R., Silverstein, M., and Kareholt, I. (2015). The role of midlife occupational complexity and leisure activity in late-life cognition. *J. Gerontol. Series B* 70, 314–321. doi: 10.1093/geronb/gbu110

Bäckman, L., Ginovart, N., Dixon, R. A., Wahlin, T.-B. R., Wahlin, A., Halldin, C., et al. (2000). Age-related cognitive deficits mediated by changes in the striatal dopamine system. *Am. J. Psychiatry* 157, 635–637. doi: 10.1176/ajp.157.4.635

Bennett, D. A., Arnold, S. E., Valenzuela, M. J., Brayne, C., and Schneider, J. A. (2014). Cognitive and social lifestyle: links with neuropathology and cognition in late life. *Acta Neuropathol.* 127, 137–150. doi: 10.1007/s00401-013-1226-2

Chan (1998). The Conceptualization and Analysis of Change Over Time: an Integrative Approach Incorporating Longitudinal Mean and Covariance Structures Analysis (LMACS) and Multiple Indicator Latent Growth Modeling (MLGM). *Org. Res. Methods* 1, 421–483. doi: 10.1177/109442819814004

Costa, D. A., Cracchiolo, J. R., Bachstetter, A. D., Hughes, T. F., Bales, K. R., Paul, S. M., et al. (2007). Enrichment improves cognition in AD mice by amyloid-related and unrelated mechanisms. *Neurobiol. Aging* 28, 831–844. doi: 10.1016/j.neurobiolaging.2006.04.009

Cristina, P. M., Gereon, N., and Ludger, P. (2010). Prevention of Dementia: Focus on Lifestyle. *Internat. J. Alzheimers Dis.* 2010, 1–9. doi: 10.4061/2010/393579

Curran, P. J., and Bollen, K. A. (2001). "The best of both worlds - Combining autoregressive and latent curve models," in *New methods for the analysis of change*, eds L. M. Collins and A. G. Sayer (Washington, D.C: American Psychological Association), 107–135. doi: 10.1037/10409-004

De Stavola, B. L., Nitsch, D., Silva, I. S., McCormack, V., Hardy, R., Mann, V., et al. (2006). Statistical issues in life course epidemiology. *Am. J. Epidemiol.* 163, 84–96. doi: 10.1093/aje/kwj003

Ding, M., Jia, N., Zhou, Y., Dong, X., and Tian, C. (2021). Associations Between Physical Activity and Cognitive Function, Daily Physical Function in Chinese With Heart Disease: a Cross-Sectional Study. *Reseach* 2021:631110. doi: 10.21203/rs.3.rs-631110/v1

Fancourt, D., and Steptoe, A. (2019). Television viewing and cognitive decline in older age: findings from the English Longitudinal Study of Ageing. *Sci. Rep.* 9, 1–8. doi: 10.1038/s41598-019-39354-4

Farfel, J. M., Nitrini, R., Suemoto, C. K., Grinberg, L. T., Ferretti, R., Leite, R., et al. (2013). Very low levels of education and cognitive reserve: a clinicopathologic study. *Neurology* 81, 650–657. doi: 10.1212/WNL.0b013e3182a08f1b

Feng, Q., Fong, J. H., Zhang, W., Liu, C., and Chen, H. (2020). Leisure Activity Engagement Among the Oldest Old in China, 1998–2018. *Am. J. Public Health* 110, 1535–1537. doi: 10.2105/AJPH.2020.305798

Haizhong, Z. (2014). Life cycle theory, life psychological cycle and assumption of psychological problems of the elderly. *Sci. Soc. Psychol.* 2, 32–35.

Hamer, M., and Stamatakis, E. (2014). Prospective Study of Sedentary Behavior, Risk of Depression, and Cognitive Impairment. *Med. Sci. Sports Exerc.* 46, 718–723. doi: 10.1249/MSS.0000000000000156

Hazzouri, A., Haan, M. N., Galea, S., and Aiello, A. E. (2011). Life-Course Exposure to Early Socioeconomic Environment, Education in Relatiosn to Late-Life Cognitive Function Among Older Mexicans and Mexican Americans. *J. Aging Health* 23, 1027–1049. doi: 10.1177/0898264311421524

Horvat, P., Richards, M., Kubinova, R., Pajak, A., Malyutina, S., Shishkin, S., et al. (2015). Alcohol consumption, drinking patterns, and cognitive function in older Eastern European adults. *Neurology* 84, 287–295. doi: 10.1212/WNL.0000000000001164

Hui-Xin, W., Anita, K., Bengt, W., and Laura, F. (2002). Late-Life Engagement in Social and Leisure Activities Is Associated with a Decreased Risk of Dementia: a Longitudinal Study from the Kungsholmen Project. *Am. J. Epidemiol.* 12, 1081–1087. doi: 10.1093/aje/155.12.1081

Hui-Xin, W., Jin, Y., Hendrie, H. C., Liang, C., Yang, L., Cheng, Y., et al. (2013). Late Life Leisure Activities and Risk of Cognitive Decline. *J. Gerontol. A Biol. Med.* 2, 205–213. doi: 10.1093/gerona/gls153

Iwasa, H., Yoshida, Y., Kai, I., Suzuki, T., Kim, H., and Yoshida, H. (2012). Leisure activities and cognitive function in elderly community-dwelling individuals in Japan: a 5-year prospective cohort study. *J. Psychosom. Res.* 72, 159–164. doi: 10.1016/j.jpsychores.2011.10.002

Jia, L., Meina, Q., Yue, F., Tan, Z., Yan, L., Cuibai, W., et al. (2020). Dementia in China: epidemiology, clinical management, and research advances. *Lancet Neurol.* 19, 81–92. doi: 10.1016/S1474-4422(19)30290-X

Kåreholt, I., Fau-Gatz, M. L. C., Fau-Parker, G. M., and Parker, M. G. (2011). Baseline leisure time activity and cognition more than two decades later. *Internat. J. Geriat. Psychiatry* 26, 1099–1166. doi: 10.1002/gps.2490

Karp, A., Paillard-Borg, S., Wang, H. X., Silverstein, M., Winblad, B., and Fratiglioni, L. (2006). Mental, physical and social components in leisure activities equally contribute to decrease dementia risk. *Dement Geriatr. Cogn. Disord* 21, 65–73. doi: 10.1159/000089919

Kashibayashi, T., Takahashi, R., Fujita, J., Fujito, R., and Kazui, H. (2021). Correlation between cerebral blood flow and olfactory function in mild cognitive impairment and Alzheimer's disease. *Internat. J. Geriat. Psychiatry* 36, 1103–1109. doi: 10.1002/gps.5527

Katja, F., Michael, R., and Christian, G. (2014). Gender-specific impact of personal health parameters on individual brain aging in cognitively unimpaired elderly subjects. *Front. Aging Neurosci.* 6:94. doi: 10.3389/fnagi.2014.00094

Kihara, T., Shimohama, S., Urushitani, M., Sawada, H., Kimura, J., Kume, T., et al. (1998). Stimulation of alpha4beta2 nicotinic acetylcholine receptors inhibits beta-amyloid toxicity. *Brain Res.* 792, 331–334. doi: 10.1016/S0006-8993(98)00138-3

Konttinen, H., Pekka, M., Karri, S., and Lei, Y. (2016). Association of socioeconomic status and cognitive functioning change among elderly Chinese people. *Age Ageing J. Br. Geriat. Soc. Br. Soc. Res. Ageing* 45, 674–680. doi: 10.1093/ageing/afw107

Lee, Y., Chi, I., and Palinkas, L. A. (2019). Retirement, leisure activity engagement, and cognition among older adults in the United States. *J. Aging Health* 31, 1212–1234. doi: 10.1177/0898264318767030

Lindenberger, U., Mayr, U., and Kliegl, R. (1993). Speed and intelligence in old age. *Psychol. Aging* 8, 207–220. doi: 10.1037//0882-7974.8.2.207

Liu, Y., and Lachman, M. E. (2020). Education and cognition in middle age and later life: the mediating role of physical and cognitive activity. *J. Gerontol. Series B* 75, e93–e104. doi: 10.1093/geronb/gbz020

Lv, Y.-B., Gao, X., Yin, Z.-X., Chen, H.-S., Luo, J.-S., Brasher, M. S., et al. (2018). Revisiting the association of blood pressure with mortality in oldest old people in China: community based, longitudinal prospective study. *BMJ* 361, 1–11. doi: 10.1136/bmj.k2158

Mao, C., Li, Z. H., Lv, Y. B., Gao, X., Kraus, V. B., Zhou, J. H., et al. (2020). Specific Leisure Activities and Cognitive Functions Among the Oldest-Old: the Chinese Longitudinal Healthy Longevity Survey. *J. Gerontol.* 75, 739–746. doi: 10.1093/gerona/glz086

Mazzonna, F., and Peracchi, F. (2012). Ageing, cognitive abilities and retirement. *SSRN Electr. J.* 56, 691–710. doi: 10.1016/j.euroecorev.2012.03.004

McArdle, J. J. (2008). Latent Growth Curve Analysis Using Structural Equation Modeling Techniques[M]. *Handbook Res. Methods Dev. Sci.* 2008, 340–366. doi: 10.1002/9780470756676.ch17

Mihailescu, S., and Drucker-Colín, R. (2000). Nicotine, brain nicotinic receptors, and neuropsychiatric disorders. *Archiv. Med. Res.* 31, 131–144. doi: 10.1016/S0188-4409(99)00087-9

National Bureau of Statistics of China (2021). *National Bureau of Statistics China.* Available online at: http://www.stats.gov.cn/ztjc/zdtjgz/zgrkpc/dqcrkpc/

Noice, T., Noice, H., and Kramer, A. F. (2014). Participatory arts for older adults: a review of benefits and challenges. *Gerontol.* 54, 741–753. doi: 10.1093/geront/gnt138

Nooyens, A., Gelder, B. V., and Verschuren, W. (2008). Smoking and cognitive decline among middle-aged men and women: the Doetinchem Cohort Study. *Am. J. Public Health* 98, 2244–2250. doi: 10.2105/AJPH.2007.130294

Ostlund, B. (2010). Watching television in later life: a deeper understanding of TV viewing in the homes of old people and in geriatric care contexts. *Scand. J. Caring Sci.* 24, 233–243. doi: 10.1111/j.1471-6712.2009.00711.x

Ott, A., Andersen, K., Dewey, M. E., Letenneur, L., Brayne, C., Copeland, J. R. M., et al. (2004). Effect of smoking on global cognitive function in nondemented elderly. *Neurology* 62, 920–924. doi: 10.1212/01.WNL.0000115110.35610.80

Pakpahan, E., Hoffmann, R., and Kröger, H. (2017). Statistical methods for causal analysis in life course research: an illustration of a cross-lagged structural equation model, a latent growth model, and an autoregressive latent trajectories model. *Internat. J. Soc. Res. Methodol.* 20, 1–19. doi: 10.1080/13645579.2015.1091641

Park, S., Choi, B., Choi, C., Kang, J. M., and J-Y, Lee (2019). Relationship between education, leisure activities, and cognitive functions in older adults. *Aging Ment. Health* 23, 1651–1660. doi: 10.1080/13607863.2018.1512083

Podsakoff, P. M., Mackenzie, S. B., Lee, J. Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J. Appl. Psychol.* 88, 879–903. doi: 10.1037/0021-9010.88.5.879

Preacher, K. J. (2015). Advances in mediation analysis: a survey and synthesis of new developments. *Annu. Rev. Psychol.* 66, 825–852. doi: 10.1146/annurev-psych-010814-015258

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., and Ferri, C. P. (2013). The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimer's Dement.* 9, 63–75. doi: 10.1016/j.jalz.2012.11.007

Pusswald, G., Elisa, T., Ilse, K.-E., Doris, M., Stefanie, K., Eduard, A., et al. (2015). Health-related quality of life in patients with subjective cognitive decline and mild cognitive impairment and its relation to activities of daily living. *J. Alzheimer's Dis.* 47, 479–486. doi: 10.3233/JAD-150284

Reas, E. T., Laughlin, G. A., Kritz-Silverstein, D., Barrett-Connor, E., and Mcevoy, L. K. (2016). Moderate, regular alcohol consumption is associated with higher cognitive function in older community-dwelling adults. *J. Prevent. Alzheimer's Dis.* 3, 105–113. doi: 10.14283/jpad.2016.89

Robinson, J. D., Skill, T., and Turner, J. W. (2004). Media usage patterns and portrayals of seniors. *Handbook Comm. Aging Res.* 2004, 423–450.

Ruining, X., and Yinge, L. (2015). The Status and Influential Factors of Cognitive Function Of the Community- dwelling Elderly. *Chin. J. Health Psychol.* 8, 1222–1224. doi: 10.13342/j.cnki.cjhp.2015.08.029

Rundek, T., and Bennett, D. A. (2006). Cognitive leisure activities, but not watching TV, for future brain benefits. *Neurology* 66, 794–795. doi: 10.1212/01.wnl.0000209497.38834.d7

Saji, N., Makizako, H., Suzuki, H., Nakai, Y., and Nakashima, T. (2020). Hearing impairment is associated with cognitive function in community-dwelling older adults: a cross-sectional study. *Archiv. Gerontol. Geriatr.* 93:104302. doi: 10.1016/j.archger.2020.104302

Scarmeas, N., and Stern, Y. (2003). Cognitive reserve and lifestyle. *J. Clin. Exp. Neuropsychol.* 25, 625–633. doi: 10.1076/jcen.25.5.625.14576

Silverstein, M., and Parker, M. G. (2002). Leisure activities and quality of life among the oldest old in Sweden. *Res. Aging* 24, 528–547. doi: 10.1177/0164027502245003

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *J. Internat. Neuropsychol. Soc. Jins* 8, 448–460. doi: 10.1017/S1355617702813248

Sunmin, L., Ichiro, K., Berkman, L. F., and Francine, G. (2003). Education, Other Socioeconomic Indicators, and Cognitive Function. *Am. J. Epidemiol.* 8, 712–720. doi: 10.1093/aje/kwg042

United Nations (2012). *World Population Prospects: The 2012 Revision.* New York, NY: United Nations Department of Economic and Social Affairs. Available online at: http://esa.un.org/unpd/wpp/index.htm

Weixiong, H. (2004). Social Communication Theories by Marx. *Teach. Res.* 8, 39–45.

Williams, L. J., Dunlop, J. P., and Abdi, H. (2012). Effect of age on variability in the production of text-based global inferences. *PLoS One* 7:e36161. doi: 10.1371/journal.pone.0036161

Wingfield, A., and Grossman, M. (2006). Language and the aging brain: patterns of neural compensation revealed by functional brain imaging. *J. Neurophys.* 96, 2830–2839. doi: 10.1152/jn.00628.2006

Wuzhen, L. (2008). The characteristics of contemporary Chinese leisure and the localization of leisure theory. *J. Guangzhou Sport Univ.* 2, 3–4. doi: 10.3969/j.issn.1007-323X.2008.02.002

Xu, M., Pirtskhalava, T., Farr, J. N., Weigand, B. M., Palmer, A. K., Weivoda, M. M., et al. (2018). Senolytics improve physical function and increase lifespan in old age. *Nat. Med.* 24, 1246–1256. doi: 10.1038/s41591-018-0092-9

Yue, Z., Nan, X., Huwei, L., and Erpeng, L. (2021). The evolution trend of availability of China's community-based care services and its impact on the cognitive function of elderly people: 2008-2018. *Internat. J. Equit. Health* 20, 1–11. doi: 10.1186/s12939-021-01544-w

Zeng, Y., Feng, Q., Hesketh, T., Christensen, K., and Vaupel, J. W. (2017). Survival, disabilities in activities of daily living, and physical and cognitive functioning among the oldest-old in China: a cohort study. *Lancet* 389, 1619–1629. doi: 10.1016/s0140-6736(17)30548-2

Zhu, X., Qiu, C., and Zeng, Y. (2017). Leisure activities, education, and cognitive impairment in Chinese older adults: a population-based longitudinal study. *Internat. Psychogeriat.* 29, 727–739. doi: 10.1017/S1041610216001769

Zimmer, Z., Hickey, T., and Searle, M. S. (1997). The Pattern of Change in Leisure Activity Behavior Among Older Adults With Arthritis1. G*erontologist* 37, 384–393. doi: 10.1093/geront/37.3.384

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 20 May 2022
doi: 10.3389/fpsyg.2022.853992

The Time Sequence of Face Spatial Frequency Differs During Working Memory Encoding and Retrieval Stages

*Anqing Wang, Enguang Chen, Hang Zhang, Chinheg H. Borjigin and Hailing Wang**

School of Psychology, Shandong Normal University, Jinan, China

Previous studies have found that P1 and P2 components were more sensitive to configural and featural face processing, respectively, when attentional resources were sufficient, suggesting that face processing follows a coarse-to-fine sequence. However, the role of working memory (WM) load in the time course of configural and featural face processing is poorly understood, especially whether it differs during encoding and retrieval stages. This study employed a delayed recognition task with varying WM load and face spatial frequency (SF). Our behavioral and ERP results showed that WM load modulated face SF processing. Specifically, for the encoding stage, P1 and P2 were more sensitive to broadband SF (BSF) faces, while N170 was more sensitive to low SF (LSF) and BSF faces. For the retrieval stage, P1 on the right hemisphere was more sensitive to BSF faces relative to HSF faces, N170 was more sensitive to LSF faces than HSF faces, especially under the load 1 condition, while P2 was more sensitive to high SF (HSF) faces than HSF faces, especially under load 3 condition. These results indicate that faces are perceived less finely during the encoding stage, whereas face perception follows a coarse-to-fine sequence during the retrieval stage, which is influenced by WM load. The coarse and fine information were processed especially under the low and high load conditions, respectively.

Keywords: face, configural processing, spatial frequency, working memory, coarse-to-fine sequence

Edited by:
Edwin James Burns,
Edge Hill University, United Kingdom

Reviewed by:
Peter James Hills,
Bournemouth University,
United Kingdom
Kristina Küper,
Bundeswehr Institute for Preventive
Medicine, Germany
Irina Shoshina,
Russian Academy of Sciences, Russia

Correspondence:
Hailing Wang
hailingwang@sdnu.edu.cn

Specialty section:
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

Received: *13 January 2022*
Accepted: *11 April 2022*
Published: *20 May 2022*

Citation:
Wang A, Chen E, Zhang H,
Borjigin CH and Wang H (2022) The
Time Sequence of Face Spatial
Frequency Differs During Working
Memory Encoding and Retrieval
Stages. Front. Psychol. 13:853992.
doi: 10.3389/fpsyg.2022.853992

INTRODUCTION

Visual stimuli perception combines multiple hierarchical levels of information. According to numerous studies, these different levels are processed at different times and follow a coarse-to-fine sequence (Boeschoten et al., 2005; Hegde, 2008; Goffaux et al., 2011; Petras et al., 2019). The global-local letter paradigm (Navon stimuli) has been widely used to investigate how these levels are processed. Results from these investigations suggest that global processing precedes the analysis of local information (Perfect et al., 2008; Gao et al., 2011).

As a complex visual stimulus, a face contains configural and featural information. Configural face information included first-order relational (i.e., the arrangement of face features with two eyes above a nose, which is above a mouth), second-order relational (i.e., the distance between facial features, such as the distance between eyes or the distance between the mouth and nose), and holistic processing (i.e., the gestalt of face features), whereas featural face information referred to the differences in face internal features, such as the shape or size of eyes and mouth (Maurer et al., 2002). Electrophysiological studies have importantly contributed to our understanding of the time course of configural and featural face processing in the human brain. Event-related potential (ERP) studies showed that configural and featural processing elicit separate responses as early as 100–250 ms after the onset of face presentation. For example, Wang et al. (2016) reported that second-order relational and featural face information elicited a larger P1 (approximately 128 ms) and P2 (approximately 248 ms), respectively, under the face-attended condition. These results, which were later replicated by several studies (Wang and Fu, 2018; Wang et al., 2020), suggest the temporal precedence of second-order relational over featural face processing and lend support to coarse-to-fine face processing. However, in the featural face sets utilized in the above-mentioned studies, the eyes or mouth in a given face were replaced by the eyes or mouth from a separate face and the feature location did not change. As Leder and Carbon (2006) pointed out, this kind of featural face manipulation requires second-order relational processing as the second-order relationships are also inherently changed. This subtle but important difference between second-order relational and featural manipulation might not have been purely dissociated in previous studies.

Notably, any early visual input contains different spatial frequencies based on the spatial frequency theory (Shulman and Wilson, 1987; Morrison and Schyns, 2001; Jeantet et al., 2018). Many previous studies have demonstrated an association between spatial frequency and global/local processing. Specifically, local details (i.e., fine information) of an image are represented by high spatial frequency (HSF) information, whereas large-scale visual details (i.e., coarse information) are represented by low spatial frequency (LSF) information (Rolls et al., 1985; Nagayama et al., 1995). Thus, previous studies have proposed that local (fine) and global (coarse) information are transmitted by relatively high and low spatial frequencies, respectively (Evans et al., 2000; Malinowski et al., 2002; Flevaris et al., 2009). Similarly, other studies investigating face-specific SF processing found that LSF and HSF information played a dominant role in face configural and featural processing, respectively (Boeschoten et al., 2005; Goffaux et al., 2005; Jeantet et al., 2019). For example, Goffaux et al. (2005) altered face second-order relational and featural information for both LSF and HSF conditions. They reported better performance during the LSF condition after changing face second-order relational information and better performance during the HSF condition after altering face featural information, suggesting that processing of second-order relational and featural information of faces can be dissociated by face spatial frequency. Thus, it is believed that configural face processing requires LSF information and featural face processing requires HSF.

It has been shown that the processing of LSF faces precedes that of HSF faces (Wang et al., 1998; Gao and Bentin, 2011; Peters et al., 2018), which supports the coarse-to-fine theory of facial processing. An fMRI study showed that most face-sensitive brain regions robustly responded to LSF faces during the early stage (75 ms) but exhibited decreased activation from 75 ms to 150 ms and, conversely, showed heightened activation during processing of HSF faces from 150 to 300 ms (Goffaux et al., 2011). Moreover, Jeantet et al. (2019) asked participants to perform a facial gender discrimination task on LSF and HSF faces and observed that LSF and HSF faces elicited a larger P1 and N170 (about 160 ms) amplitude, respectively. These results were in line with the coarse-to-fine theory of face perception, but not all studies have yielded congruent results (Goffaux et al., 2003; Mares et al., 2018). In the study by Goffaux et al. (2003), which employed a training-test paradigm, LSF faces elicited a larger N170 amplitude relative to HSF faces in the gender task but not in the familiarity task, and no difference in P1 elicited by LSF and HSF faces was observed. Taken together, these studies suggest that the time course of face spatial frequency processing is best examined with studies of ERP components. However, there is less evidence supporting this claim for the late-stage components, such as P2, which is linked to deeper processing and increasing sensory demand (Mercure et al., 2008). And there was a conflict about the P2 results. Some studies found that it was more sensitive to featural face information (Wang et al., 2016; Wang et al., 2018), while others found that it was more sensitive to second-order relational information (Mercure et al., 2008).

Furthermore, the coarse-to-fine face processing occurred when attentional resources were sufficient (Wang et al., 2016; Jeantet et al., 2019) as only one face stimulus was presented. It is unclear whether the time course of LSF and HSF face processing was impacted by the resource limitations, which can be characterized by working memory capacity. Previous studies demonstrated that working memory (WM) capacity for objects can vary based on qualitative differences in configural and featural processing. WM capacity can be measured by experimentally manipulating WM load, which is defined as the number of items maintained simultaneously (Morgan et al., 2008; Bauser et al., 2011). Generally, upright face stimuli are perceived by configural processing to a greater degree than inverted faces or other objects (Maurer et al., 2002). A behavioral study found that participants had a larger WM capacity for upright faces than for inverted faces and upright cars at a longer encoding duration, suggesting that configural processing might contribute to the increasing WM capacity when sufficient time is given for encoding (Curby and Gauthier, 2007). However, another behavioral study reported discrepant results that face configural processing was impaired under a high WM load condition relative to a WM load condition (Cheung and Gauthier, 2010).

Furthermore, within a delayed recognition task paradigm, previous studies found that WM load effects on the processing of upright faces and human body forms, which are perceived in a configural-based manner, are dissociated during the WM encoding and retrieval stages on early ERP components (Morgan et al., 2008; Bauser et al., 2011). Specifically, P1 (from load 1 to 2) and N170, which were evoked by upright faces, increased with increasing WM load during the encoding stage, whereas both P1 and N170 elicited by human body forms were not influenced by WM load at the encoding stage. But for the retrieval stage, both studies found that N170 decreased with increasing WM load and that P1 showed no differences across WM load. In addition, P3b (∼300–700 ms), which is strongly associated with WM and an attentional updating process (Polich, 2007), was suppressed by increasing WM load during the encoding and retrieval stages. In addition, Morgan et al. (2008) analyzed the N250r component, which is more negative for familiar faces than unfamiliar faces, and it is assumed that it reflects WM processes related to temporary activation of face recognition units. They found that the amplitude of N250r decreased as WM load increased at both encoding and retrieval stages. Taken together, these studies demonstrated that the neural mechanisms underlying the processing of these objects, which occur configurally, differ during WM encoding and retrieval periods, especially on P1 and N170 components. However, few studies focused on featural processing mechanisms during WM encoding and retrieval periods. Importantly, evidence supporting coarse-to-fine sequential processing during WM encoding and retrieval stages is currently lacking.

To this end, this study used a delayed recognition task to explore whether and how the WM load influences the time sequence of configural and featural face processing during WM encoding and retrieval stages. We examined configural and featural face processing by means of spatial frequency filtering. Stimuli were either unfiltered and contained all spatial frequencies (BSF), low-pass filtered (LSF), or high-pass filtered (HSF). Based on the previous ERP studies investigating configural and featural face processing (Mercure et al., 2008; Lv et al., 2015; Negrini et al., 2017; Wang et al., 2020), we analyzed the P1, N170, and P2 components to investigate the time course of LSF and HSF face processing during encoding and retrieval stages, and we analyzed P3b and N250r components to test the effect of WM load. We hypothesize that LSF processing will precede HSF processing at both memory encoding and retrieval stages if face processing occurs in a coarse-to-fine manner during both the working memory processing stages. If memory load impacts the time course of featural and configural face processing, then the ERP components elicited by faces with differing spatial frequencies would differ as a function of WM load. According to Wang and Fu (2018), coarse-to-fine processing occurs when attentional resources are sufficient; thus, in this study, we speculate that coarse-to-fine sequential processing will be found under low, but not high, WM load conditions.

MATERIALS AND METHODS

Participants

A power analysis using G*Power software version 3.01 (Faul et al., 2007) indicated that for an effect size of 0.2, at least 22 participants were required to achieve 80% power. In this study, 29 students (23 female students, aged 19.34 ± 1.21 years) completed the experiment, but one was excluded from the behavioral performance due to a data recording error. All participants were right-handed based on Edinburgh Handedness Test. All participants had normal or corrected-to-normal vision

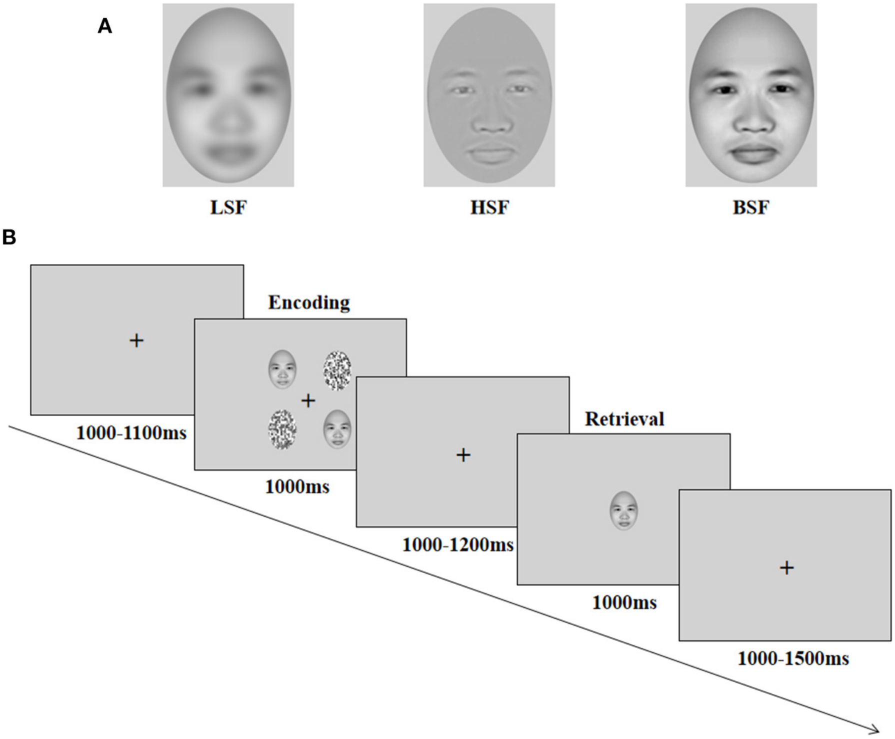


FIGURE 1 | Illustration of face stimuli **(A)** and procedure **(B)**. LSF = low spatial frequency; HSF = high spatial frequency; BSF = broadband spatial frequency. Reproduced with permission from Singular Inversions Inc., available at https://facegen.com.

without color blindness or mental illness based on self-report. All study procedures were approved by the local Institutional Review Board of the School of Psychology, Shandong Normal University. Written informed consent was collected from all participants prior to the study initiation.

Stimuli and Apparatus

Each face stimulus was presented without hair or glasses and with a neutral expression, as generated by FaceGen Modeller 3.5 (Toronto, Canada, http://en.softonic.com/) and edited using Adobe Photoshop 7.0 (Adobe Systems, San Jose, CA). According to previous studies (Holmes et al., 2005; Dale and Arnell, 2014), HSF faces were processed in Photoshop using the high-pass filter tool to select only spatial frequencies higher than 6 cycles/degree or 30 cycles/image of visual angle (i.e., a radius of 1.5 pixels); LSF faces were processed in Photoshop using the Gaussian blur tool to select only spatial frequencies lower than 1.3 cycles/degree or 6.5 cycles/image of visual angle (i.e., a radius of 3 pixels); original images with broadband and no spatial frequency filtering were used for BSF faces (**Figure 1A**). A total of 24 faces were created (8 for each spatial frequency category), with the same grayscale, size, background, luminance, and other physical characteristics. The distractors were the same as those used in previous studies (Wang et al., 2015) and were the scrambled pictures, which were constructed by sectioning each picture into 272 rectangles (in a 7 by 9 matrix) and randomly reassigning the locations of these rectangles, without replacement.

All the stimuli were presented on a 17-in Lenovo monitor (resolution: 1,024 × 768; refresh rate: 60 Hz) using E-Prime 2.0 (Pittsburgh, PA, USA). The viewing distance was 60 cm and the stimulus size was $4^\circ \times 5.5^\circ$ of the visual angle (113 × 156 pixels).

Procedure

A delayed recognition task was selected for the study, and it allowed separate ERP investigations for encoding and retrieval stages (Morgan et al., 2008). Each trial started with the presentation of a central cross for 1,000–1,100 ms, followed by the encoding display (a 2 × 2 image array) for 1,000 ms. In the encoding display, one, two, or three different face stimuli were presented, corresponding to working memory load 1, load 2, and load 3, respectively, and the remaining stimuli were scrambled distractor images. The four different stimuli were presented randomly on the four corners of the fixation cross, and the eccentricity of the stimuli on the encoding display (measured as the distance between the center of each stimulus and the horizontal or vertical line of the central fixation cross) was 3.8° for stimuli in the four positions (**Figure 1B**). Participants were instructed to remember the identity of the faces and ignore the scrambled distractors. The stimuli presentation was followed by a 1,000–1,200 ms delay. A test face was then presented on the center of the screen for 1,000 ms (retrieval display), and participants were required to judge whether the test face appeared on the encoding display by pressing the keyboard (50% trials for "yes," 50% trials for "no"). The intertrial interval was at random and ranged between 1 and 1.5 s.

A 3 (WM load: load 1, load 2, load 3) × 3 (face spatial frequency: low, high, broadband) within-subject design was adopted. Each condition had 120 trials, for a total of 1,080 trials.

Data Recording and Analysis

EEG data were collected from 64 channels with an EEG recording system produced by the NeuroScan company. Vertical electrooculography (VEOG) was recorded with two electrodes in the upper and lower parts of the left orbital frontal region, and horizontal electrooculography (HEOG) was recorded with two electrodes placed 1.5 cm laterally in both eyes. The reference electrode was located between Cz and CPz. The resistance between all electrodes and the scalp was less than 5 kΩ. The band-pass filter range was 0.01–100 Hz, and the sampling rate was 500 Hz.

EEG data were analyzed using Letswave 7 software (https://letswave.cn; Mouraux and Iannetti, 2008) operating in Matlab R2013b (Mathworks, Natick, MA, USA). At both encoding and retrieval stages, the EEG analysis window was between −200 to 700 ms, and a baseline was acquired 200 ms before the appearance of stimuli. EEG data were band-pass filtered at 0.1–30 Hz, and the average of all electrodes was used as a reference. Eye artifacts were identified by independent components analysis and removed from all EEG electrode traces, and the rejection standard was ± 75 mV. Only trials with correct responses were analyzed for the retrieval stage.

According to previous studies (Wang et al., 2015, 2016, 2020), we analyzed P1 (75–145 ms), N170 (140–200 ms), and P2 components (200–270 ms) acquired from P3/P4, P5/P6, P7/P8, PO3/PO4, PO5/PO6, PO7/PO8, and O1/O2 electrodes. Based on previous studies (Morgan et al., 2008; Bauser et al., 2011), P3b (300–600 ms) was analyzed from PZ, POZ, and OZ electrodes to examine WM load manipulation. Besides, our stimulus number is low, which might make the faces familiar. To test the familiarity effect, the mean amplitude of N250r (230–320 ms) was analyzed from P7/P8, PO7/PO8, and O1/O2 electrodes (Neumann and Schweinberger, 2008). The latency (time from start to peak) and amplitude (baseline to peak) of each component were selected as dependent variables. A 3 (face spatial frequency: BSF, LSF, HSF) × 3 (WM load: load 1, load 2, load 3) × 2 (hemisphere: left, right) repeated-measures ANOVA was performed for P1, N170, P2, and N250r components. A 3 (WM load: load 1, load 2, load 3) × 3 (face spatial frequency: BSF, LSF, HSF) repeated-measures ANOVA was performed on reaction time and accuracy performance and the P3b component. When necessary, *p* values were corrected using the Greenhouse-Geisser method. The Bonferroni correction was applied to account for multiple comparisons and *post-hoc* analyses were performed to understand interaction effects.

RESULTS

Behavioral Performance

There was a significant main effect of face spatial frequency on reaction time [$F_{(2,\ 54)} = 3.51$, $p = 0.042$, $\eta 2\ p = 0.115$]; however, *post-hoc* analyses did not detect any differences among the three SFs (LSF: 652 ms, HSF: 661 ms, BSF: 642 ms). A significant main

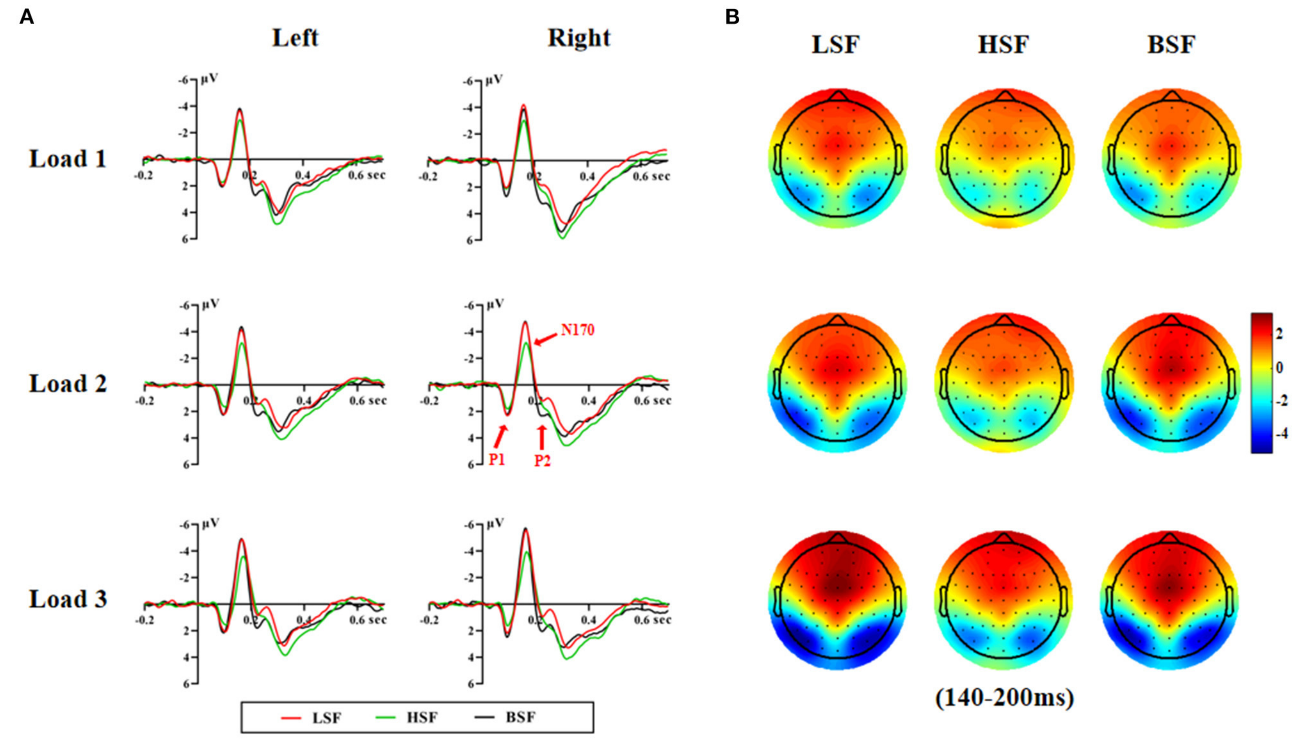


FIGURE 2 | Grand averaged ERP waveforms on the left and right hemispheres **(A)** and the topographic maps **(B)** during the encoding stage.

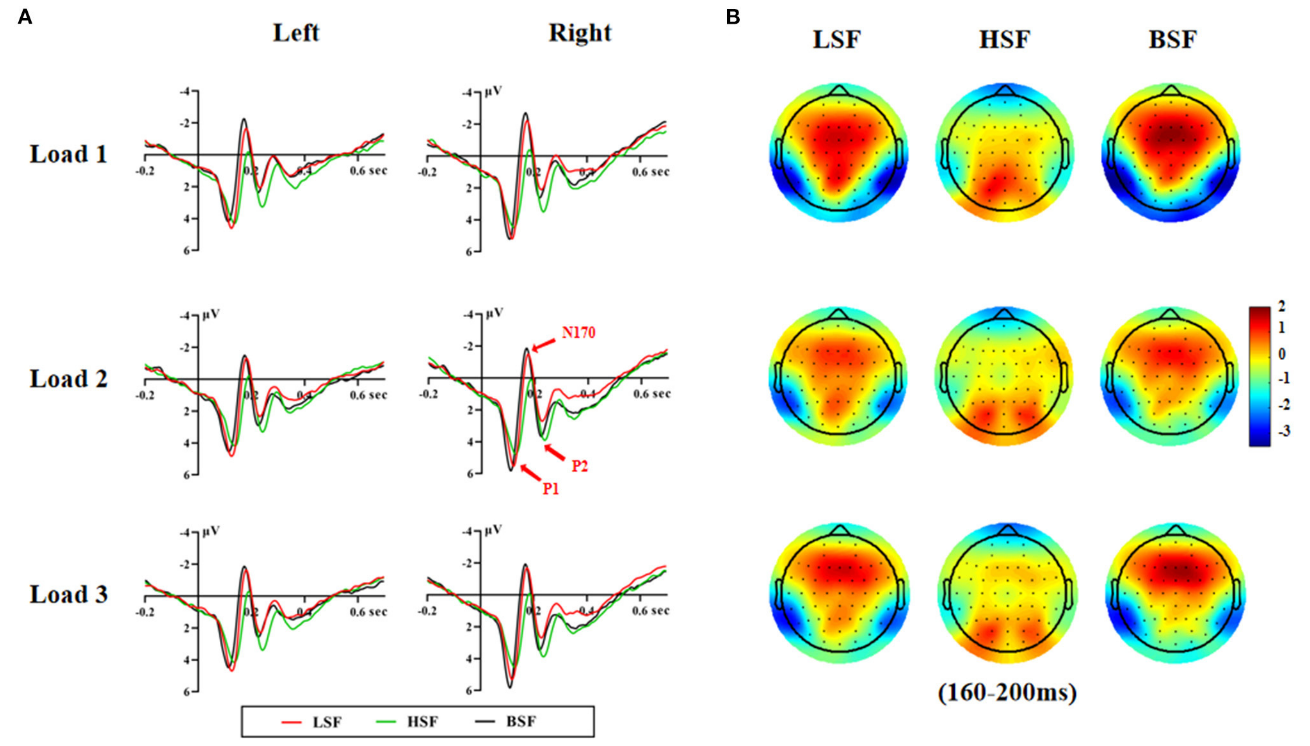


FIGURE 3 | Grand averaged ERP waveforms on the left and right hemispheres **(A)** and the topographic maps **(B)** during the retrieval stage.

effect of WM load was detected on reaction time [$F_{(2,\,54)} = 25.44$, $p < 0.001$, η2 $p = 0.485$], and *post-hoc* analyses determined that the reaction time for load 1 was shorter than for load 2 and load 3 (614 vs. 667 vs. 674 ms, $p < 0.001$). The interaction between face spatial frequency and WM load on reaction time was not significant [$F_{(4,\,108)} = 1.54$, $p = 0.209$].

There was a significant main effect of face spatial frequency on accuracy [$F_{(2,\,54)} = 39.13$, $p < 0.001$, η2 $p = 0.592$], and *post-hoc* results revealed that accuracy was higher for BSF than for LSF and HSF conditions (79 vs. 74 vs. 71%, $p < 0.05$). The decomposition of a significant main effect of WM load on accuracy [$F_{(2,\,54)} = 321.86$, $p < 0.001$, η2 $p = 0.923$] showed that accuracy decreased with increasing WM load (load 1: 88%, load 2: 75%, load 3: 62%, $p < 0.001$). Importantly, the interaction effect of face spatial frequency and WM load on accuracy was significant [$F_{(4,\,108)} = 3.22$, $p = 0.018$, η2 $p = 0.107$]. The accuracy for detection of BSF faces was higher than for LSF and HSF faces under load 1 [91 vs. 87 vs. 85%, $F_{(2,\,54)} = 17.73$, $p < 0.001$, η2 $p = 0.396$] and load 2 (80 vs. 73 vs. 70%, $F_{(2,\,54)} = 26.25$, $p < 0.001$, η2 $p = 0.493$), and it was significantly different among BSF, LSF, and HSF conditions under load 3 (67 vs. 62 vs. 58%, $F_{(2,\,54)} = 34.27$, $p < 0.001$, η2 $p = 0.559$). These results indicate that HSF appears to be less important than LSF under increasing WM load.

ERP Results

Figures 2, **3** show grand-average ERPs elicited by BSF, HSF, and LSF faces during encoding and retrieval stages, recorded from the temporal-occipital cortex. All stimuli evoked the canonical P1, N170, P2, P3b, and N250r components. **Table 1** shows the results on P1, N170, and P2 amplitude and latency during the encoding and retrieval stages.

Encoding Stage

P1

There was a significant main effect of face spatial frequency on P1 amplitude and latency [amplitude: $F_{(2,\,56)} = 7.30$, $p < 0.01$, η2 $p = 0.207$; latency: $F_{(2,\,56)} = 8.30$, $p < 0.001$, η2 $p = 0.229$]. BSF faces evoked larger and earlier P1 responses than HSF faces (3.23 vs. 2.88 μV, $p = 0.005$; 97 vs. 101 ms, $p = 0.001$).

N170

There were significant main effects of face spatial frequency [$F_{(2,\,56)} = 26.30$, $p < 0.001$, η2 $p = 0.484$] and WM load [$F_{(2,\,56)} = 61.00$, $p < 0.001$, η2 $p = 0.685$] on N170 amplitude. The N170 amplitude elicited by LSF faces and BSF faces was larger than for HSF faces (−5.16 vs. −5.23 vs. −4.00 μV, $p < 0.001$). Besides, there was a significant interaction between WM load and hemisphere on N170 amplitude [$F_{(2,\,56)} = 4.13$, $p < 0.05$, η2 $p = 0.129$]. *Post-hoc* analyses showed that the N170 amplitude elicited by load 3 was larger than that for load 2 and load 1 on both hemispheres [left: −5.07 vs. −4.47 vs. −4.03 μV, $F_{(2,\,56)} = 34.83$, $p < 0.001$, η2 $p = 0.554$; right: −5.85 vs. −4.96 vs. −4.41 μV, $F_{(2,\,56)} = 52.06$, $p < 0.001$, η2 $p = 0.650$]. These results indicated that N170 amplitude was more sensitive to configural face processing than featural face processing, which was not modulated by WM load.

There were significant main effects of spatial frequency [$F_{(2,\,56)} = 4.19$, $p = 0.033$, η2 $p = 0.130$] and WM load [$F_{(2,\,56)} = 16.20$, $p < 0.001$, η2 $p = 0.367$] on N170 latency. More importantly, the interaction between spatial frequency and WM load was significant [$F_{(4,\,112)} = 6.78$, $p < 0.001$, η2 $p = 0.195$]. The N170 latency for BSF faces (161 ms) was earlier than that for LSF (166 ms) and HSF faces (168 ms) under the load 3 condition [$F_{(2,\,56)} = 7.74$, $p < 0.001$, η2 $p = 0.217$]. There were no significant differences in N170 latency among spatial frequencies for the load 1 and load 2 conditions.

P2

There were significant main effects of spatial frequency [amplitude: $F_{(2,\,56)} = 20.53$, $p < 0.001$, η2 $p = 0.423$; latency: $F_{(2,\,56)} = 7.87$, $p = 0.002$, η2 $p = 0.219$] and WM load (amplitude: $F_{[2,\,56]} = 45.52$, $p < 0.001$, η2 $p = 0.619$) on P2 amplitude and latency. BSF faces (3.49 μV) elicited a larger P2 than LSF (2.33 μV) and HSF faces (2.56 μV, $p < 0.001$). Load 1 (3.55 μV) elicited a larger P2 than load 2 (2.76 μV) and load 3 (2.07 μV, $p < 0.001$).

There was also a significant interaction effect of spatial frequency and WM load on P2 latency [$F_{(4,\,112)} = 3.93$, $p = 0.008$, η2 $p = 0.123$]. Furthermore, the interaction of spatial frequency and WM load and hemisphere was also significant [$F_{(4,\,112)} = 3.09$, $p = 0.023$, η2 $p = 0.099$]. *Post-hoc* analysis showed that the P2 latency evoked by HSF faces was later than that for LSF and BSF faces [239 vs. 229 vs. 231 ms, $F_{(2,56)} = 14.61$, $p < 0.001$, η2 $p = 0.343$] on the right hemisphere for the load 3 condition. On the left hemisphere, the interaction between spatial frequency and WM load was not significant. These results indicated that configural face information was processed earlier than featural face information in the right hemisphere under a higher WM load.

P3b

There was a significant main effect of WM load on P3b amplitude [$F_{(2,\,56)} = 31.54$, $p < 0.001$, η2 $p = 0.530$], and *post-hoc* analyses showed that the amplitude elicited by load 1 was larger than that for load 2 and load 3 (load 1: 4.51 μV; load 2: 3.79 μV; load 3: 3.79 μV, $p < 0.001$).

N250r

Results revealed the significant main effects of WM load [$F_{(2,\,56)} = 50.87$, $p < 0.001$, η2 $p = 0.645$] and face spatial frequency [$F_{(2,\,56)} = 18.85$, $p < 0.001$, η2 $p = 0.402$] on the N250r amplitude. The N250r amplitude decreased with the increase in WM load (3.25 vs. 2.15 vs. 1.17 μV, $p < 0.01$). LSF faces elicited a lower N250r than HSF faces and BSF faces (1.41 vs. 2.65 vs. 2.50 μV, $p < 0.01$).

Retrieval Stage

P1

There were significant main effects of face spatial frequency [$F_{(2,\,56)} = 4.83$, $p < 0.05$, η2 $p = 0.147$] and WM load [$F_{(2,\,56)} = 3.65$, $p < 0.05$, η2 $p = 0.115$] on P1 amplitude. The interaction of spatial frequency and hemisphere on P1 amplitude was significant [$F_{(2,\,56)} = 5.89$, $p < 0.01$, η2 $p = 0.174$], which upon decomposition indicated that BSF faces evoked larger P1

TABLE 1 | The amplitude (μV) and latency (ms) values of the P1, N1, and P2 components for low-spatial frequency (LSF) and high-spatial frequency (HSF) face during the encoding and retrieval stages.

	P1			**N170**			**P2**		
	LSF	**HSF**	***p***	**LSF**	**HSF**	**p**	**LSF**	**HSF**	***p***
Amplitude									
Encoding	3.10 ± 0.29	2.88 ± 0.26	0.161	−5.16 ± 0.66	−4.00 ± 0.64	0.001	2.33 ± 0.77	2.56 ± 0.75	0.678
Retrieval	6.21 ± 0.34	5.80 ± 0.35	0.099	−2.26 ± 0.52	−1.08 ± 0.56	0.001	3.02 ± 0.51	4.34 ± 0.50	0.001
Latency									
Encoding	100 ± 2	101 ± 2	0.459	163 ± 2	164 ± 2	0.614	231 ± 2	235 ± 2	0.029
Retrieval	122 ± 2	125 ± 2	0.061	179 ± 2	183 ± 2	0.120	232 ± 2	240 ± 2	0.001

amplitudes than HSF faces on the right hemisphere [6.56 vs. 5.80 μV, $F_{(1, 28)} = 16.06$, $p < 0.001$, $\eta 2\ p = 0.364$].

There were also significant main effects of spatial frequency [$F_{(2, 56)} = 56.36$, $p < 0.001$, $\eta 2\ p = 0.668$] and hemisphere [$F_{(1, 28)} = 6.68$, $p < 0.05$, $\eta 2\ p = 0.193$] on P1 latency. P1 latencies elicited by BSF faces were earlier than for LSF and HSF faces (112 vs. 122 vs. 125 ms, $p < 0.001$), and they were also earlier on the right relative to the left hemisphere (118 vs. 122 ms, $p < 0.05$).

N170

There were significant main effects of face spatial frequency [$F_{(2, 56)} = 30.36$, $p < 0.001$, $\eta 2\ p = 0.520$] and WM load [$F_{(2, 56)} = 7.85$, $p < 0.01$, $\eta 2\ p = 0.219$] on N170 amplitude. There were also significant interaction effects between WM load and hemisphere [$F_{(2, 56)} = 7.15$, $p < 0.001$, $\eta 2\ p = 0.203$] and between spatial frequency and WM load [$F_{(4, 112)} = 2.85$, $p < 0.05$, $\eta 2\ p = 0.092$] on N170 amplitude (**Figure 4A**). *Post-hoc* analysis showed that the N170 amplitudes elicited by HSF faces were lower than those from LSF and BSF faces under the load 1 [−1.17 vs. −2.56 vs. −3.14 μV, $F_{(2, 56)} = 31.60$, $p < 0.001$, $\eta 2\ p = 0.530$], load 2 [−1.00 vs. −1.98 vs. −2.40 μV, $F_{(2, 56)} = 19.76$, $p < 0.001$, $\eta 2\ p = 0.414$], and load 3 conditions [−1.06 vs. −2.25 vs. −2.57 μV, $F_{(2, 56)} = 21.28$, $p < 0.001$, $\eta 2\ p = 0.432$]. These results indicated that N170 was more sensitive to configural face processing than featural face processing, especially under the load 1 condition.

There were significant main effects of face spatial frequency [$F_{(2, 56)} = 37.24$, $p < 0.001$, $\eta 2\ p = 0.571$] and hemisphere [$F_{(1, 28)} = 8.72$, $p < 0.01$, $\eta 2\ p = 0.237$] on N170 latency. N170 latency was later for LSF and HSF faces than BSF faces (179 vs. 183 vs. 171 ms, $p < 0.001$) and was later in the left hemisphere compared to the right hemisphere (180 vs. 176 ms).

P2

There were significant main effects of face spatial frequency [$F_{(2, 56)} = 14.05$, $p < 0.001$, $\eta 2\ p = 0.334$] and WM load [$F_{(2, 56)} = 10.13$, $p < 0.001$, $\eta 2\ p = 0.266$] on P2 amplitude. The interaction between WM load and hemisphere was significant [$F_{(2, 56)} = 5.61$, $p = 0.015$, $\eta 2\ p = 0.167$]. The load 1 P2 amplitude was lower than that for load 2 and load 3 [3.42 vs. 4.13 vs. 4.03 μV, $F_{(2, 56)} = 14.60$, $p < 0.001$, $\eta 2\ p = 0.343$] in the right hemisphere. Importantly, the interaction of spatial frequency and WM load was also significant [$F_{(4, 112)} = 2.75$, $p = 0.037$, $\eta 2\ p = 0.089$; **Figure 4B**]. HSF faces evoked larger P2 amplitudes than LSF and BSF faces [4.18 vs. 2.76 vs. 3.06 μV, $F_{(2, 56)} = 15.58$, $p < 0.001$, $\eta 2\ p = 0.357$] under load 1. Amplitudes were larger for HSF and BSF faces than LSF faces [4.43 vs. 3.93 vs. 3.20 μV, $F_{(2, 56)} = 10.77$, $p < 0.001$, $\eta 2\ p = 0.278$] under load 2, with no significant difference between P2 amplitudes from HSF and BSF faces. For load 3, HSF faces gave larger amplitudes than LSF faces [4.40 vs. 3.10 μV, $F_{(1, 28)} = 24.54$, $p < 0.001$, $\eta 2\ p = 0.467$], while P2 amplitude from BSF faces did not differ with that from either LSF or HSF faces. These results indicated that P2 was more sensitive to featural face processing than configural face processing, especially under the load 3 condition.

There were significant main effects of face spatial frequency [$F_{(2, 56)} = 27.90$, $p < 0.001$, $\eta 2\ p = 0.499$] on P2 latency. There were significant differences among the P2 latencies elicited by BSF, LSF, and HSF faces (229 vs. 232 vs. 240 ms, $p < 0.05$).

P3b

Results revealed a significant main effect of WM load [$F_{(2, 56)} = 6.73$, $p < 0.05$, $\eta 2\ p = 0.194$] on P3b amplitude. P3b amplitude was lower for load 3 than for load 2 and load 1 (−0.171 vs.0.048 vs.0.339 μV, $p < 0.05$).

N250r

There was a significant main effect of WM load [$F_{(2, 56)} = 17.64$, $p = 0.000$, $\eta 2\ p = 0.387$], due to load 1 eliciting a more negative N250r amplitude relative to other conditions (−0.648 vs. 0.356 vs. 0.158 μV, $p < 0.01$). In addition, the main effect of face spatial frequency was significant [$F_{(2, 56)} = 8.58$, $p = 0.002$, $\eta 2\ p = 0.235$], showing that LSF faces induced a more negative amplitude than HSF faces (−0.60 vs. 0.64 μV, $p < 0.01$).

DISCUSSION

This study investigated whether and how working memory load impacts coarse-to-fine face processing during WM encoding and retrieval stages. Our findings showed that for the encoding stage, P1 and P2 amplitudes were more sensitive to BSF relative to HSF faces, while N170 amplitude was more sensitive to LSF and BSF relative to HSF faces. P1 and N170 (under the load 3 condition) latencies were earlier for BSF than HSF faces, and P2 latency on the right hemisphere was earlier for LSF and BSF than HSF faces under the load 3 condition. The results indicated that featural face information appears to be less important than

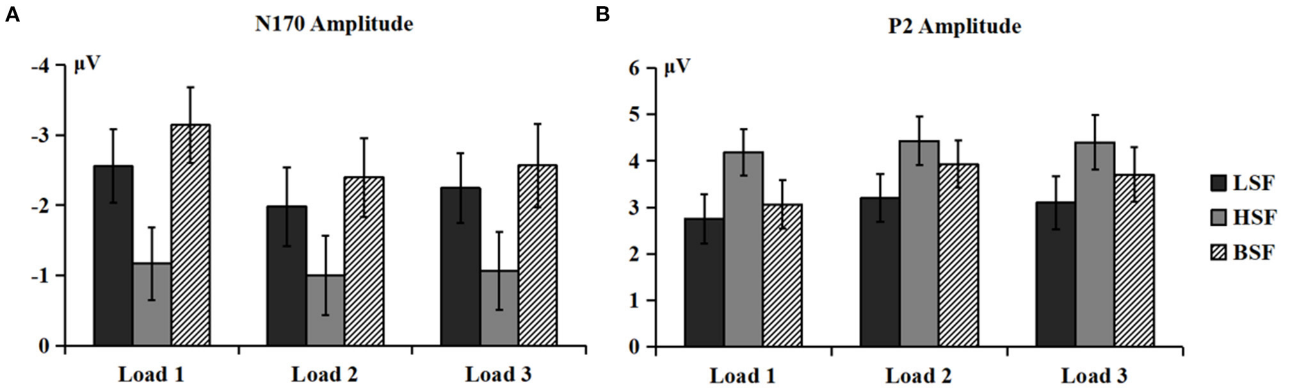


FIGURE 4 | Mean amplitudes of N170 **(A)** and P2 **(B)** components (P3/P4, P5/P6, P7/P8, PO3/PO4, PO5/PO6, PO7/PO8, and O1/O2) elicited by LSF, HSF, and BSF faces under different WM load conditions.

configural face information during the WM encoding stage (from N170 to P2 component). For the retrieval stage, P1 in the right hemisphere was more greatly influenced by BSF than HSF faces. N170 amplitude was more sensitive to BSF and LSF faces than HSF faces, especially under the load 1 condition. P2 amplitude was more influenced by HSF faces than LSF faces, especially under the load 3 condition. BSF faces had earlier latency than LSF and HSF faces among P1, N170, and P2. The results indicated that face perception follows a coarse-to-fine sequence during the WM retrieval stage. Specifically, coarse information (i.e., LSF, on N170) was processed first and more rapidly, followed by the processing of fine information (i.e., HSF, on P2). Additionally, P3b amplitude decreased as WM load increased at both the encoding and retrieval stages, which is consistent with previous studies (Morgan et al., 2008) and shows the successful manipulation of WM load. LSF faces induced more negative N250r than HSF faces during both encoding and retrieval stages, which was in line with the idea that familiar faces are processed in a configural manner relative to unfamiliar faces (Ramon and Rossion, 2012).

The LSF and HSF Face Processing During the Encoding Stage

Previous studies reported that P1 was a robust component that was associated with holistic face processing and was more sensitive to LSF faces (Nakashima et al., 2008; Jeantet et al., 2019), which was in line with the assumption that configural information plays an essential role in early face processing as it could efficiently provide the whole structure of the face (Freire et al., 2000; James et al., 2001; Itier and Taylor, 2002, 2004). Conversely, others reported that P1 was more sensitive to BSF faces (Pourtois et al., 2005; Peters et al., 2013; Craddock et al., 2015). Consistent with these latter findings, our results showed that BSF faces elicited a larger and earlier P1 relative to HSF faces at the encoding stage. BSF faces were from the upright non-filtered faces, which are processed in a more configural way compared with inverted faces (Schwaninger et al., 2003; Kimchi and Amishav, 2010; Cousins et al., 2020). Interestingly, we found that LSF and HSF faces cannot be purely dissociated based on the early P1 component during the encoding stage, which might be reconciled by the assumption that HSF faces also inevitably contain configural information, less than LSF faces but enough to produce an effect similar to LSF faces during early face processing (Halit et al., 2006). The cutoff value of spatial frequency, which varied depending on studies, might be another important factor. LSF values were fewer than 8, 5, 4, or 3 cycles/image, and HSF values were above 35, 30, 24, 22, or 15 cycles/image (Jeantet et al., 2019). Our LSF and HSF faces might include mid-band spatial frequencies (MSF), which are important in face processing (for a review, see Jeantet et al., 2018). Thus, our results suggest that P1 was less sensitive to LSF and HSF faces during the encoding stage.

The impact of spatial frequencies on N170 modulation has been controversial. Our results replicated previous findings that N170 was more negative following the presentation of BSF or LSF faces relative to HSF faces during the encoding stage (Goffaux et al., 2003; Peters et al., 2013; Yao and Zhao, 2019), but are in contrast to other studies reporting a larger N170 in response to HSF faces (Nakashima et al., 2008; Jeantet et al., 2019) or reporting no association between N170 amplitude and face spatial frequency (Holmes et al., 2005). How can these discrepancies be reconciled? One answer may stem from selective attention to spatial location. Selective attention to objects in the periphery promotes LSF processing, and attention to foveal location promotes HSF processing (Shulman and Wilson, 1987; Carrasco et al., 2006). In this study, face stimuli were randomly presented in the four corners of the fixation cross during the encoding stage, directing attention to the periphery, hence the N170 amplitude was greater for LSF faces relative to HSF faces. Thus, the current N170 reflected the configural processing during the encoding stage.

Although many studies found that P2 is more sensitive to LSF pictures than HSF pictures (Mathes and Fahle, 2007; De Cesarei et al., 2013; Yang and Chan, 2015), P2 amplitude is also augmented by visual stimuli with broadband information (Hansen et al., 2011). In this study, P2 amplitude was more sensitive to BSF faces than to LSF and HSF faces during the encoding stage. According to previous research (Craddock et al., 2015), this might be related to the fact that BSF faces were upright non-filtered faces, which had higher spectral power and could drive stronger responses compared to faces with lower or higher spectral power. In addition, BSF faces contain not only configural and featural information but also MSF information (Halit et al., 2006). Previous research found that MSF information plays an important role in face processing (Collin et al., 2012; Jeantet et al., 2018) as it is intermixed with LSF and HSF information and can be processed in a similar way to the non-filtered BSF faces (Jeantet et al., 2019). Moreover, Parker and Costen (1999) demonstrated that recognition was more accurate and rapid for MSF faces than for LSF or HSF faces, which was also supported by later ERP research (Collin et al., 2012). These combined results led us to postulate that it might be the MSF information that made P2 highly sensitive to BSF faces during the encoding stage. Thus, our finding indicated that LSF and HSF faces cannot be dissociated from P2 amplitude during the encoding stages. But, the P2 latency showed that LSF faces were processed earlier than HSF faces.

In sum, for the encoding stage, LSF and HSF faces cannot be dissociated on the P1 component but can be dissociated on the N170 and P2 components, which were more sensitive to LSF faces than to HSF faces. The result is in line with the hypothesis that the N170 is a marker for face structural encoding and is linked to configural face processing (Eimer, 2011).

The Time Course of LSF and HSF Face Processing During the Retrieval Stage

Consistent with our results at the encoding stage, the results at the retrieval stage showed that P1 was less sensitive to LSF and HSF faces and that N170 was more sensitive to LSF faces than HSF faces. However, the P2 was more positive for HSF faces relative to BSF and LSF faces during the retrieval stage, suggesting that P2 was more sensitive to face featural information. Although the sensitivity of P2 to featural processing has rarely been observed, several previous studies reported that faces with featural modifications elicited a larger P2 amplitude than faces with second-order relational modifications (Wang et al., 2016; Wang and Fu, 2018). Moreover, using Chinese and Western participants, Wang et al. (2020) found that P2 in Chinese participants was more sensitive to own-race faces with featural modifications. The reverse result was observed for Western participants, suggesting that cultural variation might account for this discrepancy.

It is worth emphasizing that the P2-enhancing effect of HSF faces was only observed at the retrieval stage and not at the encoding stage. It might be that faces in the encoding phase were processed in a parallel way not relying on in-depth visual analysis (cortical visual feedback; Mercure et al., 2008). Therefore, coarse facial information was more important during encoding, while during memory retrieval finer featural processing was adopted to match the encoded faces to the target at late stages (Bauser et al., 2011). Moreover, P2 could reflect the comparison between experimental face stimulus features and mental templates for task-related features (Potts, 2004); hence the facial features were of particular importance during the retrieval stage and evoked larger P2 waveforms. Taken together, only P2 reflected that encoding and retrieval stages were dissociated for face spatial frequencies.

In sum, for the retrieval stage, N170 was also more sensitive to LSF faces than to HSF faces, whereas P2 was more sensitive to HSF faces than to LSF faces. The processing priority of LSF faces during the retrieval stage is in line with previous neuronal findings that information processed in the LSF-sensitive magnocellular pathway has a faster cortical arrival than information processed in the HSF-sensitive parvocellular pathway (Laycock et al., 2007).

The Role of WM Load in Face SF Processing

This study found that WM load modulated face SF processing, which was reflected in the behavioral results and ERP responses. Consistent with previous studies (Morgan et al., 2008; Bauser et al., 2011), the P3b amplitude decreased as WM load increased during both the encoding and retrieval stages. Moreover, at the retrieval stage, the N170 amplitude evoked by LSF and BSF faces did not differ but was larger than that for HSF faces, especially under the load 1 condition. These results indicated that BSF faces tended to be processed in a configural manner under a lower WM load, which agrees with previous findings that N170 reflects first-order relational processing (Maurer et al., 2002). However, our behavioral results showed higher accuracy in processing LSF faces relative to HSF faces under the high WM load condition, which might be related to the limited WM capacity for faces. Previous studies showed that the maintenance capacity for faces is approximately two (Towler et al., 2016), which might not have allowed participants to process the details and features of faces in the higher WM load conditions (load 3).

However, compared with the N170 component, we found a different result for the P2 component during WM retrieval. Specifically, the P2 amplitude evoked by HSF faces was larger than that for LSF faces, especially under the load 3 condition. Besides, BSF faces tended to be processed in a configural manner under the load 1 condition, as there was no difference between BSF and LSF faces. Conversely, they tended to be processed in a featural manner under the load 2 condition, as there was no P2 difference between BSF and HSF faces. BSF faces tended to be processed in both configural and featural manners under the load 3 condition, as there were no differences in P2 evoked by faces in any SF category. These results cannot be reconciled with the proposition from Morgan et al. (2008) that resources are limited during the retrieval stage by the increasing number of faces during the encoding stage. According to Morgan et al. (2008), attentional resources are sufficient under a low WM load, and fine processing of facial features should be adopted; under a high WM load, attentional resources are insufficient, and efficient global/configural processing should be adopted. However, we found the opposite results, and the mechanisms underlying this discrepancy need to be further investigated.

Notably, we also found a hemispheric advantage for N170 during the encoding and retrieval stages. In agreement with previous results (Rossion et al., 2003; Scott and Nelson, 2006), N170 was larger in the right hemisphere than in the left hemisphere during the encoding stage. Moreover, relative to HSF faces, the LSF faces elicited an earlier P2 under the load 3 condition during the encoding stage on the right hemisphere. This is in line with previous studies showing that configural processing occurs strongly in the right hemisphere (Scott and Nelson, 2006; Calvo and Beltran, 2014; Wang et al., 2016; Worley and Boles, 2016).

Limitation

Our results might be confined to the familiar faces as the low stimuli number is likely to make the faces familiar in this study. We analyzed N250r, which is more sensitive to familiar faces than unfamiliar faces, to test the familiarity effect. Compared with HSF faces, LSF faces elicited more negative N250r in the present. This result provides evidence that familiar faces are more sensitive to configural information, even though there is an argument against this idea (Burton et al., 2015). Additionally, our data showed that the N250r amplitude decreased as the WM load increased during the encoding (from load 1 to load 2) and retrieval stages (from load 1 to load 3). The results were consistent with previous studies (Morgan et al., 2008), which employed six male faces. Further work is required to minimize the effect of familiarity. Another factor that cannot be neglected is the cutoff value of face spatial frequency. As we mentioned above, this value varied depending on studies, which might lead to different results. Compared with previous studies, our LSF (below 6.5 cycles/image) and HSF (above 30 cycles/image) faces might include the mid-band spatial frequencies [6.73–32 cycles/image in Collin et al. (2012); 5–15 cycles/image in Hsiao et al. (2005)]. Thus, this factor should be taken into consideration when comparing different studies.

CONCLUSION

This study indicated that featural face information appears to be less important than configural face information during the working memory encoding stage from 140 to 270 ms, whereas face processing follows a coarse-to-fine manner during the retrieval stage as LSF faces (N170) are processed earlier than HSF faces (P2). Furthermore, the working memory load has an impact on the time course of face processing during the retrieval stage rather than the encoding stage. The results were consistent with face processing models that suggest facial recognition seems to be based on distinct types of processing. This study further shows the dissociation of face processing manner during the encoding and retrieval stages.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the local Institutional Review Board of the School of Psychology, Shandong Normal University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HW designed the research. HZ and EC obtained the data. AW, EC, and CB analyzed the data. AW and HW wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China [Grant Numbers 32171042 and 31700940] for HW.

ACKNOWLEDGMENTS

We would like to express their gratitude to EditSprings (https://www.editsprings.com/) for the expert linguistic services provided.

REFERENCES

Bauser, D. A. S., Mayer, K., Daum, I., and Suchan, B. (2011). Encoding/retrieval dissociation in working memory for human body forms. *Behav. Brain Res.* 220, 65–73. doi: 10.1016/j.bbr.2011.01.032

Boeschoten, M. A., Kemner, C., Kenemans, J. L., and van Engeland, H. (2005). The relationship between local and global processing and the processing of high and low spatial frequencies studied by event-related potentials and source modeling. *Cogn. Brain Res.* 24, 228–236. doi: 10.1016/j.cogbrainres.2005.01.021

Burton, A. M., Schweinberger, S. R., Jenkins, R., and Kanufmann J. M. (2015). Arguments against a configural processing account of familiar face recognition. *Psychol. Sci.* 10, 482–496. doi: 10.1177/1745691615583129

Calvo, M. G., and Beltran, D. (2014). Brain lateralization of holistic versus analytic processing of emotional facial expressions. *NeuroImage.* 92, 237–247. doi: 10.1016/j.neuroimage.2014.01.048

Carrasco, M., Loula, F., and Ho, Y. X. (2006). How attention enhances spatial resolution: evidence from selective adaptation to spatial frequency. *Percept. Psychophys.* 68, 1004–1012. doi: 10.3758/BF03193361

Cheung, O. S., and Gauthier, I. (2010). Selective interference on the holistic processing of faces in working memory. *J. Exper. Psychol.* 36, 448–461. doi: 10.1037/a0016471

Collin, C. A., Therrien, M. E., Campbell, K. B., and Hamm, J. P. (2012). Effects of band-pass spatial frequency filtering of face and object images on the amplitude of n170. *Perception.* 41, 717–732. doi: 10.1068/p7056

Cousins, R., Pettigrew, A., Ferrie, O., and Hanley, J. R. (2020). Understanding the role of configural processing in face emotion recognition in Parkinson's disease. *J. Neuropsychol.* 15, 8–26. doi: 10.1111/jnp.12210

Craddock, M., Martinovic, J., and Mueller, M. M. (2015). Early and late effects of objecthood and spatial frequency on event-related potentials and gamma band activity. *Bmc Neurosci.* 16. doi: 10.1186/s12868-015-0144-8

Curby, K. M., and Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychon. Bull. Rev.* 14, 620–628. doi: 10.3758/BF03196811

Dale, G., and Arnell, K. M. (2014). Lost in the forest, stuck in the trees: dispositional global/local bias is resistant to exposure to high and low spatial frequencies. *PLoS ONE.* 9, e98625. doi: 10.1371/journal.pone.0098625

De Cesarei, A., Mastria, S., and Codispoti, M. (2013). Early spatial frequency processing of natural images: An ERP study. *PLoS ONE.* 8, e65103. doi: 10.1371/journal.pone.0065103

Eimer, M. (2011). "*The Face-Sensitive N170 Component of the Event-Related Brain Potential*", Oxford Handbook of Face Perception. Oxford University Press, p. 329–344. doi: 10.1093/oxfordhb/9780199559053.013.0017

Evans, M. A., Shedden, J. M., Hevenor, S. J., and Hahn, M. C. (2000). The effect of variability of unattended information on global and local processing: Evidence for lateralization at early stages of processing. *Neuropsychologia.* 38, 225–239. doi: 10.1016/S0028-3932(99)00080-9

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Method.* 39, 175–191. doi: 10.3758/BF03193146

Flevaris, A., Bentin, S., and Robertson, L. (2009). Attention to hierarchical level influences attentional selection of spatial scale. *J. Vision.* 9, 224–224. doi: 10.1167/9.8.224

Freire, A., Lee, K., and Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, 29, 159–170. doi: 10.1068/p3012

Gao, Z., and Bentin, S. (2011). Coarse-to-fine encoding of spatial frequency information into visual short-term memory for faces but impartial decay. *J. Exper. Psychol.* 37, 1051–1064. doi: 10.1037/a0023091

Gao, Z., Flevaris, A. V., Robertson, L. C., and Bentin, S. (2011). Priming global and local processing of composite faces: revisiting the processing-bias effect on face perception. *Attent. Percept. Psychophys.* 73. doi: 10.3758/s13414-011-0109-7

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., and Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception.* 34, 77–86. doi: 10.1068/p5370

Goffaux, V., Jemel, B., Jacques, C., Rossion, B., and Schyns, P. G. (2003). ERP evidence for task modulations on face perceptual processing at different spatial scales. *Cognit. Sci.* 27, 313–325. doi: 10.1207/s15516709cog2702_8

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., and Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebr. Cortex.* 21, 467–476. doi: 10.1093/cercor/bhq112

Halit, H., de Haan, M., Schyns, P. G., and Johnson, M. H. (2006). Is high-spatial frequency information used in the early stages of face detection? *Brain Res.* 1117, 154–161. doi: 10.1016/j.brainres.2006.07.059

Hansen, B. C., Jacques, T., Johnson, A. P., and Ellemberg, D. (2011). From spatial frequency contrast to edge preponderance: the differential modulation of early visual evoked potentials by natural scene stimuli. *Visual Neurosci.* 28, 221–237. doi: 10.1017/S095252381100006X

Hegde, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progr. Neurobiol.* 84, 405–439. doi: 10.1016/j.pneurobio.2007.09.001

Holmes, A., Winston, J. S., and Eimer, M. (2005). The role of spatial frequency information for ERP components sensitive to faces and emotional facial expression. *Cognit. Brain Res.* 25, 508–520. doi: 10.1016/j.cogbrainres.2005.08.003

Hsiao, F. J., Hsieh, J. C., Lin, Y. Y., and Chang, Y. (2005). The effects of face spatial frequencies on cortical processing revealed by magnetoencephalography. *Neurosci. Lett.* 380, 54–59. doi: 10.1016/j.neulet.2005.01.016

Itier, R., and Taylor, M. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs. *NeuroImage.* 15, 353–372. doi: 10.1006/nimg.2001.0982

Itier, R., and Taylor, M. (2004). Face recognition memory and configural processing: a developmental ERP study using upright, inverted, and contrast-reversed faces. *J. Cognit. Neurosci.* 16, 487–502. doi: 10.1162/089892904322926818

James, M. S., Johnstone, S. J., and Hayward, W. G. (2001). Event-related potentials, configural encoding, and feature-based encoding in face recognition. *J. Psychophysiol.* 15, 275–285. doi: 10.1027//0269-8803.15.4.275

Jeantet, C., Caharel, S., Schwan, R., Lighezzolo-Alnot, J., and Laprevote, V. (2018). Factors influencing spatial frequency extraction in faces: a review. *Neurosci. Biobehav. Rev.* 93, 123–138. doi: 10.1016/j.neubiorev.2018.03.006

Jeantet, C., Laprevote, V., Schwan, R., Schwitzer, T., Maillard, L., Lighezzolo-Alnot, J., et al. (2019). Time course of spatial frequency integration in face perception: an ERP study. *Int. J. Psychophysiol.* 143, 105–115. doi: 10.1016/j.ijpsycho.2019.07.001

Kimchi, R., and Amishav, R. (2010). Faces as perceptual wholes: The interplay between component and configural properties in face processing. *Visual Cognit.* 18, 1034–1062. doi: 10.1080/13506281003619986

Laycock, R., Crewther, S. G., and Crewther, D. P. (2007). A role for the 'magnocellular advantage' in visual impairments in neurodevelopmental and psychiatric disorders. *Neurosci. Biobehav. Rev.* 31, 363–376. doi: 10.1016/j.neubiorev.2006.10.003

Leder, H., and Carbon, C. (2006). Face-specific configural processing of relational information. *Bri. J. Psychol.* 97, 19–29. doi: 10.1348/000712605X54794

Lv, J., Yan, T., Tao, L., and Zhao, L. (2015). The role of configural processing in face classification by race: an ERP study. *Front. Human Neurosci.* 9. doi: 10.3389/fnhum.2015.00679

Malinowski, P., Hubner, R., Keil, A., and Gruber, T. (2002). The influence of response competition on cerebral asymmetries for processing hierarchical stimuli revealed by ERP recordings. *Exper. Brain Res.* 144, 136–139. doi: 10.1007/s00221-002-1057-1

Mares, I., Smith, M. L., Johnson, M. H., and Senju, A. (2018). Revealing the neural time-course of direct gaze processing via spatial frequency manipulation of faces. *Biol. Psychol.* 135, 76–83. doi: 10.1016/j.biopsycho.2018.03.001

Mathes, B., and Fahle, M. (2007). The electrophysiological correlate of contour integration is similar for color and luminance mechanisms. *Psychophysiol.* 44, 305–322. doi: 10.1111/j.1469-8986.2007.00501.x

Maurer, D., Grand, R. L., and Mondloch, C. J. (2002). The many faces of configural processing. *Trends Cognit. Sci.* 6, 255–260. doi: 10.1016/S1364-6613(02)01903-4

Mercure, E., Dick, F., and Johnson, M. H. (2008). Featural and configural face processing differentially modulate ERP components. *Brain Res.* 1239, 162–170. doi: 10.1016/j.brainres.2008.07.098

Morgan, H. M., Klein, C., Boehm, S. G., Shapiro, K. L., and Linden, D. E. J. (2008). Working memory load for faces modulates P300, N170, and N250r. *J. Cognit. Neurosci.* 20, 989–1002. doi: 10.1162/jocn.2008.20072

Morrison, D. J., and Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychon. Bull. Rev.* 8, 454–469. doi: 10.3758/BF03196180

Mouraux, A., and Iannetti, G. D. (2008). Across-trial averaging of event-related EEG responses and beyond. *Magn. Reson. Imag.* 26, 1041–1054. doi: 10.1016/j.mri.2008.01.011

Nagayama, R., Yoshida, H., and Toshima, T. (1995). Interrelationship between the facial expression and familiarity: analysis using spatial filtering and inverted presentation. *Shinrigaku kenkyu: Japanese J. Psychol.* 66, 327–335. doi: 10.4992/jjpsy.66.327

Nakashima, T., Kaneko, K., Goto, Y., Abe, T., Mitsudo, T., Ogata, K., et al. (2008). Early ERP components differentially extract facial features: Evidence for spatial frequency-and-contrast detectors. *Neurosci. Res.* 62, 225–235. doi: 10.1016/j.neures.2008.08.009

Negrini, M., Brkic, D., Pizzamiglio, S., Premoli, I., and Rivolta, D. (2017). Neurophysiological correlates of featural and spacing processing for face and non-face stimuli. *Front. Psychol.* 8, 333. doi: 10.3389/fpsyg.2017.00333

Neumann, M. F., and Schweinberger, S. R. (2008). N250r and N400 ERP correlates of immediate famous face repetition are independent of perceptual load. *Brain Res.* 1239, 18–190. doi: 10.1016/j.brainres.2008.08.039

Parker, D. M., and Costen, N. P. (1999). One extreme or the other or perhaps the golden mean? Issues of spatial resolution in face processing. *Curr. Psychol.* 18, 118–127. doi: 10.1007/s12144-999-1021-3

Perfect, T. J., Weston, N. J., Dennis, I., and Snell, A. (2008). The effects of precedence on Navon-induced processing bias in face recognition. *Quar. J. Exper. Psychol.* 61, 1479–86. doi: 10.1080/17470210802034678

Peters, J. C., Goebel, R., and Goffaux, V. (2018). From coarse to fine: Interactive feature processing precedes local feature analysis in human face perception. *Biol. Psychol.* 138, 1–10. doi: 10.1016/j.biopsycho.2018.07.009

Peters, J. C., Vlamings, P., and Kemner, C. (2013). Neural processing of high and low spatial frequency information in faces changes across development: qualitative changes in face processing during adolescence. *Eur. J. Neurosci.* 37, 1448–1457. doi: 10.1111/ejn.12172

Petras, K., ten Oever, S., Jacobs, C., and Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage*. 186, 103–112. doi: 10.1016/j.neuroimage.2018.10.086

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Potts, G. F. (2004). An ERP index of task relevance evaluation of visual stimuli. *Brain Cogn.* 56, 5–13. doi: 10.1016/j.bandc.2004.03.006

Pourtois, G., Dan, E. S., Grandjean, D., Sander, D., and Vuilleumier, P. (2005). Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces: Time course and topographic evoked-potentials mapping. *Human Brain Mapp.* 26, 65–79. doi: 10.1002/hbm.20130

Ramon, M., and Rossion, B. (2012). Hemisphere-dependent holistic processing of familiar faces. *Brain Cogn.* 78, 7–13. doi: 10.1016/j.bandc.2011.10.009

Rolls, E. T., Baylis, G. C., and Leonard, C. M. (1985). Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus in the monkey. *Vision Res.* 25, 1021–1035. doi: 10.1016/0042-6989(85)90091-4

Rossion, B., Joyce, C. A., Cottrell, G. W., and Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *NeuroImage.* 20, 1609–1624. doi: 10.1016/j.neuroimage.2003.07.010

Schwaninger, A., Ryf, S., and Hofer, F. (2003). Configural information is processed differently in perception and recognition of faces. *Vision Res.* 43, 1501–1505. doi: 10.1016/S0042-6989(03)00171-8

Scott, L. S., and Nelson, C. A. (2006). Featural and configural face processing in adults and infants: A behavioral and electrophysiological investigation. *Perception.* 35, 1107–1128. doi: 10.1068/p5493

Shulman, G. L., and Wilson, J. (1987). Spatial frequency and selective attention to spatial location. *Perception.* 16, 103–111. doi: 10.1068/p160103

Towler, J., Kelly, M., and Eimer, M. (2016). The focus of spatial attention determines the number and precision of face representations in working memory. *Cerebral Cortex.* 26, 2530–2540. doi: 10.1093/cercor/bhv083

Wang, H., and Fu, S. (2018). Spatial attention modulates the temporal sequence of hemispheric asymmetry in configural and featural face processing. *Neuropsychologia.* 111, 269–275. doi: 10.1016/j.neuropsychologia.2018.01.029

Wang, H., Guo, S., and Fu, S. (2016). Double dissociation of configural and featural face processing on P1 and P2 components as a function of spatial attention. *Psychophysiology.* 53, 1165–1173. doi: 10.1111/psyp.12669

Wang, H., Qiu, R., Li, W., Li, S., and Fu, S. (2020). Cultural differences in the time course of configural and featural processing for own-race faces. *Neuroscience.* 446, 157–170. doi: 10.1016/j.neuroscience.2020.08.003

Wang, H., Sun, P., Ip, C., Zhao, X., and Fu, S. (2015). Configural and featural face processing are differently modulated by attentional resources at early stages: an event-related potential study with rapid serial visual presentation. *Brain Res.* 1602, 75–84. doi: 10.1016/j.brainres.2015.01.017

Wang, L., Mo, Y., and Qi, F. (1998). Automatic localization of the centre of iris. *J. Infr. Millim. Waves.* 17, 349–356. doi: 10.1023/A:1022531910444

Worley, M. M., and Boles, D. B. (2016). The face is the thing: faces, not emotions, are responsible for chimeric perceptual asymmetry. *Brain Cogn.* 21, 672–688. doi: 10.1080/1357650X.2015.1136319

Yang, L., and Chan, L. L. H. (2015). "An ERP study about the effects of different spatial frequencies and orientations on human brain activity", in *International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE.

Yao, Q., and Zhao, L. (2019). Using spatial frequency scales for processing own-race and other-race faces: an ERP analysis. *Neurosci. Lett.* 705, 167–171. doi: 10.1016/j.neulet.2019.04.059

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 20 May 2022
doi: 10.3389/fnins.2022.745269

Heuristic Vetoing: Top-Down Influences of the Anchoring-and-Adjustment Heuristic Can Override the Bottom-Up Information in Visual Images

Fallon Branch[1], *Erin Park*[1,2] *and Jay Hegdé*[1]*

[1] Department of Neuroscience and Regenerative Medicine, Medical College of Georgia, Augusta University, Augusta, GA, United States, [2] Department of Psychological Sciences, Augusta University, Augusta, GA, United States

When making decisions under uncertainty, human subjects do not always act as rational decision makers, but often resort to one or more mental "shortcuts", or heuristics, to arrive at a decision. How do such "top-down" processes affect real-world decisions that must take into account empirical, "bottom-up" sensory evidence? Here we use recognition of camouflaged objects by expert viewers as an exemplar case to demonstrate that the effect of heuristics can be so strong as to override the empirical evidence in favor of heuristic information, even though the latter is random. We provided the viewers a random number that we told them was the estimate of a drone reconnaissance system of the probability that the visual image they were about to see contained a camouflaged target. We then showed them the image. We found that the subjects' own estimates of the probability of the target in the image reflected the random information they were provided, and ignored the actual evidence in the image. However, when the heuristic information was not provided, the same subjects were highly successful in finding the target in the same set of images, indicating that the effect was solely attributable to the availability of heuristic information. Two additional experiments confirmed that this effect was not idiosyncratic to camouflage images, visual search task, or the subjects' prior training or expertise. Together, these results demonstrate a novel aspect of the interaction between heuristics and sensory information during real-world decision making, where the former can be strong enough to veto the latter. This 'heuristic vetoing' is distinct from the vetoing of sensory information that occurs in certain visual illusions.

Keywords: Bayesian inference, camouflage-breaking, camouflage-learning, cognitive rules of thumb, judgment and decision making, mental shortcuts, visual search

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Shenghua Luan,
Institute of Psychology (CAS), China
Sergio Da Silva,
Federal University of Santa Catarina,
Brazil

Correspondence:
Jay Hegdé
jhegde@augusta.edu

Specialty section:
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

Received: *21 July 2021*
Accepted: *28 March 2022*
Published: *20 May 2022*

Citation:
Branch F, Park E and Hegdé J
(2022) Heuristic Vetoing: Top-Down
Influences of the
Anchoring-and-Adjustment Heuristic
Can Override the Bottom-Up
Information in Visual Images.
Front. Neurosci. 16:745269.
doi: 10.3389/fnins.2022.745269

INTRODUCTION

A large body of previous research has shown that visual perception can be understood as statistical inference, whereby the brain arrives at a likely interpretation of a given visual scene by jointly evaluating the information it receives from the eyes, what it knows about the visual world, and the potential risks and rewards of a given interpretation (for reviews, see Geisler and Kersten, 2002;

Kersten et al., 2004). More generally, studies have shown that statistical (Bayesian) inference provides a useful, quantitative framework of quantitatively understanding the outcome in many sensorimotor tasks. For instance, Bayesian framework can accurately predict the outcomes even on a 'retail,' *i.e.*, trial-to-trial basis, which makes it useful for the study in many aspects of real-world decision making in which the decisions must be made on a case-by-case basis based on the information about a given case. Indeed, in many cases, the brain functions much like a perfectly rational decision maker, *i.e.,* an Ideal Observer, that combines the various aforementioned probabilistic factors in a computationally optimal fashion (Geisler and Kersten, 2002; Kersten et al., 2004; Geisler, 2011). Remarkably, it turns out that even in case of the phenomena such as visual illusions which, at first blush, might appear to violate the rules of rationality, the perceptual outcome accurately reflects the inferences of a rational decision maker, *i.e.*, that of a Bayesian Ideal Observer (Geisler and Kersten, 2002; Kersten et al., 2004; Geisler, 2011).

On the other hand, research has also shown human rationality in decision making has its limits (Tversky and Kahneman, 1974; Kahneman et al., 1982; Simon, 1982; Kahneman, 2013; Thaler, 2015). One influential line of research in bounded rationality, established by Tversky and Kahneman, has shown that human subjects often resort to 'mental shortcuts' or heuristics when making judgments and decisions under uncertainty (Tversky and Kahneman, 1974; Kahneman et al., 1982; Kahneman, 2013). The overall motivation for this study was to further elucidate these deviations from Bayesian optimality. More specifically, the present study aimed to characterize the *interaction* between the heuristic factors on the one hand and the effects of other, possibly countervailing factors on the other hand (also see below).

Extensive previous research has established that using heuristics is a natural tendency of the human mind (for overviews, see Gigerenzer and Gaissmaier, 2011; Kahneman, 2013). It is known to occur in naïve subjects as well as highly trained experts (Ericsson, 2018), and has been found in every area of human decision-making examined so far (Gigerenzer and Gaissmaier, 2011; Kahneman, 2013). While the use of heuristics does have its advantages (Kahneman, 2013; Gigerenzer, 2015), the main disadvantage is that judgments (or estimates, in statistical parlance) based on heuristics can result in systematic errors, or biases (Tversky and Kahneman, 1974).

Classical studies of heuristics have typically characterized the decision-making behavior using a paradigm where subjects are presented with vignettes of conceptual or hypothetical problem scenarios and asked to make judgments about the problem (Kahneman, 2013; Raab and Gigerenzer, 2015). For instance, in their classical study of the anchoring and adjustment (AAA) heuristic, Tversky and Kahneman asked two groups of high school students to estimate the product of the sequence of numbers from 1 to 8 within five seconds (Tversky and Kahneman, 1974). One group was presented the descending sequence ($8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$), and the other group was presented the ascending sequence ($1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$). The median estimates for the ascending and descending sequences were 512 and 2,250, respectively (the correct answer being 40,320), depending on the group. But decision-making under real-world conditions can be substantially different, in three interrelated respects: First, the decisions cannot be based on the cognitive (or 'top-down') information alone, but must take into account 'bottom-up' empirical information gleaned from the sensory faculties (Samei and Krupinski, 2010). Second, oftentimes real-world decisions must be made not in the aggregate, but on a case-by-case basis based on information specific to the problem at hand. Third, the observer's ability to glean and evaluate the sensory information can affect the decisions. However, the role of heuristics during such real-world, "retail" decision-making by experts remains unclear.

To help address this issue, we used recognition of camouflaged objects, or "camouflage-breaking," by expert observers as an exemplar case. We have previously shown when an object of interest, or target, is effectively camouflaged against its background, naïve, untrained observers cannot recognize the camouflaged target (or "break" its camouflage) (Chen and Hegdé, 2012a,b). However, subjects can be trained in the laboratory to become expert camouflage-breakers (Chen and Hegdé, 2012a). Thus, camouflage-breaking is an excellent model system for studying real-world, retail decision-making by experts. We therefore examined the effects of the AAA heuristic on camouflage-breaking. As described below, we used a straightforward modification of the classical AAA paradigm (Tversky and Kahneman, 1974) to characterize the effects of AAA on visual search for a camouflaged target in a camouflage scene. For this reason, we also present and discuss our results using AAA as the primary framework of understanding.

EXPERIMENT 1: CHARACTERIZATION OF THE EFFECT OF THE ANCHORING AND ADJUSTMENT HEURISTIC ON CAMOUFLAGE-BREAKING IN VISUAL SCENES

Materials and Methods

Subjects

All procedures used in this study were duly reviewed and approved in advance by the Institutional Review Board (IRB) of Augusta University in Augusta, GA, where this study was carried out. All subjects were adult volunteers who had normal or corrected-to-normal vision, and provided written informed consent prior to participating in the study.

Prior to their participation in these experiments, we used our previously described deep-training method (Chen and Hegdé, 2012a) to train the subjects to break camouflage using the same background texture (*e.g.,* foliage, see **Figure 1**) as the texture they would encounter during the present study (see Chen and Hegdé, 2012a for details). All the subjects who participated in this study had an asymptotic camouflage-breaking performance of $d' >= 1.95$ ($p < 0.05$) for the background texture that they were to encounter during this study (Chen and Hegdé, 2012a).

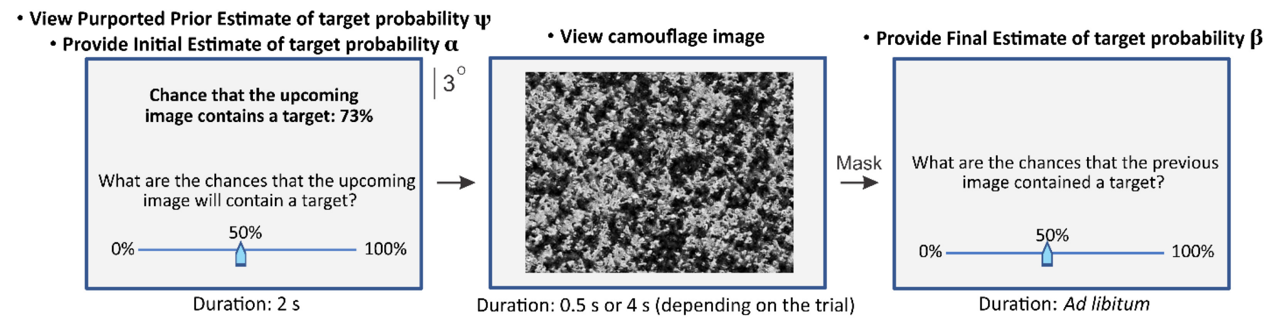


FIGURE 1 | Task paradigm of Experiment 1. The three panels from left to right in this figure are shown in the temporal order they were presented during each trial. The starting position of the on-screen slider (**left** and **right** panels, **bottom**) was always 50%. Panels not drawn to exact scale. See text for details.

Six subjects trained to asymptotic levels participated in Experiment 1.

We digitally synthesized the camouflaged visual scenes used in this study *de novo* as we have previously described (Chen and Hegdé, 2012a). Briefly, each scene consisted of a textured background with or without a single foreground object of interest, *i.e.*, the search target. We created background textures that captured key statistical properties of real-world textures using the texture synthesis algorithm of Portilla and Simoncelli, 1999). For instance, to create the background texture type we named "foliage", we used a real-world photograph of foliage as input, and synthesized a large number of images that captured the key statistical properties of the input texture (see, *e.g.*, **Figure 1**, *center*), so that the output images had the same statistical properties, but were pixelwise non-identical to each other. To create a camouflaged scene with a target for this experiment, we digitally textured a 3-D model of a human face using one of the output images, and composited it, without shadows or occlusion, against a different output image. An equal number of additional output images served as scenes without the target, so that the stimulus during each given trial had a 50% chance of containing a target (see Chen and Hegdé, 2012a for details).

Procedure

Prior to the actual data collection, subjects received detailed, illustrated instructions about the trial procedures. Subjects were encouraged to carry out practice trials before starting the actual trials to familiarize themselves with the procedure. The data from the practice trials were discarded.

Experiment 1 consisted of four conditions. During conditions in which explicit anchoring information was externally provided (conditions 1 and 2, **Table 1**; also see below), each trial began when the subject indicated readiness by pressing a key on the computer's keyboard, upon which the subject was shown, for 2s, an on-screen message stating the percent chance (which ranged between 0 and 100%, depending on the trial) that the image they were about see contained the search target, *i.e.*, a single camouflaged face (**Figure 1**, *left panel, top*). For convenience, we will refer to this estimate as "purported prior estimate ψ" or, equivalently, "anchoring information". The subjects were told that this probability was determined by a drone system that reconnoitered the scene for this target. But in actuality, these were pseudorandom numbers generated *de novo* by a random number generator during each trial (also see below).

Subjects were then given *ad libitum* time to provide an initial estimate of their perceived probability that the upcoming image contained the search target ("subject's initial estimate α") using an on-screen slider (**Figure 1**, *left panel, bottom*). A previously unseen camouflaged scene was then presented for 0.5 s or 4 s, depending on the trial (**Figure 1**, *middle panel*), followed by a 0.5 s random-dot mask (not shown). After this, subjects were given *ad libitum* time to estimate the probability that the scene they just viewed contained a target ("subject's final estimate β"; **Figure 1**, *right panel*).

The conditions in which no explicit anchoring information was provided (conditions 3 and 4; see **Table 1**), were identical to conditions 1 and 2, respectively, except that the purported prior estimate was blank ("–").

Each trial block consisted of eight trials (four conditions × two stimulus durations) presented in a randomly interleaved fashion. Each subject performed at least four blocks of trials over one or more days.

Rationale for using random numbers for purported prior probabilities ψ. As noted above, an overall goal of the present study was to characterize the effect of the subjects' anchoring information ψ on their probability estimates. This meant, on the one hand, that we needed to manipulate ψ. On the other hand, we had to ensure that ψ conveyed no systematic information about the target status of the stimulus, so as to prevent confounding effects. Using random ψ values was a principled way of meeting both of these requirements.

TABLE 1 | Experimental conditions in Experiment 1.

Condition #	Anchoring Information	Target status of the image
1	Provided	Target absent
2	Provided	Target present
3	Not provided	Target absent
4	Not provided	Target present

It is important to note that our IRB has determined that our use of random numbers does not amount to deception under the applicable regulations and policies.

Data Analysis

Data were analyzed using scripts custom-written for R[1] and Matlab[2] platforms. Area under the ROC curve (AUC) was calculated using the default options in the AUC function of the R library *DescTools* (Signorell et al., 2020).

Post hoc Power Analyses

These analyses were carried out using the R library *pwr*. Before initiating the present study, we carried out *a priori* power analyses to determine the subject recruitment target. To do this, we used the empirically observed fit of the data from a pilot study (Branch et al., 2022) as the expected effect size, and calculated the total number of trials (pooled across all subjects). The results indicated that at least 47 trials (pooled across all subjects and repetitions) would be needed to achieve a statistical power of 0.90. *A posteriori* power analyses using the actual data indicated that our data achieved a power of > 0.95 for the regression analyses in each of the three experiments.

RESULTS AND DISCUSSION

Effect of the Anchoring and Adjustment Heuristic on Camouflage-Breaking in Visual Scenes: Experiment 1

Prior to participating in this experiment, subjects were trained to criterion in the camouflage-breaking task (mean $d' = 2.08$; median = 1.96; SEM = 0.13) as described in Materials and Methods. The background texture used in this experiment was synthesized from using real-world pictures of natural foliage. The target, when present, was a human head, and was also textured using a different image of the same texture type (*i.e.,* "foliage"). The camouflage images used in this experiment were a random subset of the same large superset of $>10^4$ images from which the images used in the training of the subjects were also drawn. That is, the subjects were tested in this experiment using the same type of target and background texture that were used during their prior training.

Trials Without Anchoring Information

Our task paradigm required the subjects to provide an initial estimate α of the chances that the camouflage image they had not seen yet (but were about to see) contained a target. For convenience, we will refer to this starting estimate of the subjects as their anchored position. When the purported prior estimate ψ was not provided to the subjects during a given trial, the subjects had no explicit information on which to base their initial estimates. For convenience, we will refer to these trials as those in which anchoring information was unavailable or trials without anchoring information.

As expected, when the anchoring information was unavailable, the subjects tended to estimate the target probability at around 50% on average before they viewed the image (*Subjects' Initial Estimates*α; *x-axis* in **Figure 2A**). After viewing the image, the subjects' final estimates β of target probability were broadly distributed (*y-axis* in **Figure 2A**), indicating that viewing the image substantially altered their estimates of target probability.

Classical studies have shown that in AAA based on vignettes, subjects start with an initial judgment "anchored" based on the anchoring information, and arrive at their final estimate by adjusting their estimate until they are satisfied with it (Tversky and Kahneman, 1974). The biases, or errors, in these judgments arise from the fact that the subjects' final judgments tend to be influenced by their initial judgments.

To determine if this also occurs in the absence of anchoring information, we plotted the size of adjustment δ_i during a given trial *i* (*i.e.*, the amount by which the subjects adjusted their final estimate β_i relative to their initial estimate α_i during a given trial *i*; $\delta_i = \beta_i - \alpha_i$) as a function of their initial estimate α_i during that trial (**Figure 2B**). The two quantities were significantly anticorrelated ($r = -0.57$, $df = 142$, $p < 0.05$) indicating that, in this case, the anchored position did contribute to the final estimate even in the absence of the anchoring information. That is, adjustment from an anchored position can occur even in the absence of explicit anchoring information akin to that provided in the classical studies of Tversky and Kahneman (1974). Thus, the anchoring process is dissociable from anchoring information *per se*.

Subjects Break Camouflage Accurately When the Anchoring Information Is Unavailable

The fact that the AAA effect did occur (albeit on a much smaller scale) when the anchoring information was unavailable raises an important issue: The subjects had to come up with their initial estimates α before they had seen the image for that trial. They provided their final estimates β after they had viewed the stimulus. The fact that β values were significantly correlated with the corresponding α values straightforwardly means that the initial values influenced the subjects' final estimates. The net effect, if any, of such image-irrelevant factors, by definition, is to degrade camouflage-breaking performance. Were the expert subjects able to overcome the biasing influence of their own initial estimates enough to accurately detect camouflaged targets in the images?

To help answer this question, we carried out a receiver operating characteristic (ROC) analysis of the subjects' final responses. The resulting ROC curve is shown in **Figure 2C** (*solid blue line*). The diagonal represents random performance. In this case, the area under the curve (AUC) is 0.5. The actual AUC was significantly above random levels (AUC = 0.92; randomization test, $p < 0.05$, *i.e.*, 0 out of 1,000 rounds of randomization). Thus, even though the subjects' initial positions α did have a biasing effect on their final estimates, the subjects successfully overcame this effect in their final estimates and detected the camouflaged target highly accurately.

---

[1] r-project.org

[2] Mathworks.com

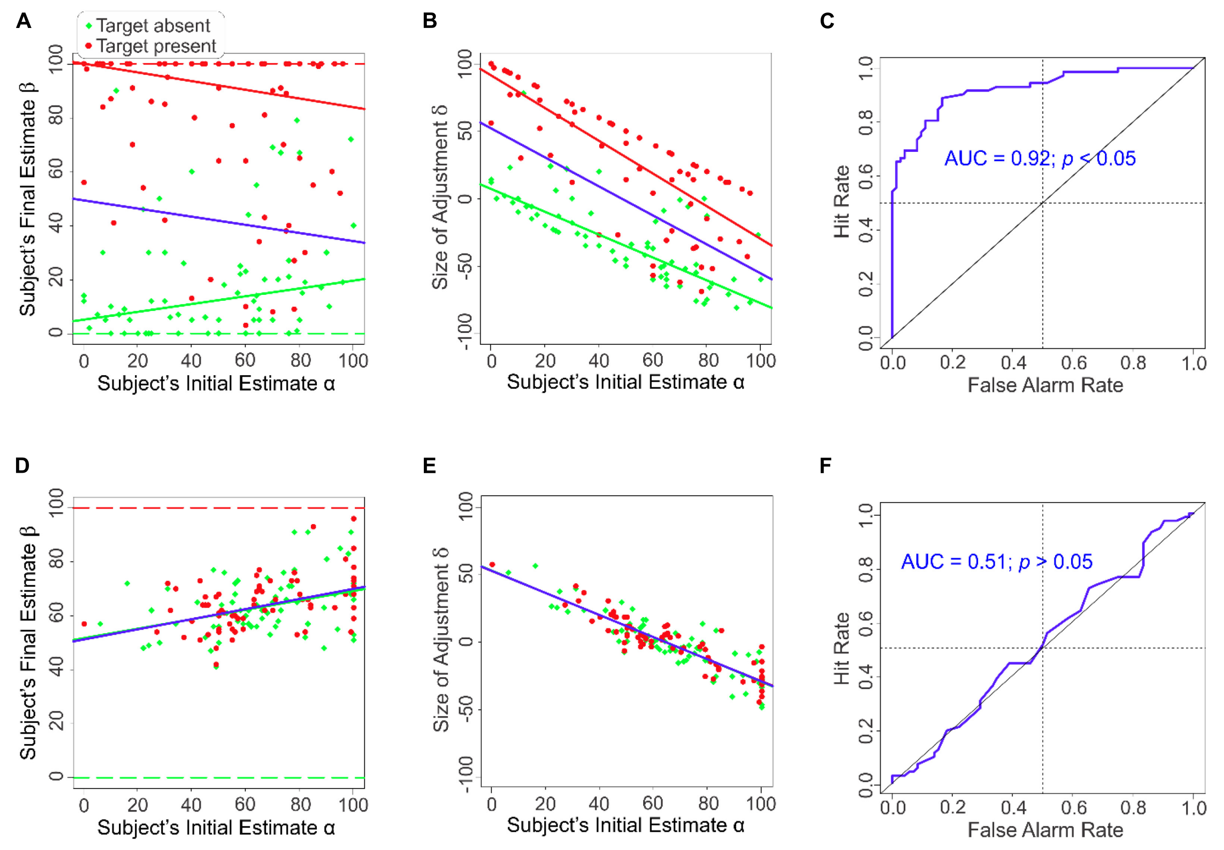


FIGURE 2 | Task performance with or without anchoring information in Experiment 1. Panels **(A–C)** results when the external anchoring information was not provided (*i.e.*, Conditions 3 and 4). Panels **(D,E)** results when the external anchoring information was provided (*i.e.*, Conditions 1 and 2). **(A)** Subjects' final estimates as a function of their initial estimates in the absence of anchoring information. **(B)** The magnitude of the subjects' adjustment δ as a function of their initial estimate α in the absence of anchoring information. **(C)** ROC analysis of the subjects' final estimates in the absence of anchoring information. **(D)** The subjects' final estimates as a function of their initial estimates in the presence of anchoring information. **(E)** The magnitude of the subjects' adjustment δ as a function of their initial estimate α the presence of anchoring information. **(F)** ROC analysis of the subjects' final estimates in the presence of anchoring information. Regression lines that best account for the data are shown in a color-coded fashion in panels **(A,B,D,E)** (*red*, target present; *green*, target absent; *blue*, all data points). Note that in panels d and e, the *blue* line largely overlaps, and therefore obscures, the *red* and the *green* lines. The *dashed lines* in panels **(A,D)** denote the expected responses (*red*, target present; *green*, target absent).

TABLE 2A | Contribution of the various explanatory variables to the size of adjustment d when anchoring information was unavailable in Experiment 1 (Conditions 3 and 4).

Row #	Explanatory variable	Estimated coefficient	Standard error	*t* value	*p* value
1	Subjects' initial estimate α	−0.04	0.08	−0.52	0.60
2	Target status θ (target present *vs*. target absent)	29.32	2.19	13.40	<0.001
3	Reaction time *r*	−0.0003	0.0004	0.69	0.49

TABLE 2B | Contribution of the various explanatory variables to the size of adjustment d when anchoring information was available in Experiment 1 (Conditions 1 and 2).

Row #	Explanatory variable	Estimated coefficient	Standard error	*t* value	*p* value
1	Subjects' initial estimate α	0.19	0.04	5.43	<0.001
2	Target status θ (target present *vs*. target absent)	0.20	0.79	0.25	0.80
3	Reaction time *r*	0.0002	0.0003	0.80	0.43

To help determine the contributions of various underlying factors to the final estimates γ, we carried out a regression analysis (see "Materials and Methods" section). When the anchoring information was unavailable (**Table 2A**), the target status θ was a highly significant contributor to the final estimates γ (row 2). Indeed, no other explanatory variable accounted for a significant proportion of the final estimates (rows 1 and 3).

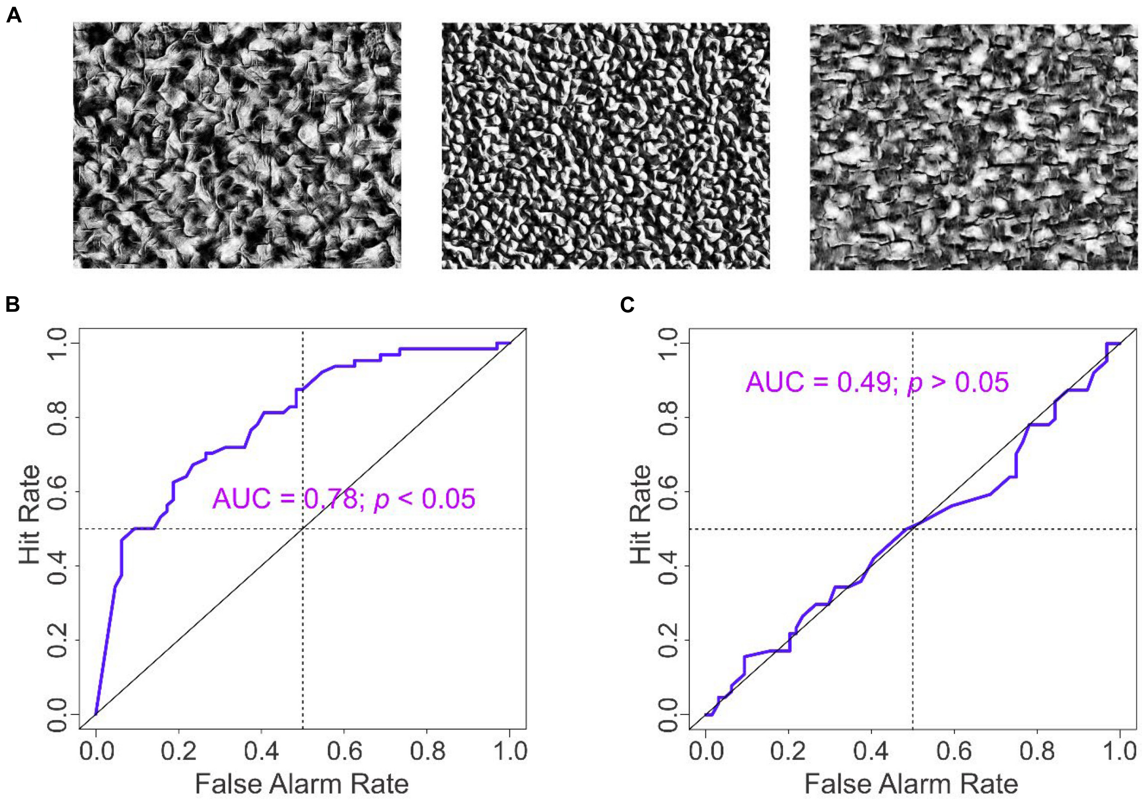


FIGURE 3 | Results of Experiment 2. **(A)** exemplar stimuli used in Experiment 2. **(B,C)** ROC analysis of the subjects' final estimates in the absence and presence of anchoring information, respectively.

In the Presence of Anchoring Information, the Subjects' Camouflage-Breaking Performance Is at Random Levels

When the anchoring information, *i.e.*, the purported prior estimates ψ, were available, the subjects' initial estimates α were highly correlated with prior estimates (correlation coefficient $r = 0.95$, $df = 142$, $p < 0.05$; not shown), indicating that the purported prior estimate did succeed in producing a strong anchoring effect as expected. That is, subjects were strongly influenced by this 'top-down' information and tended to anchor their own initial estimates on this information. Recall that the purported prior estimates ψ were random.

The subjects were then shown, in a randomized order, the same set of images as those shown when the anchoring information was unavailable. Thus, the differences in outcome between the two pairs of conditions, if any, were not attributable to the images *per se*.

Note that, after viewing the image, the subjects were required to estimate the chance that the image they had just viewed contained a target, and that the sole relevant source of information for estimating this quantity was the image itself. If the subjects solely relied on the image information, their final estimates β would conform to the ground truth about the given image (*red* and *green dashed lines* in **Figure 2D**). However, the subjects' actual final estimates of the target status of images substantially varied from the ground truth, regardless of whether the images were positive or negative for the target (*red* and *green symbols* in **Figure 2D**).

To help characterize the relationship of the magnitude of adjustment δ to the anchored position in the presence of anchoring information, we plotted the size of adjustment δ_i during each given trial i as a function of their initial estimate α_i during that trial (**Figure 2E**). We found that δ was highly anticorrelated with α, regardless of the target status θ of the image ($r = -0.89$, $df = 142$, $p < 0.05$; **Figure 2E**). This straightforwardly suggests that the reason why the final estimates were *uncorrelated* with the target status θ of the image (**Figure 2D**) was that the subjects arrived at their final estimates β by adjusting from their anchored positions α (**Figure 2E**), which themselves were highly correlated with the random ψ values ($r = 0.53$, $df = 142$, $p < 0.05$; not shown).

Post hoc modeling of the subjects' final estimates confirmed that the actual target status of the image indeed played an insignificant role in the subjects' final estimates of the target (**Table 2B**, row 2). Indeed, the only predictor that significantly accounted for the final estimates were the subjects' initial estimates α (row 1). Receiver operating characteristic (ROC) analysis indicated that subjects' performance was indistinguishable from random (**Figure 2F**). Note that this effect is not attributable to the subjects' intrinsic inability to break camouflage to begin with, because when the anchoring information was unavailable, the same subjects broke camouflage highly accurately using the same set of images.

The result that the subjects performed at random levels is consistent with the fact that the anchoring information ψ that their decisions were based on was itself random. This result is nonetheless surprising, because it suggests that trained subjects can altogether ignore task-relevant empirical information in camouflage scenes when they have access to anchoring information. One plausible explanation for this is that the subjects were under time pressure so that they were unable to scrutinize the images sufficiently well. Previous studies have shown that time pressure can induce subjects to resort to using heuristics (Kahneman et al., 1982; Kahneman, 2013). However, our *post hoc* analyses indicated that the stimulus duration did not significantly contribute to the outcome, regardless of the target status (row 3, **Tables 2A,B**). Moreover, subjects often took less than the allotted time before responding (data not shown; also see Experiment 2 below).

EXPERIMENT 2: DOES THE EFFECT OF ANCHORING AND ADJUSTMENT GENERALIZE TO OTHER EXPERIMENTAL CONDITIONS?

Materials and Methods

Subjects

Four subjects trained to asymptotic levels participated in Experiment 2.

Procedure

This experiment was identical to Experiment 1, except in the following three respects. First, three new background textures ("fruit," "nuts," and "mushrooms"; see **Figure 3A**; also see **Table 3**) were used as background textures, and counter-rotated across trials, blocks, and subjects. Second, novel, naturalistic 3-D objects, called "digital embryos" that the subjects had not seen before were used as targets in 50% of randomly interleaved trials, also on a counter-rotating basis (not shown). Third, the subjects were allowed to view the stimuli for an unlimited duration and were allowed to end the stimulus presentation and proceed to the next phase of the trial by pressing a designated button (not shown).

RESULTS AND DISCUSSION

Anchoring and Adjustment Effects Are Reproducible Across Disparate Experimental Conditions

To determine whether and to what extent the AAA effect generalizes across to other experimental parameters, we carried out Experiment 2, in which we systematically varied the background texture and the search targets (see "Materials and Methods" section for details; also see **Figure 3A**).

TABLE 3 | Experimental conditions in Experiment 2.

Condition #	Anchoring Information	Target status of the image
1	Provided	Target absent
2	Provided	Target present
3	Not provided	Target absent
4	Not provided	Target present

We found that all of the key results of Experiment 1 were reproducible in this experiment as well (**Figures 3B,C**). For instance, when the purported prior estimates ψ were available, the magnitude of adjustment δ was strongly anticorrelated with α regardless of the target status θ of the image when the anchoring information was available ($r = -0.79$, $df = 126$, $p < 0.05$; not shown). When the prior information was unavailable, the anticorrelation between δ and α was weaker, albeit still statistically significant ($r = -0.44$, $df = 126$, $p < 0.05$; not shown). Finally, the subjects' camouflage-breaking performance was highly accurate when anchoring information was unavailable (AUC = 0.78, $p < 0.05$), but was at random levels when anchoring information was available (AUC = 0.49, $p > 0.05$). The results of the regression analyses for this experiment (**Tables 4A,B**) were qualitatively similar to those from Experiment 1. Thus, the results of Experiment 1 were essentially reproducible in Experiment 2.

EXPERIMENT 3: VISUAL PATTERN DETECTION PERFORMANCE OF NAÏVE, NON-PROFESSIONAL SUBJECTS WITH *VS.* WITHOUT ANCHORING INFORMATION

Materials and Methods

Subjects

Eleven naïve, non-professional subjects (as opposed to trained camouflage-breakers used in Experiments 1 and 2) participated in Experiment 3.

Procedure

This experiment was identical to Experiments 1 and 2, except where specified otherwise. The subjects performed a target detection task as in Experiments 1 and 2, except that the target in this experiment was a Gabor patch (8 cycles/degree, $\sigma = 1°$) embedded in dynamic random dot noise (Kersten, 1984) (dot density, dot size = 1 pixel^2; 50% ON, 50% OFF; refresh rate = 60 Hz; see **Figure 4**). Prior to the experiment, subjects received detailed instructions and viewed exemplar images with or without Gabor patches (clearly discernible when present), so that subjects knew what to look for. Collectively, these procedures helped ensure that no prior training or visual pattern recognition expertise was needed in order for the subjects to perform the task (see **Table 5**). To help add stimulus uncertainty, the spatial

TABLE 4A | Contribution of the various explanatory variables to the final estimates γ when anchoring information was available in Experiment 2 (Conditions 3 and 4): *Post hoc* general linear modeling (GLM) of the contributions of the various explanatory variables to the response variable (*i.e.*, final estimates γ of subjects).

Row #	Explanatory variable	Estimated coefficient	Standard error	*t* value	*p* value
1	Subjects' initial estimate α	0.04	0.15	0.26	0.80
2	Target status θ (target present *vs*. target absent)	9.39	1.78	5.27	<0.001
3	Reaction time *r*	0.0005	0.007	0.07	0.94

TABLE 4B | Contribution of the various explanatory variables to the final estimates γ when anchoring information was available in Experiment 2 (Conditions 1 and 2): *Post hoc* general linear modeling (GLM) of the contributions of the various explanatory variables to the response variable (*i.e.*, final estimates γ of subjects).

Row #	Explanatory variable	Estimated coefficient	Standard error	*t* value	*p* value
1	Subjects' initial estimate α	0.04	0.07	0.55	0.58
2	Target status θ (target present *vs*. target absent)	−0.11	1.07	−0.11	0.92
3	Reaction time *r*	−0.01	0.006	−1.70	0.09

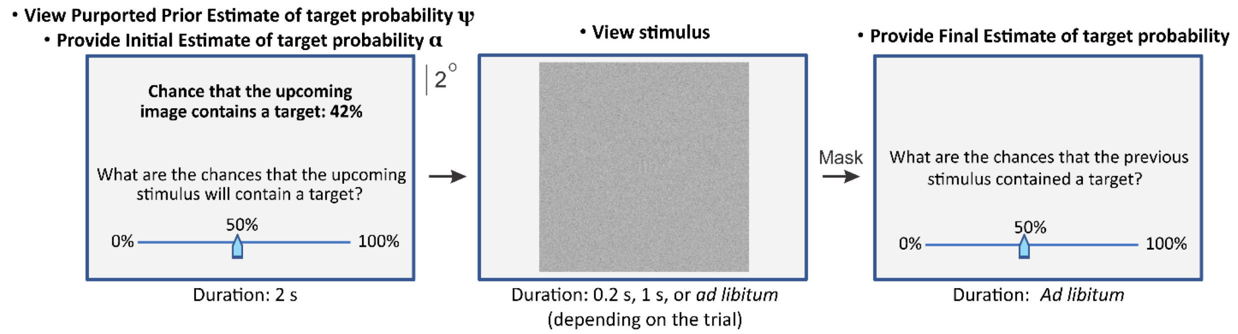


FIGURE 4 | Task paradigm of Experiment 3. In this experiment, the visual stimulus was a dynamic random dot stimulus (dRDS), one static frame of which is shown in this figure (**middle panel**). In 50% of the randomly interleaved trials, the dRDS contained in Gabor patch at the subject's contrast threshold (Kersten, 1984). See text for details.

TABLE 5 | Experimental conditions in Experiment 3.

Condition #	Anchoring Information	Target status of the stimulus
1	Provided	Gabor patch absent
2	Provided	Gabor patch present
3	Not provided	Gabor patch absent
4	Not provided	Gabor patch present

location and orientation of the Gabor patch (when present) were randomly jittered from one trial to the next.

We customized the contrast of the Gabor patch for each subject, so as to help ensure that the stimulus was sufficiently ambiguous and to help minimize the variations in task performance related to task difficulty across subjects. We carried out a preliminary experiment to determine the contrast threshold for each subject. To do this, we presented the Gabor patch (with the same parameters as above), one per trial at systematically varying contrasts. Subjects viewed the stimulus *ad libitum,* followed by a random dot mask, and used an on-screen slider to report the probability that the stimulus contained the Gabor patch target. We fitted a logistic contrast response function (Harvey, 1997) to the data (**Supplementary Figure 1A**). We took the point of inflection of the fitted function, at which the slope of the function was maximal, as the contrast threshold for the given subject (Campbell and Green, 1965). The distribution of contrast thresholds for all subjects is shown in **Supplementary Figure 1B**.

For each subject, the Gabor patch target in Experiment 3 was presented at their contrast threshold. The subject performed the target detection as in Experiments 1 and 2, except that the target was the Gabor patch, instead of a camouflaged target.

RESULTS AND DISCUSSION

Anchoring and Adjustment Effects Are Reproducible in Naïve, Untrained Subjects Performing a Simple Detection Task

To determine if this overriding effect of AAA is specific to experts such as highly trained camouflage-breakers, we tested naïve, non-professional subjects using a variation of the above task that required neither training nor expertise in pattern recognition (Experiment 3; see "Materials and Methods" section for details). This experiment was identical to Experiments 1 and 2, except that the subjects were required to report whether a dynamic random dot stimulus contained a Gabor patch presented at the subject's empirically determined contrast threshold (see **Figure 4**; also see **Supplementary Figure 1**). The subjects were told that the prior information provided to them was the probability that the image they were about to see did contain the Gabor target, as determined by a previous viewer.

The results of this experiment (**Figure 5**) were qualitatively similar to those of Experiments 1 and 2 (**Figures 2**,**3**, respectively). Moreover, each individual subject in Experiment 3 detected the target accurately in the absence of the anchoring information, but performed at chance levels in the presence of anchoring information (**Figure 6**). Thus, the ability of the AAA heuristic to override the empirical information generalized across stimuli, tasks, and the subject's training/expertise in pattern recognition.

Two additional aspects of Experiments 1-3 are worth noting and are clearest from the results of Experiment 3. First, the subjects' use of the AAA heuristic is not attributable to time pressure *per se*, because the subjects performed highly accurately under otherwise identical conditions when anchoring information was not available (**Figures 3**, **5**). Second, the anchoring effects in this experiment were not attributable to the *requirement* to report the initial estimate *per se*, because the subjects were required to make this report regardless of whether anchoring information was present (**Tables 6A,B**). When the anchoring information ψ was available, the amount of adjustment δ was highly anticorrelated with the initial values α ($r = -0.89$; $df = 838$; $p < 0.05$; **Figure 5A**), and was not significantly influenced by the presence of the Gabor patch θ (1-way ANCOVA; α: $F(1,836) = 3088.47$, $p < 2.0 \times 10^{-16}$; θ: $F(2,836) = 0.973$, $p = 0.32$). When the anchoring information was unavailable, the anticorrelation was more modest, albeit still significant ($r = -0.30$; $df = 838$; $p < 0.05$; **Figure 5C**), arguably because the subjects took into account the presence of the Gabor patch θ when the anchoring information α was unavailable (1-way ANCOVA; α: $F(1,836) = 118.13$, $p < 2 \times 10^{-16}$; θ: $F(2,836) = 351.99$, $p < 2 \times 10^{-16}$). Thus, the anchoring process itself is dissociable from the anchoring information it is based on, in that the former can occur without the latter.

GENERAL DISCUSSION

A New Principle of Top-Down *vs.* Bottom-Up Interaction: Anchoring and Adjustment Heuristic Can 'Veto' Visual Information

We show that, in each of the three experiments, the subjects fail to detect the target when anchoring information is available. But when anchoring information is unavailable, the same subjects detect the target highly accurately using the same set of images. This straightforwardly implies that the anchoring information causes the subjects to ignore the image information in favor of the anchoring information when the latter is available. That is, the heuristic information can override or veto the image information in visual pattern recognition tasks.

Our results demonstrate that there are certain conditions, such as the availability of strong anchoring information in the present case, under which heuristic decision-making is the default mode, and not the strategy of last resort, of decision-making under uncertainty. This is because when both sets of information were available, the subjects' decisions were dominated by the heuristic information. This finding is particularly important, because the resulting errors were large enough to reduce the subjects' camouflage-breaking performance to chance levels.

Another notable aspect of our results also show that the biasing effects of AAA, previously demonstrated in the aggregate for subject groups evaluating verbal vignettes (Tversky and Kahneman, 1974; Kahneman et al., 1982; Thaler, 1993; Rieskamp and Hoffrage, 2008), persist in 'retail', case-by-case decision-making. Case-by-case decision scenarios are common in the real world, so that the heuristic influences demonstrated by our study are likely to be prevalent under real-world conditions.

Our results also show that the anchoring can occur, albeit to a lesser extent, in the absence of externally provided anchoring information. That it is, even when no anchoring information is externally provided, the subjects' final estimates

TABLE 6A | Contribution of the various explanatory variables to the final estimates γ when anchoring information was available in Experiment 3 (Conditions 1 and 2): *Post hoc* general linear modeling (GLM) of the contributions of the various explanatory variables to the response variable (*i.e.*, final estimates γ of the subjects).

Row #	Explanatory variable	Estimated coefficient	Standard error	*t* value	*p* value
1	Subjects' initial estimate α	0.48	0.11	4.40	<0.001
2	Target status θ (target present *vs*. target absent)	−0.69	0.75	−0.92	0.36
3	Reaction time *r*	3.01×10^{-3}	2.47×10^{-3}	1.22	0.22

TABLE 6B | Contribution of the various explanatory variables to the final estimates γ when anchoring information was available in Experiment 3 (Conditions 3 and 4): *Post hoc* general linear modeling (GLM) of the contributions of the various explanatory variables to the response variable (*i.e.*, final estimates γ of the subjects).

Row #	Explanatory variable	Estimated coefficient	Standard error	*t* value	*p* value
1	Subjects' initial estimate α	−1.10	0.10	−10.65	<0.001
2	Target status θ (target present *vs*. target absent)	38.98	2.08	18.72	<0.001
3	Reaction time *r*	5.67×10^{-3}	6.62×10^{-3}	0.86	0.39

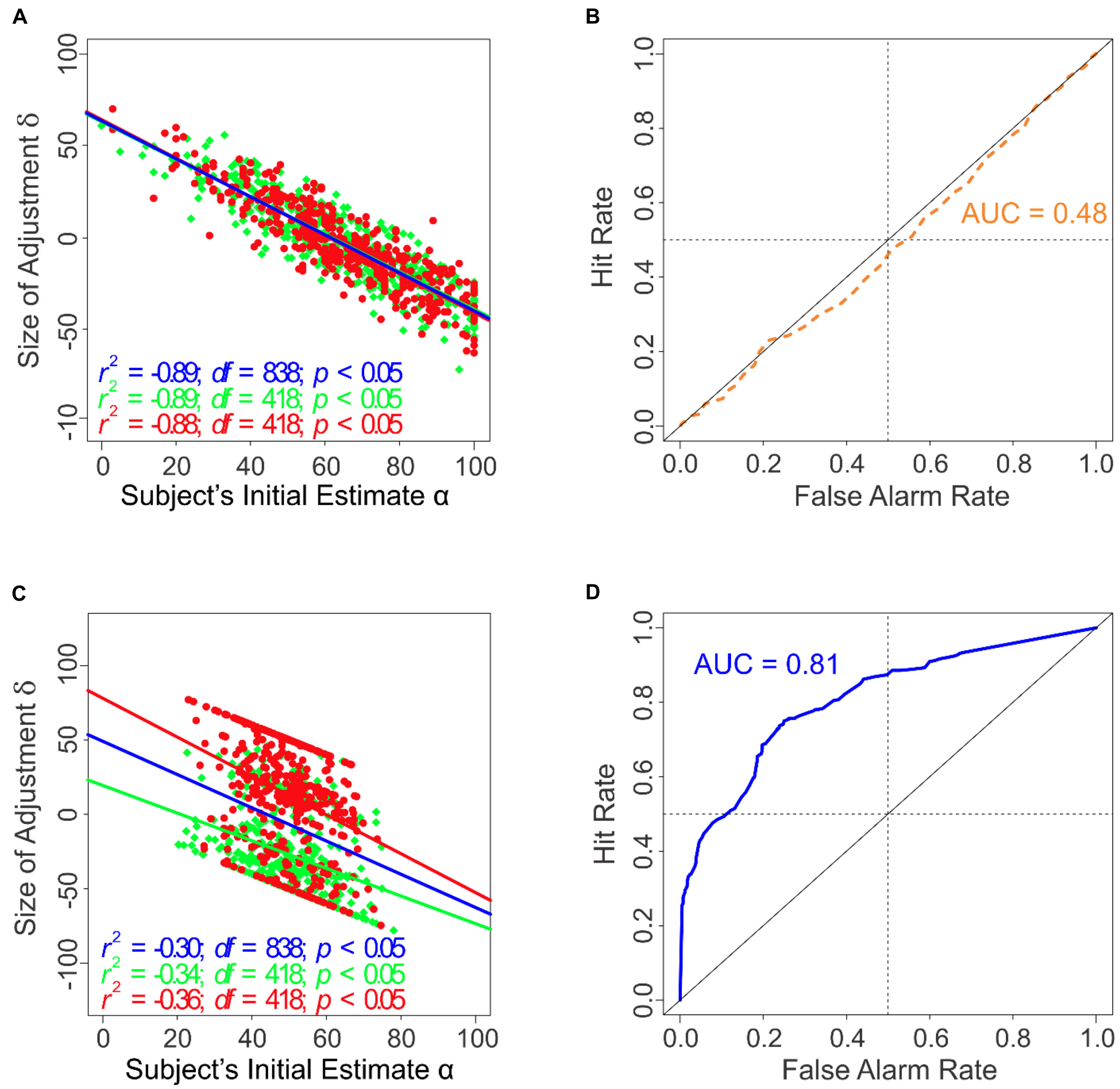


FIGURE 5 | Task performance of subjects with or without anchoring information in Experiment 3. The various panels in this figure are drawn using the same plotting conventions as the corresponding panels in earlier figures. **(A)** The magnitude of the subjects' adjustment δ as a function of their initial estimate α in the absence of anchoring information. Note that the *blue* regression line in this panel largely overlaps, and therefore obscures, the *red* and the *green* regression lines. **(B)** ROC analysis of the subjects' final estimates in the presence of anchoring information. **(C)** The magnitude of the subjects' adjustment δ as a function of their initial estimate α in the absence of anchoring information. **(D)** ROC analysis of the subjects' final estimates in the absence of anchoring information.

are anticorrelated, albeit modestly, with their initial estimates, suggesting that the subjects start from an anchored position even when not induced to do so by externally provided information (see **Figures 2B**, **5C**). It is plausible that the process of providing the initial estimates itself had the implicit effect of anchoring the subjects' initial judgments. In any event, this internal anchoring was not strong enough to significantly affect the subjects' performance (see **Figures 2C**, **5D**). More significantly, this effect demonstrates that the anchoring process is dissociable from the anchoring information *per se*. This is important, because this suggests that requiring subjects to make an initial decision can affect their final decision in any task.

Our results raise the possibility that the AAA heuristic can, in principle, affect any task involving visual search. This has serious implications for real-world tasks involving visual search, such as airport baggage screening and medical image perception. Indeed, we have recently found a similar AAA 'veto' effect in practicing radiologists examining mammograms (Branch et al., 2022).

Why Disbelieve Your Own Eyes?

A striking aspect of our results is the fact that subjects effectively disbelieve their own eyes in favor of what they hear from an external source, such as a drone or a previous viewer. In all three

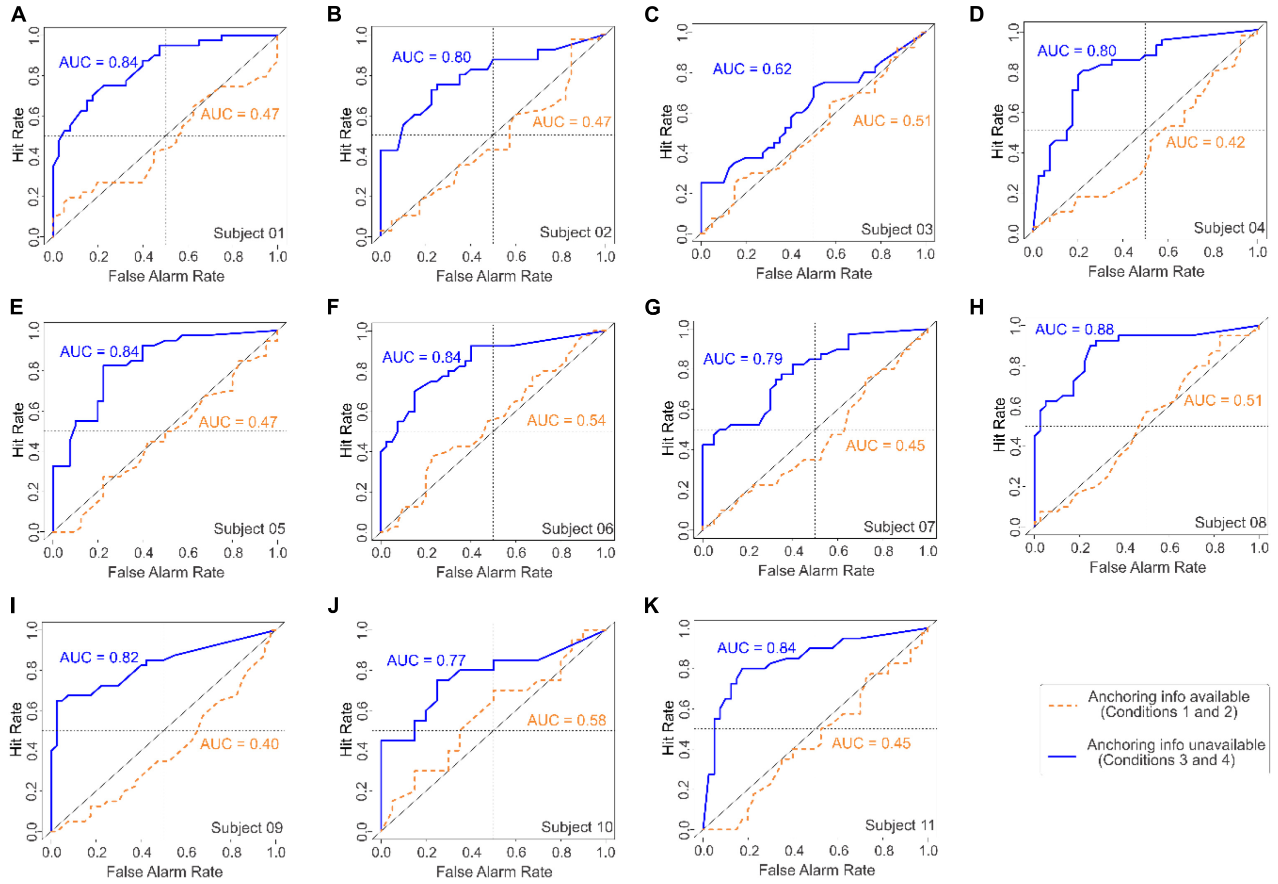


FIGURE 6 | ROC analyses of the responses of each of the 11 individual subjects in Experiment 3 (panels **A-K**). In each panel, the ROC curves for Gabor detection performances with or without anchoring information (dashed *brown* and *solid blue curves*, respectively) are shown, as are the corresponding AUC values (*brown* and *blue* type, respectively). In each panel, the *diagonal* represents chance performance (AUC = 0.5). See text for details.

experiments, subjects accurately detected the target in the absence of prior information, indicating that the subjects were able to detect the target to begin with, but when the prior information was available, they essentially ignored what they saw in favor what they were told.

The veto effect is all the more striking in the cases of Experiments 1 and 2, where the subjects were expert camouflage-breakers. We have previously reported that expert camouflage-breakers are so skilled in their task that they can detect the camouflaged target even after brief viewing the stimulus, even as briefly as 50 ms, which does not permit extended scrutiny or eye movements (Chen and Hegdé, 2012a; Branch et al., 2021). In this specific sense, detecting the target is relatively easy for the expert subjects, so that the subjects could easily cross-check the prior information against the visual evidence. It is therefore surprising that the subjects – judging by the results – fail to, or choose not to, do such cross-checking. A detailed examination of the cognitive costs of such cross-checking, including the costs imposed by task difficulty, are needed to help clarify the reasons behind this surprising effect.

To be sure, what is surprising here is that the heuristic effect can be so strong, and not that expert camouflage-breakers resort to heuristic decision-making in the first place. After all, heuristic decision-making is notoriously resistant to expertise training; experts in every profession examined to date are known to resort to heuristic decision-making (Gigerenzer and Gaissmaier, 2011; Kahneman, 2013; Ericsson, 2018). But previous studies have neither systematically examined the interaction between the heuristic information versus the sensory evidence. Our study examined this effect and found the veto effect.

Still, why does the veto occur at all? Why do subjects ignore the physical evidence in the images? While our study did not examine this important question for practical reasons, one plausible explanation is that the veto itself is, at least in part, a reflection of the so-called authority bias or halo effect, whereby experts and laypeople alike abide by what they consider expert opinions (Milgram, 1963; Stasiuk et al., 2016; Zaleskiewicz and Gasiorowska, 2021). This may also explain, at least in part, why the subjects apparently do not begin to disregard the prior information even upon a relatively large number of trials in which the prior information does not jive with the empirical evidence before the subjects' very eyes. The present study did not examine this important issue for practical reasons, in part because it would require, among other things, a detailed quantification of both the perceived reliability of the prior information during a given trial, and the updating of the perceived reliability from one trial to the next. Further studies are needed to examine these important issues in detail.

Possible Limitations of Heuristic Vetoing and Other Caveats

It is important to emphasize that what our results demonstrate is that under certain conditions, e.g., when the heuristic information is strong and the bottom-up information is ambiguous or otherwise weak, the heuristic information *can* override the visual information. This is not to say, however, that heuristic information always *does* override visual information. The uncertainty of the visual information in our experiments was arguably high enough, *i.e*., the sensory information was weak enough, that the strong top-down information was able to override it.

It is intuitively obvious, on the other hand, that there exist conditions where the opposite is true, *i.e*., the bottom-up information overrides the top-down information. For instance, if the visual targets in our experiments were easily detectable, *e.g.*, if the Weber contrast of the Gabor patches in Experiment 3 were 1.0 and that of the background were 0.0, subjects would readily ignore the prior information and go with the image information instead. For practical reasons, the present study did not examine this possibility. Further studies are needed to empirically establish this possibility.

It is also intuitively obvious that under most real-world conditions, the strength of the stimulus information would be somewhere between the aforementioned two extremes. While the vetoing effect would be obscured in such cases, the underlying heuristic-visual interaction is unlikely to disappear altogether. Instead, the behavioral outcomes under these conditions are likely to reflect a complex interplay of the two influences, when both are present.

Heuristic-Visual Interaction Is Distinct From Visual Illusions

It is instructive to compare and contrast heuristic vetoing with certain visual illusions. For instance, in the hollow face illusion or the Ames room illusion, the brain's built-in assumptions about the relevant visual objects override the visual information (Geisler and Kersten, 2002; Hartung et al., 2005; Kroliczak et al., 2006; Parpart et al., 2018). These visual illusions are analogous to the heuristic vetoing, in two main respects. First, in both cases, image information is overshadowed by top-down factors. Second, both represent special cases, where the image information is ambiguous, usually in highly specific ways. For example, the Ames room has to be constructed in specific ways to facilitate the brain's tendency to assume the room is symmetrical. In the case of heuristic vetoing, the visual target presumably must be difficult enough to find for the vetoing effect to show through. Thus, visual illusions are special cases just as heuristic vetoing is.

On the other hand, heuristic vetoing is distinctly different, in the sense that it is clearly not built-in, but externally induced. In the present case, for instance, the anchoring effect is induced by the anchoring information provided to the subject. The built-in assumptions in the aforementioned visual illusions are typically so strong that it is not possible generally to volitionally alter these influences.

Concluding Remarks: Heuristic Vetoing in Perspective

Given the aforementioned fact that heuristic vetoing is self-evidently a rather special case in the vein of visual illusions, one reasonable perspective about our study is that it is a proof-of-principle study that reveals that heuristics can, in principle, veto the visual evidence. Also, given the fact that heuristics are ubiquitous in human judgments, what is ultimately surprising about our results is not that they reveal a heuristic effect, but that they reveal a veto effect.

DATA AVAILABILITY STATEMENT

The data supporting the conclusions of this article will be made available by the authors upon reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board (IRB) of Augusta University, Augusta, GA, United States. The participants gave written informed consent prior to participating in the study.

AUTHOR CONTRIBUTIONS

FB, EP, and JH designed the experiment, analyzed the data and prepared the manuscript. FB and EP collected the data. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by grant #W911NF-15-1-0311 from the Army Research Office (ARO) to JH.

ACKNOWLEDGMENTS

We thank our colleagues, especially Alan Saul and Eugene Bart, for helpful discussions and for comments on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.745269/full#supplementary-material

REFERENCES

Branch, F., Lewis, A. J., Santana, I. N., and Hegde, J. (2021). Expert camouflage-breakers can accurately localize search targets. *Cogn. Res. Princ. Implic.* 6:27. doi: 10.1186/s41235-021-00290-5

Branch, F., Santana, I. N., and Hegdé, J. (2022). Biasing influence of 'Mental Shortcuts' on diagnostic decision-making: radiologists can overlook breast cancer in mammograms when prior diagnostic information is available. *Diagnostics* 12:105. doi: 10.3390/diagnostics12010105

Campbell, F. W., and Green, D. G. (1965). Optical and retinal factors affecting visual resolution. *J. Physiol.* 181, 576–593. doi: 10.1113/jphysiol.1965.sp007784

Chen, X., and Hegdé, J. (2012a). Learning to break camouflage by learning the background. *Psychol. Sci.* 23, 1395–1403. doi: 10.1177/0956797612445315

Chen, X., and Hegdé, J. (2012b). Neural mechanisms of camouflage-breaking: a human FMRI study. *J. Vision* 12:582. doi: 10.1167/12.9.582

Ericsson, K. A. (2018). *The Cambridge Handbook of Expertise and Expert Performance*. New York, NY: Cambridge University Press.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Res.* 51, 771–781. doi: 10.1016/j.visres.2010.09.027

Geisler, W. S., and Kersten, D. (2002). Illusions, perception and Bayes. *Nat. Neurosci.* 5, 508–510. doi: 10.1038/nn0602-508

Gigerenzer, G. (2015). *Simply Rational: Decision Making in the Real World*. Oxford: Oxford University Press.

Gigerenzer, G., and Gaissmaier, W. (2011). Heuristic decision making. *Annu. Rev. Psychol.* 62, 451–482.

Hartung, B., Schrater, P. R., Bulthoff, H. H., Kersten, D., and Franz, V. H. (2005). Is prior knowledge of object geometry used in visually guided reaching? *J. Vis.* 5, 504–514. doi: 10.1167/5.6.2

Harvey, L. O. Jr. (1997). Efficient estimation of sensory thresholds with ML-PEST. *Spat. Vis.* 11, 121–128. doi: 10.1163/156856897x00159

Kahneman, D. (2013). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgment Under Uncertainty : Heuristics and Biases*. Cambridge: Cambridge University Press.

Kersten, D. (1984). Spatial summation in visual noise. *Vision Res.* 24, 1977–1990. doi: 10.1016/0042-6989(84)90033-6

Kersten, D., Mamassian, P., and Yuille, A. (2004). Object perception as bayesian inference. *Annu. Rev. Psychol.* 55, 271–304. doi: 10.1146/annurev.psych.55.090902.142005

Kroliczak, G., Heard, P., Goodale, M. A., and Gregory, R. L. (2006). Dissociation of perception and action unmasked by the hollow-face illusion. *Brain Res.* 1080, 9–16. doi: 10.1016/j.brainres.2005.01.107

Milgram, S. (1963). Behavioral study of obedience. *J. Abnorm. Psychol.* 67, 371–378. doi: 10.1037/h0040525

Parpart, P., Jones, M., and Love, B. C. (2018). Heuristics as Bayesian inference under extreme priors. *Cogn. Psychol.* 102, 127–144. doi: 10.1016/j.cogpsych.2017.11.006

Portilla, J. S., and Simoncelli, E. P. (1999). "Texture modeling and synthesis using joint statistics of complex wavelet coefficients," in *Proceedings of the IEEE Workshop on Statistical and Computational Theories of Vision* (Fort Collins, CO).

Raab, M., and Gigerenzer, G. (2015). The power of simplicity: a fast-and-frugal heuristics approach to performance science. *Front. Psychol.* 6:1672. doi: 10.3389/fpsyg.2015.01672

Rieskamp, J., and Hoffrage, U. (2008). Inferences under time pressure: how opportunity costs affect strategy selection. *Acta Psychol. (Amst)* 127, 258–276. doi: 10.1016/j.actpsy.2007.05.004

Samei, E., and Krupinski, E. (2010). *The Handbook of Medical Image Perception and Techniques.* Cambridge: Cambridge University Press.

Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., Arachchige, C., et al. (2020). *DescTools: Tools for Descriptive Statistics. R Package Version* 0.99, 32 Edn. Vienna: The R Foundation for Statistical Computing.

Simon, H. A. (1982). *Models of Bounded Rationality*. Cambridge: MIT Press.

Stasiuk, K., Bar-Tal, Y., and Maksymiuk, R. (2016). The effect of physicians' treatment recommendations on their epistemic authority: the medical expertise bias. *J. Health Commun.* 21, 92–99. doi: 10.1080/10810730.2015.1049308

Thaler, R. H. (1993). *Advances in Behavioral Finance*. New York, NY: Russell Sage Foundation.

Thaler, R. H. (2015). *Misbehaving : the Making of Behavioral Economics*. New York, NY: W.W. Norton & Company.

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124

Zaleskiewicz, T., and Gasiorowska, A. (2021). Evaluating experts may serve psychological needs: Self-esteem, bias blind spot, and processing fluency explain confirmation effect in assessing financial advisors' authority. *J. Exp. Psychol. Appl.* 27, 27–45. doi: 10.1037/xap0000308

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 24 May 2022
doi: 10.3389/fpsyg.2022.784188

Spatial Memory and Blindness: The Role of Visual Loss on the Exploration and Memorization of Spatialized Sounds

Walter Setti[1], Luigi F. Cuturi[1], Elena Cocchi[2] and Monica Gori[1]*

[1] *Unit for Visually Impaired People (U-VIP), Italian Institute of Technology, Genoa, Italy,* [2] *Istituto David Chiossone, Genoa, Italy*

Spatial memory relies on encoding, storing, and retrieval of knowledge about objects' positions in their surrounding environment. Blind people have to rely on sensory modalities other than vision to memorize items that are spatially displaced, however, to date, very little is known about the influence of early visual deprivation on a person's ability to remember and process sound locations. To fill this gap, we tested sighted and congenitally blind adults and adolescents in an audio-spatial memory task inspired by the classical card game "Memory." In this research, subjects (blind, $n = 12$; sighted, $n = 12$) had to find pairs among sounds (i.e., animal calls) displaced on an audio-tactile device composed of loudspeakers covered by tactile sensors. To accomplish this task, participants had to remember the spatialized sounds' position and develop a proper mental spatial representation of their locations. The test was divided into two experimental conditions of increasing difficulty dependent on the number of sounds to be remembered (8 vs. 24). Results showed that sighted participants outperformed blind participants in both conditions. Findings were discussed considering the crucial role of visual experience in properly manipulating auditory spatial representations, particularly in relation to the ability to explore complex acoustic configurations.

Keywords: audio-spatial skills, blindness, development, working memory, user-friendly technologies, acoustic perception

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Antonio Prieto,
National University of Distance
Education (UNED), Spain
Nora Turoman,
Université de Genève, Switzerland

****Correspondence:***
Walter Setti
walter.setti@iit.it

Specialty section:
This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: *27 September 2021*
Accepted: *21 April 2022*
Published: *24 May 2022*

Citation:
Setti W, Cuturi LF, Cocchi E and Gori M (2022) Spatial Memory and Blindness: The Role of Visual Loss on the Exploration and Memorization of Spatialized Sounds. Front. Psychol. 13:784188. doi: 10.3389/fpsyg.2022.784188

HIGHLIGHTS

- A novel task, the Audio-Memory, presented in the form of a game to evaluate audio-spatial memory abilities in sighted and blind individuals.
- Sighted outperformed the blind participants.
- Blind people encounter limitations ascribed to congenital blindness in processing auditory spatial representations and exploring complex acoustic configurations.

INTRODUCTION

In everyday life, abilities such as comprehension, reasoning, or learning are achieved through memory processes that allow the human brain to retain spatial and non-spatial information. The cognitive system devoted to the temporary storage and manipulation of information is the working memory system (WM) (Palmer, 2000). Historically, the most supported model of WM was proposed by Baddeley (1992), who divided WM into three separate subsystems: the central executive component (involved in high-order cognitive functions), the phonological loop and the visuo-spatial sketchpad (VSSP) that are used for the storage and processing of verbal and visuo-spatial information, respectively. Logie (1995) proposed an additional division of the VSSP into two subcomponents: the "inner scribe," which refers to spatial components of information, and the "visual cache" for processing visual features of objects. One of the main functions ascribed to the VSSP of WM is mental imagery, a cognitive function that leads to internal representations (Cornoldi and Vecchi, 2003) of the objects composing the surrounding space. This function corresponds to a quasi-perceptual experience occurring in the absence of actual stimuli for the relevant perception (Kosslyn, 1980; Finke and Freyd, 1989; Rinck and Denis, 2004). Mental imagery is directly involved in cognitive functions such as learning (Yates, 1966), problem-solving, reasoning (Féry, 2003) and original and creative thought (LeBoutillier and Marks, 2003). The nature of these representations has long been the subject of research and debate. Kosslyn's theory (Kosslyn, 1980), the most supported in this context, posits that mental images are "picture-like" representations, as confirmed by studies involving mental rotation and mental scanning of haptic spatial layouts (Farah et al., 1988; Vingerhoets et al., 2002). Although Kosslyn's initial theory assumed that imagery processes partially overlap with perceptual mechanisms, evidence has shown that imagery cannot be equated with visual perception (Cornoldi and Vecchi, 2003). Visual mental images are not mere copies of visual input but rather the end product of a series of constructive processes based on memory retrieval mechanisms (Pietrini et al., 2004). Therefore, visuo-spatial mental imagery can originate from different sensory and perceptual inputs (e.g., visual, haptic, acoustic and verbal) (Cornoldi and Vecchi, 2003). Supporting this view, neuroimaging and electrophysiological studies generally indicate that the maintenance of information in spatial WM is not modality–specific and does not strictly depend on the encoding sensory modality (Lehnert and Zimmer, 2006, 2008).

Studies with congenitally blind individuals can provide fundamental insights into the role of vision in spatial memory abilities within the imagery debate. Visually impaired individuals can generate and manipulate mental images through long-term memory, haptic exploration, or verbal description (Zimler and Keenan, 1983; Lederman and Klatzky, 1990; Carreiras and Codina, 1992). Visual features such as dimension, shape, or texture can be perceived through touch and conveyed in internal images. Thus, the absence of sight does not impede an efficient visuospatial system functioning.

Blind individuals show deficits in memory tasks requiring large sequences of mental manipulation of stored information, namely active memory tasks (Vecchi et al., 1995, 2004; Vecchi, 1998). When the experimental demand requires only maintaining small amounts of information instead (i.e., passive memory tasks), their abilities usually do not differ significantly from sighted people (Cornoldi and Vecchi, 2003; Setti et al., 2018, 2019). Nevertheless, blind individuals might also show limitations when only passive memory processes are involved, such as memorizing 2D spatial layouts (Vecchi, 1998). In fact, vision remains the preferred sensory modality that facilitates the accomplishment of visuospatial working memory tasks, especially when great demands on memory are required.

Blind individuals do show limitations when asked to continuously process the mental image of a previously learned spatial layout (Juurmaa and Lehtinen-Railo, 1994) or when performance can be enhanced through active manipulation of spatial information (Setti et al., 2018). Moreover, blind individuals encounter difficulties using perspective in mental representations (Arditi and Dacorogna, 1988) and in elaborating the third dimension when learning a haptic spatial layout (Vecchi, 1998). When increasing the number of items to be actively processed, thus increasing task demand, blind individuals demonstrate inferior performance compared to sighted individuals (Vecchi, 1998; Vanlierde and Wanet-Defalque, 2004; Cattaneo et al., 2008). Vision is the best sensory modality through which the brain processes several items simultaneously (De Beni and Cornoldi, 1988), and as such, lack of vision impacts this ability (Cornoldi and Vecchi, 2003).

There is evidence that vision plays a crucial role in guiding the maturation of spatial cognition (Hart and Moore, 2017). In early visual deprivation, the remaining intact sensory modalities are recruited to process spatial information. In tactile tasks such as object recognition and immediate hand-pointing localization, visually impaired individuals perform as well or better than sighted controls (Morrongiello et al., 1994; Rossetti et al., 1996; Sunanto and Nakata, 1998). In the auditory processing of space, blind individuals exhibit enhanced abilities for azimuthal localization (King and Parsons, 1999; Roder et al., 1999; Gougoux et al., 2004; Doucet et al., 2005) and relative distance discrimination (Voss et al., 2004; Kolarik et al., 2013). At the same time, blind people show significant impairments for auditory spatial tasks such as vertical localization, absolute distance discrimination and spatial bisection (Zwiers et al., 2001; Lewald, 2002; Gori et al., 2014). Thus, early visual deprivation affects performance in tasks requiring a complex representation of space and it has been argued that these deficits reflect a lack of visual calibration over touch and audition in processing spatial information (Gori et al., 2014). According to the cross-sensory calibration hypothesis, vision calibrates the other sensory modalities to process spatial information. In other words, the brain learns from vision how to evaluate objects' orientation and proprioceptive position through alternate sensory modalities such as audition and touch (Cappagli et al., 2017; Cuturi et al., 2017). Another explanation for the decreased performance of visually impaired individuals in complex spatial tasks is that it originates from a compromised spatial memory. This hypothesis is supported by studies demonstrating that blind children have limitations in spatial recall (Millar, 1976) and the simultaneous processing of multiple representations (Puspitawati et al., 2014). These results do not suggest that mnemonic skills in general are impaired in blind individuals. Blind individuals ably perform temporal tasks that require participants to understand and remember the temporal order of sound presentations (Vercillo et al., 2016). They show limitations only in tasks requiring complex spatial judgments (Bertonati et al., 2020) where the spatial presentation of stimuli position is fundamental to accomplish the task (Gori et al., 2014).

Lack of visual experience may also lead to the differential use of spatial reference frames to encode the information to be memorized. The two main frames of reference used to represent the location of entities in space are egocentric and allocentric (Ruggiero and Iachini, 2010). The first defines locations of items in the surrounding environment from the observer's perspective and in relation to observer's position. Conversely, allocentric reference frames encode spatial information by considering external landmarks and spatial relationships among the items regardless of observer's position. In the context of spatial memory, previous research has demonstrated that spatial information is organized according to reference frames defined by the layout itself and not by egocentric experience (Mou and McNamara, 2002). Depending on the task to be accomplished, sighted individuals can rely on allocentric frames of reference to orient themselves or to represent and memorize spatial information (Ruggiero et al., 2012; Pasqualotto et al., 2013; Iachini et al., 2014). In contrast, early visual deprivation results in significant impairments in tasks that require an allocentric representation of space (Thinus-Blanc and Gaunet, 1997; Arnold et al., 2013; Gori et al., 2014).

Most research investigating spatial memory in vision loss has been carried out in the haptic domain (Vecchi et al., 1995; Bonino et al., 2008, 2015). Although haptic information plays a substantial role in processing objects proximal to the observer, auditory information allows visually impaired individuals to process surrounding information, including items that are not directly reachable by the observer. For instance, in spatial navigation, sensory substitution devices can aid the ability to build mental maps by integrating auditory and self-motion information (Jicol et al., 2020). In the study, visually impaired individuals could take advantage of visual information converted to acoustic cues more efficiently than sighted participants when performing both egocentric and allocentric navigation tasks, thus indicating that multisensory cueing of space may reduce blindness-related deficits. Conversely, in the context of spatial memory, Setti et al. (2018) demonstrated that congenitally blind individuals show limitations when asked to manipulate spatial information in recalling sequences of spatialized sounds in the acoustic sensory domain.

To deeply investigate audio-spatial memory and exploration strategies in blindness, we focus on comparing how blind and sighted individuals construct and manipulate a dynamic auditory structure in a spatial memory task. In the context of this study, the term "dynamic" refers to a spatial structure whose configuration needs to be continuously updated. This aspect reflects everyday life experiences where surrounding acoustic information constantly changes and provides a fundamental sensory cue for blind people to represent the surrounding environment. We tested ability to hold spatialized sounds in memory and update the mental representation of their locations. With this goal in mind, inspiration was taken from the card game "Memory," which works on attention, memory and concentration (da Cunha et al., 2016). By playing this card game, it is possible to improve concentration, train short-term memory, strengthen associations between concepts, and classify objects grouped by similar traits. In its original form, the game consists of covered cards lying on a table, and the goal is to find pairs among the cards. We adapted the "Memory" game to the auditory domain by employing a vertical array of speakers named ARENA2D (see **Figure 1**). We call this novel task "Audio-Memory." Participants were required to match sounds that were spatially displaced over the audio-tactile device. To investigate the impact of memory load on performance, we designed two experimental conditions by increasing the number of sounds paired, using four pairs in the 4-pair condition and 12 pairs in the 12-pair condition (named 4-pair and 12-pair conditions, respectively, for brevity). With the Audio-Memory task, we addressed the following research questions:

1. To what extent does early visual deprivation influence audio-spatial memory skills?
2. What is the exploration strategy used by the two groups when asked to explore a complex auditory structure to construct a spatial representation of sound dispositions?

Since vision is crucial for spatial processing (Alais and Burr, 2004; Burr et al., 2009; Hart and Moore, 2017) we hypothesized that a lack of visual experience would affect audio-spatial memory skills in blind individuals. Due to the great cognitive load required to manipulate spatial information and the difficulties in using the spatial relations among the sounds, we expected sighted to outperform blind participants. Furthermore, in the context of spatial exploration, we hypothesized that congenitally blind participants would explore the layout of speakers differently compared to the sighted group. Considering how the lack of visual experience affects spatial processing in blind individuals, we expected them to show a slower and more sequential exploration of the spatially distributed items compared to sighted participants.

RESULTS

We tested blind and sighted adolescents and adults. All groups performed two experimental tasks, consisting of four and twelve pairs of sounds to be matched (**Figure 2**). To deliver sounds, we used ARENA2D, a 5-by-5 matrix of speakers covered by tactile sensors, each constituting a haptic block (Ahmad et al., 2019; Setti et al., 2019; **Figure 1**).

Participants sat facing ARENA2D at a distance of 30 cm. Memory performance was evaluated using the Score reached at the end of the test and the Number of attempts to pair the two stimuli once positions were discovered. Finally, we defined the Audio-Anchor index as an expression of spatial exploration strategy that measured how often participants started two consecutive attempts by touching the same haptic block (see "Data Analysis" for further details). Statistical analyses were conducted using RStudio (Version 1.1.463) and data are shown as means and standard error (**Figures 3**, **4**, **5**). Given the small sample size, statistical analyses were conducted with non-parametric tests based on permutations. We first ran a non-parametric MANCOVA (*adonis2()* R function) with Score,

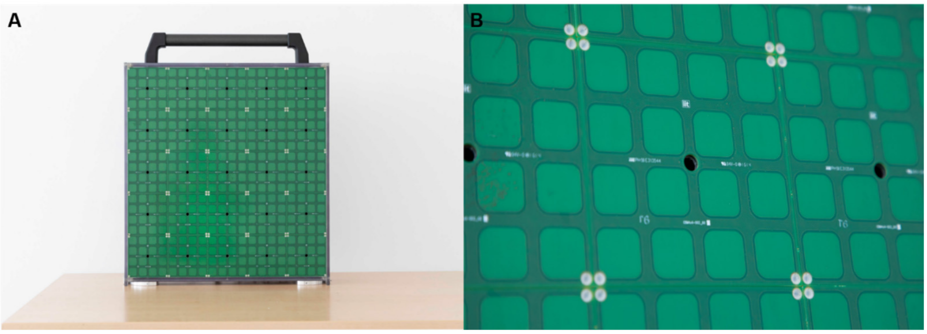


FIGURE 1 | ARENA2D at two levels of detail. **(A)** Presents the device. ARENA2D is a vertical surface (50 × 50 cm) composed of 25 haptic blocks, each with a loudspeaker in the center, arranged in the form of a matrix. **(B)** Shows a single haptic block in detail. The black hole is the speaker from which the sound is emitted. The blocks are covered by 16 (4 × 4 matrix) tactile sensors (2 × 2 cm^2) that register the position of each touch.

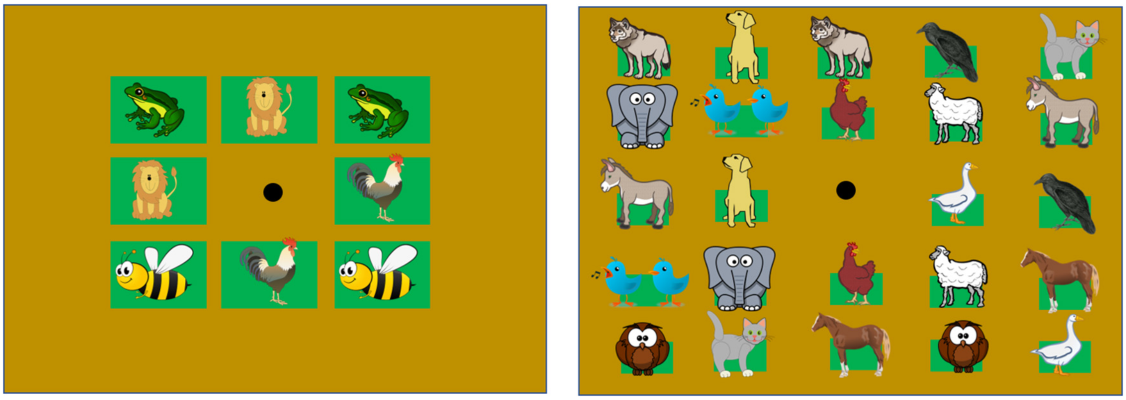


FIGURE 2 | Grids used in experimental conditions. The two grids differ in the size of the apertures for each auditory item. The apertures on the grids represented in the left column are 10 cm × 10 cm, equal to the haptic block size. The apertures on the grids represented in the right panel are 4 cm × 4 cm. Depicted animals placed inside the squares, refer to the position of the animal calls in each grid (images downloaded from a royalty-free website, https://publicdomainvectors.org/). The black dot at the center indicates the speaker emitting feedback sounds.

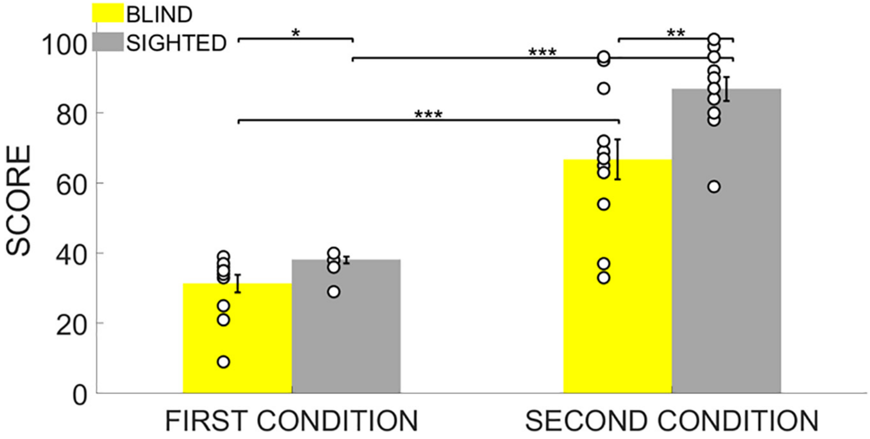


FIGURE 3 | Score. Data are presented as mean and standard error for each group. The white circles on the bars represent the individual data. The Score reached by the participants was lower in the 4-pair condition and the sighted outperformed the blind group in both experimental conditions. * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$.

Number of attempts and Audio-Anchor as dependent variables, *Group* (either blind or sighted) as between-subject, *Difficulty* (either easy or hard) as within-subject and *Age* (the age of the participants in decimal number of years) as a covariate. Follow-up analyses were conducted only for the significant interactions with ANOVAs based on permutations (*aovp()* R function). Finally, *post hoc* analyses were run with unpaired Student's *t*-tests based on permutations (*perm.t.test()* R function) and *Bonferroni* corrections were used to correct for multiple comparisons (see the Statistical Analyses section for further details).

Results of the non-parametric MANCOVA (number of permutations = 999) highlighted a significant main effect of the *Group* [$F(1,40) = 9.124$, $p = 0.001$, $\eta^2 = 0.243$], a significant main effect of *Difficulty* [$F(1,40) = 72.62$, $p = 0.001$, $\eta^2 = 0.909$] and a significant interaction *Group* * *Difficulty* [$F(1,40) = 3.35$, $p = 0.032$, $\eta^2 = 0.202$] but no significant main effect of *Age* [$F(1,40) = 0.44$, $p = 0.626$], nor significant interactions *Group** *Age* [$F(1,40) = 0.915$, $p = 0.367$], *Age* * *Difficulty* [$F(1,40) = 0.07$ $p = 0.939$] nor *Group* * *Age* * *Difficulty* [$F(1,40) = 1.87$, $p = 0.154$]. Thus, given that the only significant interaction was *Group* * *Difficulty* we ran a follow-up non-parametric ANOVA, separately for the Score the Number of attempts and the Audio-Anchor, with *Group* as between-subject and *Difficulty* as within-subject factors and by putting together adolescents and adults in both groups. The results of MANCOVA indeed highlighted that the performance in the task were not affected by participant ages.

The ANOVA of the Score, showed a significant main effect of the *Group* (iterations = 5,000, $p < 0.001$, $\eta^2 = 0.24$), a significant main effect of the *Difficulty* (iterations = 5,000, $p < 0.001$, $\eta^2 = 0.76$) and a significant interaction *Group* * *Difficulty* (iterations = 2,131, $p = 0.041$, $\eta^2 = 0.07$). As expected, *post hoc* analyses first revealed that both blind and sighted participants reached a higher score in the first condition because it was easier in terms of number of stimuli to be paired (unpaired two-tailed *t*-test based on permutations: Welch's $t = 5.66$, $p < 0.001$, Cohen's $d = 1.01$, Welch's $t = 2.49$, $p < 0.001$, Cohen's $d = 5.67$ for both the blind and sighted groups, respectively). Interestingly, we found that, compared to blind participants, sighted participants reached a higher Score in both the 4-pair (unpaired two-tailed *t*-test based on permutations: Welch's $t = 2.50$, $p = 0.031$, Cohen's $d = 1.02$) and 12-pair condition (unpaired two-tailed *t*-test based on permutations: Welch's $t = 3.02$, $p = 0.01$, Cohen's $d = 1.23$) compared to the blind group (**Figure 3**).

The results of the ANOVA on the Number of attempts also highlighted a significant main effect of the *Group* (iterations = 5,000, $p < 0.001$, $\eta^2 = 0.21$), a significant main effect of the *Difficulty* (iterations = 5,000, $p < 0.001$, $\eta^2 = 0.53$) and a significant interaction *Group* * *Difficulty* (iterations = 2,329, $p = 0.04$, $\eta^2 = 0.07$). As expected, also in this case, *post hoc* analyses first revealed that both blind and sighted participants needed more attempts to end the task in the second condition due to the greater number of sounds to be paired (unpaired two-tailed *t*-test based on permutations: Welch's $t = 5.35$, $p < 0.001$, Cohen's $d = 2.18$, Welch's $t = 4.79$, $p < 0.001$, Cohen's $d = 1.96$ for both the blind and sighted groups, respectively).

Post hoc analyses revealed no significant difference between the two groups in the Number of Attempts required to end the task in the first condition (unpaired two-tailed *t*-test based on permutations: Welch's $t = 1.98$, $p = 0.21$). Nevertheless, the blind group needed more attempts to end the task in the second condition compared to the sighted group (unpaired two-tailed *t*-test based on permutations: Welch's $t = 2.815$, $p = 0.035$, Cohen's $d = 1.15$). Results for Number of Attempts confirmed that blind participants did not hold item locations in memory as efficiently as sighted participants, especially when the cognitive load of the task increased in the second condition (**Figure 4**).

Finally, we used the Audio-Anchor index to compare the strategies of the two groups when exploring the device and completing the task. The results of the ANOVA highlighted a main effect of the *Group* (iterations = 5,000, $p < 0.001$, $\eta^2 = 0.16$) and of the *Difficulty* (iterations = 4,735, $p < 0.001$, $\eta^2 = 0.09$), but no significant interaction between the two (iterations = 238, $p = 0.3$). *Post hoc* analysis revealed no significant difference between the first and second conditions in the tendency to use the Audio-Anchor as an exploration strategy, regardless of the experimental group (unpaired two-tailed *t*-test based on permutations: Welch's $t = 1.88$, $p = 0.063$). However, the analyses indicated that blind participants were more prone to using this exploratory strategy in both the first and second conditions compared to the sighted group (unpaired two-tailed *t*-test based on permutations: Welch's $t = 2.76$, $p < 0.01$, Cohen's $d = 0.56$) (**Figure 5**).

DISCUSSION

This research investigated how blind people memorize, learn, and process acoustic spatial information and complex auditory spatial structures. To this aim, we adapted the card game "Memory" to be administered with blind and sighted participants in the form of an experimental paradigm with spatialized acoustic items instead of cards. Participants were asked to pair animal calls that were spatially displaced in two experimental conditions of increasing difficulty (the first one with eight and the second one with 24 items). In comparison to sighted individuals, we hypothesized that blind participants would encounter more difficulties when asked to process complex audio-spatial representations. In support of our hypotheses, we observed that the sighted group outperformed the blind group in both conditions by reaching a higher Score (**Figure 3**).

Furthermore, in the more difficult condition with twelve pairs to be matched, blind subjects needed more attempts to proceed with the task and returned more times on the same haptic blocks than the sighted group (**Figure 4**). This suggests that lack of vision may lead to difficulties in integrating the spatial positions of sounds into a coherent and functional spatial representation. Moreover, these results indicate that the absence of visual experience affects the employment of functional spatial exploration strategies in the discovery and memorization of non-visual auditory spatial structures regardless of the experimental condition (**Figure 5**).

Observed differences between blind and sighted participants can be related to difficulties in combining the spatial position of sound sources in a coherent and functional mental

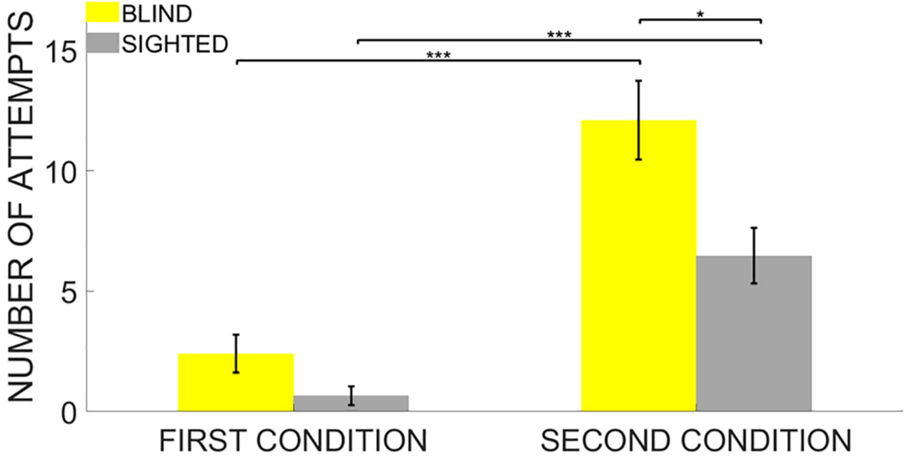


FIGURE 4 | Number of Attempts. Data are presented as mean and standard error for each group. The white circles on the bars represent the individual data. Even though both groups needed more attempts to end the task in the 4-pair compared to the 12-pair conditions, the sighted group needed fewer attempts to pair the items once their locations have been discovered on ARENA2D but only in the second condition. No significant difference between the groups was found in the first condition instead. * indicates $p < 0.05$, *** indicates $p < 0.001$.

representation. The Audio-Memory indeed, requires the active manipulation of spatial information, a mental operation generally affected by congenital blindness. In the haptic modality, blind individuals can construct a mental representation of a tactile layout and remember the locations of the targets on their surface (Vecchi et al., 2005; Cattaneo et al., 2008). The same ability has been observed with acoustic items (Setti et al., 2018). Following the exploration of a complex and meaningful spatial auditory scene, when asked to recall the position of the items composing the layout one by one, blind and sighted participants performed equally well (Setti et al., 2018). Thus, after the exploration of a spatial arrangement, blind people should overall be able to recall the positions of all items composing a certain configuration (Vecchi, 1998; Cornoldi et al., 2000). Mental representations can indeed be built even in the absence of external visual inputs. In general, when the task demand is the simple memorization of items and the cognitive load imposed by the task is not high, blind and sighted individuals perform similarly (Cattaneo et al., 2008; Cattaneo and Vecchi, 2011). Nevertheless, good performance of blind individuals in spatial memory tests, even passive, strongly depends on the demands on memory and on the amount of spatial elaboration that is needed to perform the task. Visual perception is indeed the "preferred modality" in visuo-spatial working memory and previous studies highlighted that also with simple 2D patterns, blind participants performed poorly (Vecchi, 1998). In Setti et al. (2018), we found that blind participants could remember the spatial positions of the stimuli embedded in ARENA2D, the same device used in the current study. The better performance of the sighted group was only ascribed to a better use of the spatial relations among the sounds and not to the simple memorization of their locations. The same pattern of results was confirmed in another work that relied on an acoustic virtual reality system (Setti et al., 2021a), where the blind group could easily remember sounds' locations after a spatial exploration of the virtual environment. In the current study, the task required participants to generate a mental spatial image of the audio spatial structure of the items to be remembered. Thus, it is a more complex task than the simple memorization of items' locations. Specifically, participants were required to memorize and manipulate the mental representation of a complex and dynamic acoustic layout. The dynamic aspect requires a continuous updating process occur while performing the task by maintaining locations in memory. When a new item is discovered, participants must remember its location and, at the same time, update the spatial representation of the scene by adding the uncovered sound's position. Conversely, when two items were paired, their sites were covered by cardboard squares, thus removing them from the represented scene. These processes progressively increased the cognitive load imposed by the experimental paradigm as the subject proceeded toward the end of the task. In this sense, the differences observed between blind and sighted participants may reflect the greater need for blind individuals to use executive functions affected by increasing cognitive load. Along these lines, De Beni and Cornoldi (1988) observed that congenitally blind individuals experience more difficulties in spatial WM tasks that have high memory demands than sighted individuals.

Previous research has also shown that blind people have difficulty dealing with multiple haptic stimuli presented simultaneously (Vecchi et al., 1995). Our results lead to similar conclusions in the context of spatial memory of acoustic items. In the study presented here, we observed that blind participants tended to use an audio-anchor strategy to explore the audio-spatial structure more than sighted participants (**Figure 5**). In other words, blind participants were more likely to build their spatial representation of the auditory structure piece-by-piece by referring all spatial locations to a previously explored position on ARENA2D. Thus, in comparison to sighted, blind individuals seem to be less able to organize and maintain spatialized auditory information thus suggesting that the absence of visual experience confines WM abilities to a more sequential and slower processing of spatial information (Pascual-Leone and Hamilton, 2001; Cattaneo et al., 2008; Ruggiero and Iachini, 2010). As a result of their visual experience, sighted people can better code spatial information in the form of global, externally based representations (Cornoldi et al., 1993; Cattaneo et al., 2008). In line with the calibration theory on the development of multisensory processing of spatial information (Gori et al., 2012), visual experience appears to be fundamental for developing a functional representation of spatial information in structured patterns (i.e., chunks). Previous research on the simultaneous manipulation of multiple stimuli suggests that visual experience is needed to acquire such ability even if stimuli are not visually conveyed (De Beni and Cornoldi, 1988). In this context, we interpret our results as evidence of the influence of visual experience in multisensory processing of simultaneous stimulation. The ability to process different sounds simultaneously and represent spatial information in the form of structured patterns may have helped sighted participants in updating the spatial representation of stimuli's locations during the execution of the Audio-Memory task. Such ability seems to be compromised in blind participants, as expressed by their greater tendency to start consecutive trials by exploring previously discovered items. Similar to previous observations in the context of haptic spatial memory (Ruggiero and Iachini, 2010), blind individuals show limited functional strategies to process spatial information, thus suggesting that their performance required a greater involvement of executive resources compared to sighted.

Since participants generated the auditory feedback through their arm movements, spatial information emerged from sensorimotor contingencies' coupling. In other words, to touch the sensors and emit the sound corresponding to each haptic block, the participant had first to reach the location with their arm. The movement of the arm could have been used as a cue to identify and consequently remember sound positions because the stimulus was generated after the touch. Past research showed that audition provides informative feedback on limb movement, enhancing localization skills after training (Bevilacqua et al., 2016; Cuppone et al., 2019). In the Audio-Memory task presented here, participants coupled their arm movement with spatial acoustic feedback.

Finally, independent of the presence of visual disability, we did not observe significant differences in the performance between adults and adolescents. Given that 12-year-old pre-adolescents have already reached an adult-like performance in a variety of sensory and cognitive tasks (Vuontela et al., 2003; Peterson et al., 2006; Luna, 2009; Scheller et al., 2021) we did not expect age-related differences in performance on the Audio-Memory task. Our results confirm similar age-related achievements for sighted and blind individuals but we cannot exclude that such differences might be present in a younger population tested with the Audio-Memory task. Finally, given that the task is designed in the form of a game, our experimental paradigm would be suitable to pursue studies in this direction to elucidate the influence of blindness in the development of audio-spatial memory skills. As shown here, the adaptation of a game in an experimental protocol allows the use of such a procedure with visually impaired individuals across a wide age range, including those in late childhood and potentially also with younger children. Beyond scientific settings, the Audio-Memory may be adapted for educational purposes as a tool to speed learning and development of new concepts and associations, facilitating the inclusion of visually impaired individuals in educational contexts.

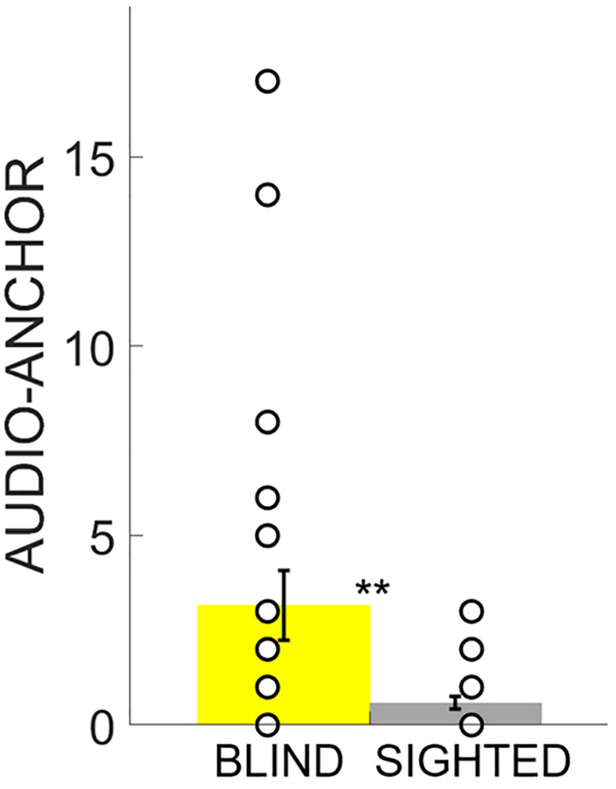


FIGURE 5 | Audio-Anchor. Data are presented as the mean and standard error. The white circles on the bars represent the individual data. The blind group relied more on the use of the Audio-Anchor regardless of the experimental condition. ** indicates $p < 0.01$.

CONCLUSION

This study evaluated spatial and memory skills of blind and sighted individuals and their strategies for exploring complex auditory structures. Early visual deprivation affects the processing and exploration of spatial items embedded in a complex acoustic structure. With higher cognitive demands (such as those required in the 12-pair condition), blind subjects needed more attempts to update the spatial information learned during the task than sighted participants. Furthermore, blind participants relied more on the audio-anchor strategy to explore ARENA2D and build a functional, unified and constantly updated spatial representation. In line with previous findings, limitations previously observed in the haptic domain (Vecchi et al., 2005), held for the auditory modality, thus confirming the pivotal role of visual experience in the active manipulation of memorized spatial information. The current paradigm, designed in the form of a game, can be

TABLE 1 | Blind participants' clinical details.

Participant	Gender	Age	Pathology	Residual vision
S1	M	14	Uveitis	Lights and shadows
S2	F	13	Retinopathy	Light and shadows
S3	F	12	Retinopathy of prematurity	No vision
S4	F	15	Leber's amaurosis	No vision
S5	F	15	Cataract	No vision
S6	M	52	Retinopathy of prematurity	No vision
S7	F	30	Retinopathy of prematurity	No vision
S8	F	12	Glaucoma	Lights and shadows
S9	F	42	Leber's amaurosis	No vision
S10	M	25	Retinitis pigmentosa	No vision
S11	F	52	Retinitis pigmentosa	No vision
S12	F	24	Retinitis pigmentosa	No vision

TABLE 2 | Stimuli employed in both experimental conditions.

Sound	Condition
Bee	1
Lion	1
Frog	1
Rooster	1
Wolf	2
Dog	2
Crow	2
Cat	2
Elephant	2
Birds	2
Hen	2
Sheep	2
Donkey	2
Goose	2
Owl	2
Horse	2

used as a starting point to define novel procedures for cognitive evaluation and rehabilitation. In addition, the Audio-Memory task might be suitable for developing multisensory training to enhance spatial representation through the coupling of auditory and proprioceptive cues. These procedures might be used in those clinical conditions where using the auditory modality can be more effective than vision, such as in the context of visual impairment or cognitive and neuropsychological impairments.

MATERIALS AND METHODS

Participants

Twelve congenitally blind (nine females; age range: 12–52 years, mean age ± SD: 25.5 ± 15.29 years, ethnicity: Caucasian) and twelve sighted (nine females; age range: 12–54 years, mean age ± SD: 25.83 ± 15.86 years, ethnicity: Caucasian) individuals took part in the experiment. In the recruitment process, we used a broad age range because of general difficulties in recruiting congenitally blind individuals. Clinical details relative to their visual impairment are given in **Table 1**. Blind adults were recruited from our institute database and blind adolescents from the "Istituto David Chiossone" based in Genoa, Italy. The local health service ethics committee approved the experiments (Comitato Etico, ASL 3, Genoa, Italy). Parental or adult informed written consent for the study was obtained in all cases. All experiments were performed under The Declaration of Helsinki. None of the sighted or blind participants had any additional sensory or cognitive disabilities.

Setup and Stimuli

The test was performed using a vertical array of speakers, arranged in the form of a matrix (50 × 50 × 10 cm) called ARENA2D (see **Figure 1** for details) that allowed for the serial emission of spatialized sounds.

This device is comprised of 25 haptic blocks (10 × 10 cm^2), each covered by a 4 × 4 matrix of tactile sensors (2 × 2 cm^2 each) (see **Figure 1** for details). When a tactile sensor of a haptic block is touched, touch position is registered. For each haptic block, the sound is emitted from the speaker belonging to the haptic block itself (see the black holes in **Figure 1B**), thus sounds are spatially distributed over the surface of ARENA2D. All blocks are connected in cascade through USB cables [see technical details in Setti et al. (2019)]. Two cardboard grids were used to allow for haptic exploration of the device while performing the task. The grids were developed in collaboration with rehabilitators from the David Chiossone Institute (Genoa, Italy). Following rehabilitators' indications, apertures were as big as the haptic blocks in the 4-pair condition (10 × 10 cm^2), while we used smaller apertures for the 12-pair condition (4 × 4 cm^2) to facilitate haptic exploration and coding of each position's device. To avoid performance being influenced by localization abilities, we chose the size of apertures to overcome auditory localization error previously observed in blind individuals, i.e., 3 cm (Cappagli et al., 2017).

The sounds chosen were distinctive animal calls to ensure ease of discrimination (**Table 2**). All sounds were downloaded from an online database of common licensed sounds[1], equalized and reproduced from the speakers at the same volume. All sound clips lasted 3 s to support easy recognition by participants. Feedback sounds about the performance were emitted from the central speaker of ARENA2D (**Figure 1**): a "Tada!" sound when two items were matched and a recorded voice saying "NO" otherwise. At the end of the task, a jingle was played from the central speaker to make the test engaging. The sound pressure level (SPL) was maintained at 70 dB and the Root Mean Square (RMS) level was calibrated to be the same across the various signals.

Experimental Procedure

The experimental protocol was an adapted audio version of the classic card game "Memory," designed to be performed by blind individuals. Cards were replaced with sounds (animal calls) and

---

[1] https://freesound.org/

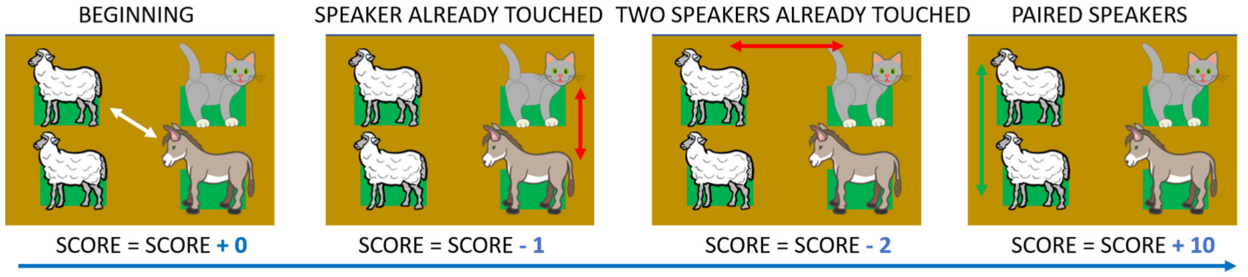


FIGURE 6 | Score, example of calculation. Score is an index that decreases when participants press a panel they have previously chosen. When two blocks are touched for the first time a Score of 0 is allocated. If they have already touched one or both blocks, the Score decreases by one or two, respectively. When a pair is found the Score increases by ten. In the example, if the starting value were equal to zero, the final Score would be: 0 – 1 – 2 + 10 = 7. Depicted animals were downloaded from a royalty-free images web archive (https://publicdomainvectors.org/).

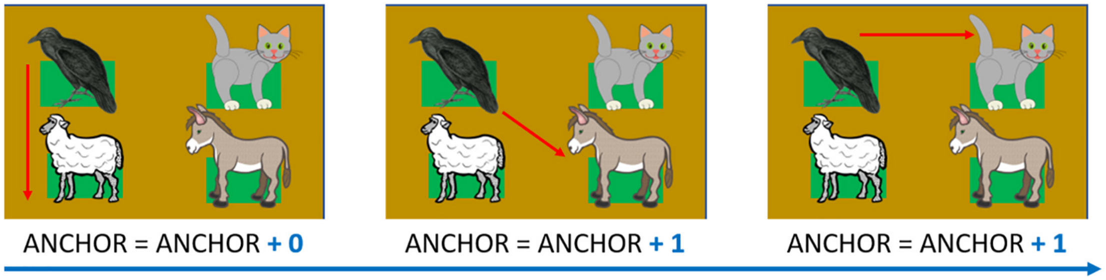


FIGURE 7 | Audio-Anchor, example of calculation. The index equals zero at the beginning of the test. The index increases with more attempts starting with the same haptic block. In the presented example, the final value would be: 0 + 1 + 1 = 2. Depicted animals were downloaded from a royalty-free images web archive (https://publicdomainvectors.org/).

two experimental conditions (**Figure 2**) considered. In the 4-pair condition, participants searched for four pairs of identical animal calls; in the 12-pair condition, there were 12 pairs to be matched. We used different sets of animal calls for each condition. To differentiate between experimental conditions, we took advantage of two cardboard grids that differed in the number of apertures and shape (**Figure 2**). Overall, the 12-pair condition required increased memory load.

During the experiment, subjects sat on a chair at a distance of 30 cm from the device, whose position was adjusted to align the subject's nose with the grid's central aperture. None of the participants had previously interacted with ARENA2D, and the group of sighted participants entered the room already blindfolded. The experimenter guided the subject's hands to explore the grid with eight apertures and counted them with the participants by guiding their hands over the grid and its apertures. After this phase, participants freely touched ARENA2D with both hands to familiarize themselves with the device. During the actual test, subjects were instructed to use the index finger on their dominant hand to select items and explore the device. Before starting the experiment, participants practiced with a trial session. Using the 4-pair condition grid, the experimenter guided the participant's hand, first over two unpaired free slots (that emitted different sounds), and then over two paired items, to familiarize them with the task and the feedback sounds. After this practice session, participants listened to and identified each animal call. The experimenter confirmed that all participants recognized all animal calls. After the recognition phase, the test started with the 4-pair condition (**Figure 2**, left panel). Once this session had finished, the grid for the second experimental condition was placed over the device (**Figure 2**, right panel). Subjects explored this grid by counting the free apertures with the experimenter's help and then exploring the device with no guidance. When the subjects were confident with the grid, the 12-pair condition started. All the subjects were instructed not to move the head through the experiment. In the case of head movements, the experimenter stopped the test to adjust participant's head. The test did not have a fixed duration since it was self-paced. However, each session lasted 25 min on average.

Data Analysis

In both conditions, to quantify subjects' performance, three parameters were calculated to measure memory and exploration strategy. Details of parameters follows.

Score

The number of touches on the same haptic block (**Figure 6**). The more participants touched the same haptic block, the lower their Score. This parameter quantified the overall memory performance in the test. Score was calculated as follows. For each attempt, if both haptic blocks touched were selected for the first time, the total Score was neither increased nor decreased. In fact, in this case, memory processes did not influence item choice as they were not previously discovered. If, in an attempt, only one of the two haptic blocks had already been touched, the total Score decreased by one. If both haptic blocks had already been touched in an attempt, the total Score was decreased by two. In this and the previous case, the Score was decreased to account for inferior performance in recalling the position of the previously uncovered item. When a pair was found, regardless of the number of touches per each haptic block, the total Score was increased by 10. **Figure 6** shows a detailed example of how Score is calculated.

Number of Attempts

This quantifies how many attempts the subject required to pair two identical items once their positions had been discovered on the ARENA2D. The higher the value, the more attempts were needed to pair the sounds, and therefore, the worse the performance. This parameter quantifies the ability to maintain the spatial locations of uncovered items in memory. The Number of Attempts to pair sounds once their locations were discovered on the device was calculated for each possible pairing. For each participant, we summed up the number of trials to pair each couple of sounds (four and twelve pairs for the 4-pair and 12-pair condition, respectively). Then, we averaged these numbers and we obtained a mean number of trials to pair two sounds for each participant. Finally, these means were mediated across all the subjects.

Audio-Anchor

This index accounts for how many consecutive attempts the participant makes by starting with the same haptic block and measures exploration strategy. In other words, this index evaluates how many times the participants started consecutive attempts from the first haptic block of the last pair that they tapped. As this strategy is increasingly adopted, the index increases. For instance, suppose that the participant encounters, in an attempt, the cat meow first and the dog bark after. The Audio-anchor index would increase by one if they began the subsequent attempt again from the same cat's meow position. The index increases until the child starts an attempt by touching a different stimulus position. As in a previous study (Setti et al., 2021b) the audio-anchor provides a measurement of how well the person construct their spatial representation of sound disposition. Thus, the greater the final index, the less the mutual relationship among the stimuli are used (see **Figure 7** for details).

Statistical Analyses

All analyses were carried out in RStudio, Version 1.1.463 (R Core Team (2020), 2020) with non-parametric tests given the small sample size. We first ran a two-way repeated measures MANCOVA based on permutations (*adonis2()* R function) using the R package "vegan" (Dixon, 2003) with Score, Number of attempts and Audio-Anchor as dependent variables, *Group* (either blind or sighted) as between-subject, *Difficulty* (either easy or hard) as within-subject and *Age* (i.e., the age of the participants in decimal number of years) as covariate to check for the influence of age on overall performances. Since we did not find any effect nor interaction given by the age, follow-up ANOVAs, conducted only for the significant interactions, were run with permutation tests [*aovp()* function using the R package "lmPerm" (Wheeler and Torchiano, 2016)]. The lmPerm package use permutation tests to obtain *p*-values for linear models when data do not follow a normal distribution (Wheeler, 2010). In reporting the results of non-normally distributed data, permutation test *p*-values are reported. Finally, *post hoc* analyses were run with two-tailed unpaired Student's *t*-tests based on permutations (*perm.t.test()* R function). Effect sizes were calculated in terms of partial eta-squared (η^2) for ANCOVAs (η^2: small, $> = 0.01$; medium, $> = 0.06$; large, $> = 0.14$) and as Cohen's d value for the t-tests (small, $> = 0.2$; medium, $> = 0.5$; large, $> = 0.8$). Bonferroni correction was used to test the significance of multiple comparison *post hoc* tests ($p < 0.05$ was considered significant).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Comitato Etico, ASL 3, Genoa, Italy. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

WS, LC, and MG conceived the studies and designed the experiments. EC helped in the recruitment of blind participants. All authors wrote and reviewed the manuscript.

FUNDING

This research was supported by the "Istituto David Chiossone," based in Genoa. This research was also partially supported by the MYSpace project (PI MG), which has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (grant agreement No 948349).

ACKNOWLEDGMENTS

We would like to thank Claudio Campus for the technical assistance. Finally, a special thank to all the people who took part in the experiments.

REFERENCES

Ahmad, H., Setti, W., Campus, C., Capris, E., Facchini, V., Sandini, G., et al. (2019). The Sound of Scotoma: audio Space Representation Reorganization in Individuals With Macular Degeneration. *Front. Integr. Neurosci.* 2019:44. doi: 10.3389/fnint.2019.00044

Alais, D., and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262. doi: 10.1016/j.cub.2004.01.029

Arditi, R., and Dacorogna, B. (1988). Optimal foraging on arbitrary food distributions and the definition of habitat patches. *Am. Natural.* 1988:284825. doi: 10.1086/284825

Arnold, A. E. G. F., Burles, F., Krivoruchko, T., Liu, I., Rey, C. D., Levy, R. M., et al. (2013). Cognitive mapping in humans and its relationship to other orientation skills. *Exp. Brain Res.* 224, 359–372. doi: 10.1007/s00221-012-3316-0

Baddeley, A. (1992). Working Memory. *Science* 255, 556–559. doi: 10.4249/scholarpedia.3015

Bertonati, G., Tonelli, A., Cuturi, L. F., Setti, W., and Gori, M. (2020). Assessment of spatial reasoning in blind individuals using a haptic version of the Kohs Block Design Test. *Curr. Res. Behav. Sci.* 1:100004. doi: 10.1016/j.crbeha.2020.100004

Bevilacqua, F., Boyer, E. O., Françoise, J., Houix, O., Susini, P., Roby-Brami, A., et al. (2016). Sensori-motor learning with movement sonification: perspectives from recent interdisciplinary studies. *Front. Neurosci.* 2016:385. doi: 10.3389/fnins.2016.00385

Bonino, D., Ricciardi, E., Bernardi, G., Sani, L., Gentili, C., Vecchi, T., et al. (2015). Spatial imagery relies on a sensory independent, though sensory sensitive, functional organization within the parietal cortex: a fMRI study of angle discrimination in sighted and congenitally blind individuals. *Neuropsychologia* 68, 59–70. doi: 10.1016/j.neuropsychologia.2015.01.004

Bonino, D., Ricciardi, E., Sani, L., Gentili, C., Vanello, N., Guazzelli, M., et al. (2008). Tactile spatial working memory activates the dorsal extrastriate cortical pathway in congenitally blind individuals. *Archives Italiennes de Biologie* 146, 133–146. doi: 10.1162/jocn_a_00208

Burr, D., Banks, M. S., and Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Exp. Brain Res.* 198, 49–57. doi: 10.1007/s00221-009-1933-z

Cappagli, G., Cocchi, E., and Gori, M. (2017). Auditory and proprioceptive spatial impairments in blind children and adults. *Dev. Sci.* 20:3. doi: 10.1111/desc.12374

Carreiras, M., and Codina, B. (1992). Spatial cognition of the blind and sighted: visual and amodal hypotheses. *Cahiers de Psychologie Cognitive*. 12:1.

Cattaneo, Z., and Vecchi, T. (2011). *Blind Vision: The Neuroscience of Visual Impairment*. Cambridge, MA: Mit press. doi: 10.1007/s00221-007-0982-4

Cattaneo, Z., Vecchi, T., Cornoldi, C., Mammarella, I., Bonino, D., Ricciardi, E., et al. (2008). Imagery and spatial processes in blindness and visual impairment. *Neurosci. Biobehav. Rev.* 32, 1346–1360. doi: 10.1016/j.neubiorev.2008.05.002

Cornoldi, C., Bertuccelli, B., Rocchi, P., and Sbrana, B. (1993). Processing Capacity Limitations in Pictorial and Spatial Representations in the Totally Congenitally Blind. *Cortex* 29, 675–689. doi: 10.1016/S0010-9452(13)80290-0

Cornoldi, C., Rigoni, F., Venneri, A., and Vecchi, T. (2000). Passive and active processes in visuo-spatial memory: double dissociation in developmental learning disabilities. *Brain Cogn.* 43, 117–120.

Cornoldi, C., and Vecchi, T. (2003). Visuo-spatial working memory and individual differences. *Book* 2003, 1–182. doi: 10.4324/9780203641583

Cuppone, A. V., Cappagli, G., and Gori, M. (2019). Audio-Motor Training Enhances Auditory and Proprioceptive Functions in the Blind Adult. *Front. Neurosci.* 2019:1272. doi: 10.3389/fnins.2019.01272

Cuturi, L., Cappagli, G., Finocchietti, S., Cocchi, E., and Gori, M. (2017). New rehabilitation technology for visually impaired children and adults based on multisensory integration. *J. Vis.* 2017:592. doi: 10.1167/17.10.592

da Cunha, S. N. S., Travassos Junior, X. L., Guizzo, R., and de Sousa Pereira-Guizzo, C. (2016). The digital memory game: an assistive technology resource evaluated by children with cerebral palsy. *Psicologia* 2016:9. doi: 10.1186/s41155-016-0009-9

De Beni, R., and Cornoldi, C. (1988). Imagery limitations in totally congenitally blind subjects. *J. Exp. Psychol.* 14, 650–655. doi: 10.1037/0278-7393.14.4.650

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *J. Veget. Sci.* 14:927. doi: 10.1658/1100-92332003014[0927:vaporf]2.0.co;2

Doucet, M. E., Guillemot, J. P., Lassonde, M., Gagné, J. P., Leclerc, C., and Lepore, F. (2005). Blind subjects process auditory spectral cues more efficiently than sighted individuals. *Exp. Brain Res.* 160, 194–202. doi: 10.1007/s00221-004-2000-4

Farah, M. J., Hammond, K. M., Levine, D. N., and Calvanio, R. (1988). Visual and spatial mental imagery: dissociable systems of representation. *Cogn. Psychol.* 1988:6. doi: 10.1016/0010-0285(88)90012-6

Féry, Y. A. (2003). Differentiating visual and kinesthetic imagery in mental practice. *Can. J. Exp. Psychol.* 2003:87408. doi: 10.1037/h0087408

Finke, R. A., and Freyd, J. J. (1989). Mental Extrapolation and Cognitive Penetrability: reply to Ranney and Proposals for Evaluative Criteria. *J. Exp. Psychol.* 1989:403. doi: 10.1037/0096-3445.118.4.403

Gori, M., Giuliana, L., Sandini, G., and Burr, D. (2012). Visual size perception and haptic calibration during development. *Dev. Sci.* 15, 854–862. doi: 10.1111/j.1467-7687.2012.2012.01183.x

Gori, M., Sandini, G., Martinoli, C., and Burr, D. C. (2014). Impairment of auditory spatial localization in congenitally blind human subjects. *Brain* 2014:311. doi: 10.1093/brain/awt311

Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R. J., and Belin, P. (2004). Neuropsychology: pitch discrimination in the early blind. *Nature* 430, 309–309. doi: 10.1038/430309a

Hart, R. A., and Moore, G. T. (2017). The development of spatial cognition: a review. *Image Env.* 2017:26. doi: 10.4324/9780203789155-26

Iachini, T., Ruggiero, G., and Ruotolo, F. (2014). Does blindness affect egocentric and allocentric frames of reference in small and large scale spaces? *Behav. Brain Res.* 273, 73–81. doi: 10.1016/j.bbr.2014.07.032

Jicol, C., Lloyd-Esenkaya, T., Proulx, M. J., Lange-Smith, S., Scheller, M., O'Neill, E., et al. (2020). Efficiency of Sensory Substitution Devices Alone and in Combination With Self-Motion for Spatial Navigation in Sighted and Visually Impaired. *Front. Psychol.* 2020:1443. doi: 10.3389/fpsyg.2020.01443

Juurmaa, J., and Lehtinen-Railo, S. (1994). Visual Experience and Access to Spatial Knowledge. *J. Vis. Impair. Blind.* 88, 157–170.

King, A. J., and Parsons, C. H. (1999). Improved auditory spatial acuity in visually deprived ferrets. *Eur. J. Neurosci.* 1999:821. doi: 10.1046/j.1460-9568.1999.00821.x

Kolarik, A. J., Cirstea, S., and Pardhan, S. (2013). Evidence for enhanced discrimination of virtual auditory distance among blind listeners using level and direct-to-reverberant cues. *Exp. Brain Res.* 224, 623–633. doi: 10.1007/s00221-012-3340-0

Kosslyn, S. M. (1980). *Image and Mind*. Cambridge, MA: Harvard University Press, doi: 10.1017/CBO9780511551277

LeBoutillier, N., and Marks, D. F. (2003). Mental imagery and creativity: a meta-analytic review study. *Br. J. Psychol.* 2003:712603762842084. doi: 10.1348/000712603762842084

Lederman, S. J., and Klatzky, R. L. (1990). Haptic classification of common objects: knowledge-driven exploration. *Cogn. Psychol.* 1990:90009. doi: 10.1016/0010-0285(90)90009-S

Lehnert, G., and Zimmer, H. D. (2006). Auditory and visual spatial working memory. *Mem. Cogn.* 34, 1080–1090. doi: 10.3758/BF03193254

Lehnert, G., and Zimmer, H. D. (2008). Modality and domain specific components in auditory and visual working memory tasks. *Cogn. Proc.* 2008:6. doi: 10.1007/s10339-007-0187-6

Lewald, J. (2002). Vertical sound localization in blind humans. *Neuropsychologia* 40, 1868–1872. doi: 10.1016/S0028-3932(02)00071-4

Logie (1995). Visuo-spatial Working Memory. *VisuoSpatial Work. Mem.* 1995:746. doi: 10.1002/acp.746

Luna, B. (2009). Developmental Changes in Cognitive Control through Adolescence. *Adv. Child Dev. Behav.* 2009:9. doi: 10.1016/S0065-2407(09)03706-9

Millar, S. (1976). Spatial representation by blind and sighted children. *J. Exp. Child Psychol.* 1976:6. doi: 10.1016/0022-0965(76)90074-6

Morrongiello, B. A., Humphrey, G. K., Timney, B., Choi, J., and Rocca, P. T. (1994). Tactual object exploration and recognition in blind and sighted children. *Perception* 1994:230833. doi: 10.1068/p230833

Mou, W., and McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *J. Exp. Psychol.* 28, 162–170. doi: 10.1037/0278-7393.28.1.162

Palmer, S. (2000). Working memory: a developmental study of phonological recoding. *Memory* 8, 179–193. doi: 10.1080/096582100387597

Pascual-Leone, A., and Hamilton, R. (2001). The metamodal organization of the brain. *Prog. Brain Res.* 2001:1. doi: 10.1016/S0079-6123(01)34028-1

Pasqualotto, A., Spiller, M. J., Jansari, A. S., and Proulx, M. J. (2013). Visual experience facilitates allocentric spatial representation. *Behav. Brain Res.* 236, 175–179. doi: 10.1016/j.bbr.2012.08.042

Peterson, M. L., Christou, E., and Rosengren, K. S. (2006). Children achieve adult-like sensory integration during stance at 12-years-old. *Gait Post*. 2006:003. doi: 10.1016/j.gaitpost.2005.05.003

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W. H. C., Cohen, L., et al. (2004). Beyond sensory images: object-based representation in the human ventral pathway. *Proc. Natl. Acad. Sci. U S A* 2004:400707101. doi: 10.1073/pnas.0400707101

Puspitawati, I., Jebrane, A., and Vinter, A. (2014). Local and Global Processing in Blind and Sighted Children in a Naming and Drawing Task. *Child Dev*. 2014:12158. doi: 10.1111/cdev.12158

R Core Team (2020). *R: A language and environment for statistical computing. In R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rinck, M., and Denis, M. (2004). The metrics of spatial distance traversed during mental imagery. *J. Exp. Psychol.* 2004:1211. doi: 10.1037/0278-7393.30.6.1211

Roder, B., Teder-Sälejärvi, W., Sterr, A., Rosler, F., Hillyard, S. A., and Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature* 400, 162–166. doi: 10.1038/22106

Rossetti, Y., Gaunet, F., and Thinus-Blanc, C. (1996). Early visual experience affects memorization and spatial representation of proprioceptive targets. *NeuroReport* 1996:25. doi: 10.1097/00001756-199604260-00025

Ruggiero, G., and Iachini, T. (2010). The Role of Vision in the Corsi Block-Tapping Task: Evidence From Blind and Sighted People. *Neuropsychology* 24, 674–679. doi: 10.1037/a0019594

Ruggiero, G., Ruotolo, F., and Iachini, T. (2012). Egocentric/allocentric and coordinate/categorical haptic encoding in blind people. *Cogn. Proc*. 2012:6. doi: 10.1007/s10339-012-0504-6

Scheller, M., Proulx, M. J., de Haan, M., Dahlmann-Noor, A., and Petrini, K. (2021). Late- but not early-onset blindness impairs the development of audio-haptic multisensory integration. *Dev. Sci*. 2021:13001. doi: 10.1111/desc.13001

Setti, W., Cuturi, L. F., Cocchi, E., and Gori, M. (2018). A novel paradigm to study spatial memory skills in blind individuals through the auditory modality. *Sci. Rep*. 2017, 1–10. doi: 10.1038/s41598-018-31588-y

Setti, W., Cuturi, L. F., Engel, I., Picinali, L., and Gori, M. (2021a). The Influence of Early Visual Deprivation on Audio-Spatial Working Memory. *Neuropsychology*. 2021:776. doi: 10.1037/neu0000776

Setti, W., Cuturi, L. F., Sandini, G., and Gori, M. (2021b). Changes in audio-spatial working memory abilities during childhood: the role of spatial and phonological development. *PLoS One* 16:260700. doi: 10.1371/journal.pone.0260700

Setti, W., Cuturi, L. F., Maviglia, A., Sandini, G., and Gori, M. (2019). ARENA:a novel device to evaluate spatial and imagery skills through sounds. *Med. Measur. Appl.* 2019:8802160. doi: 10.1109/MeMeA.2019.8802160

Sunanto, J., and Nakata, H. (1998). Indirect Tactual Discrimination of Heights by Blind and Blindfolded Sighted Subjects. *Percept. Motor Skills* 86, 383–386. doi: 10.2466/pms.1998.86.2.383

Thinus-Blanc, C., and Gaunet, F. (1997). Representation of space in blind persons: vision as a spatial sense? *Psychol. Bull.* 121, 20–42. doi: 10.1037/0033-2909.121.1.20

Vanlierde, A., and Wanet-Defalque, M. C. (2004). Abilities and strategies of blind and sighted subjects in visuo-spatial imagery. *Acta Psychol.* 116, 205–222. doi: 10.1016/j.actpsy.2004.03.001

Vecchi, T. (1998). Visuo-spatial imagery in congenitally totally blind people. *Memory* 6, 91–102. doi: 10.1080/741941601

Vecchi, T., Monticellai, M. L., and Cornoldi, C. (1995). Visuo-spatial working memory: structures and variables affecting a capacity measure. *Neuropsychologia* 33, 1549–1564. doi: 10.1016/0028-3932(95)00080-M

Vecchi, T., Richardson, J. T. E., and Cavallini, E. (2005). Passive storage versus active processing in working memory: evidence from age-related variations in performance. *Eur. J. Cogn. Psychol.* 17, 521–539. doi: 10.1080/09541440440000140

Vecchi, T., Tinti, C., and Cornoldi, C. (2004). Spatial memory and integration processes in congenital blindness. *Neuroreport* 15, 2787–2790.

Vercillo, T., Burr, D., and Gori, M. (2016). Early visual deprivation severely compromises the auditory sense of space in congenitally blind children. *Dev. Psychol.* 52, 847–853. doi: 10.1037/dev0000103

Vingerhoets, G., De Lange, F. P., Vandemaele, P., Deblaere, K., and Achten, E. (2002). Motor imagery in mental rotation: an fMRI study. *NeuroImage*. 2002:1290. doi: 10.1006/nimg.2002.1290

Voss, P., Lassonde, M., Gougoux, F., Fortin, M., Guillemot, J. P., and Lepore, F. (2004). Early- and late-onset blind individuals show supra-normal auditory abilities in far-space. *Curr. Biol*. 2004:51. doi: 10.1016/j.cub.2004.09.051

Vuontela, V., Steenari, M., and Koivisto, J. (2003). Audiospatial and Visuospatial Working Memory in 6 – 13 Year Old School Children. *Learn. Memory* 10, 74–81. doi: 10.1101/lm.53503.which

Wheeler, B., and Torchiano, M. (2016). *lmPerm: Permutation tests for linear models. R package version 1.1-2. In Cran*.

Wheeler, R. E. (2010). Permutation Tests for Linear Models in R. *R Document.* 2010, 1–36.

Yates, A. J. (1966). Psychological Deficit. *Annu. Rev. Psychol*. 1966:551. doi: 10.1146/annurev.ps.17.020166.000551

Zimler, J., and Keenan, J. M. (1983). Imagery in the congenitally blind: how visual are visual images? *J. Exp. Psychol.* 9, 269–282. doi: 10.1037/0278-7393.9.2.269

Zwiers, M. P., Van Opstal, A. J., and Cruysberg, J. R. (2001). A spatial hearing deficit in early-blind humans. *J. Neurosci.* 2001:2001. doi: 10.1523/jneurosci.21-09-j0002.2001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 14 June 2022
doi: 10.3389/fpsyg.2022.806093

Modifying Sensory Afferences on Tablet Changes Originality in Drawings

Fabien Bitu[1], Béatrice Galinon-Mélénec[2] and Michèle Molina[1]*

[1] Laboratory of Psychology of Caen Normandy (LPCN EA7452), University of Caen Normandy, Caen, France, [2] Identity and Differentiation of Spaces, Environment and Societies (UMR IDEES 6266 CNRS), University Le Havre Normandy, Le Havre, France

According to some recent empirical studies revealing that creativity is linked to sensorimotor components, the current research was aimed at evaluating whether sensory afferences could modulate originality in drawing of children and adolescents. Sixty-nine children from 1st, 3rd, 6th, and 8th grades were required to produce a man who exists and a man who doesn't exist with fingers or stylus on a tablet and with a pen on paper. Drawings were assessed with an originality scale comparing original drawings to unoriginal ones. Since, in comparison to drawings made on paper with a pen, drawing with fingers enhances proprioceptive information, this condition was expected, according to cognitive load theory, to favor originality in drawing by reducing cognitive resources devoted to motor control of the graphic gesture (lowering intrinsic load). On the contrary, since the use of a stylus involves a proprioceptive loss of information, which enhances intrinsic load by increasing cognitive resources devoted to motor control, it was expected that drawing with a stylus on the tablet would lead to the least original drawings. Results only partially confirmed these hypotheses. While the use of fingers on the tablet led to the highest original scores, using a stylus on the tablet did not impair originality in drawing of children and adolescents. On the opposite, the use of a stylus led 3rd–8th graders to perform better than with pen on paper. This modulation of the tool on originality does not confirm the hypotheses formulated in accordance with the cognitive load framework. However, it could be explained according to an embodied perspective of creativity considering the creative process as relying on a sensorimotor prediction process in which sensory afferences are central to generating and evaluate creative ideas. This research opens new avenues on creativity and proposes to consider the development of predictive motor control as a significant part of creativity development.

Keywords: creativity, tablet, children, adolescent, sensory afference, cognitive load

OPEN ACCESS

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Denis Alamargot,
Université Paris-Est Créteil Val de Marne, France
Safinah Ali,
Massachusetts Institute of Technology, United States

****Correspondence:***
Fabien Bitu
fabien.bitu@unicaen.fr

Specialty section:
This article was submitted to Cognition,
a section of the journal
Frontiers in Psychology

Received: *31 October 2021*
Accepted: *18 May 2022*
Published: *14 June 2022*

Citation:
Bitu F, Galinon-Mélénec B and Molina M (2022) Modifying Sensory Afferences on Tablet Changes Originality in Drawings.
Front. Psychol. 13:806093.
doi: 10.3389/fpsyg.2022.806093

INTRODUCTION

How can new technologies enhance our knowledge of the creative process? Multiple attempts have been made to understand and describe human creativity through the years, from the mystic idea of creativity breathed by gods with Plato to the philosophical conception of the creative genuine (Duff, 1767; Galton, 1879, 1883; James, 1880). But it is only in the 1950s that creativity was investigated as a field in psychological studies (Guilford, 1956, 1967), leading around 40 years later to a consensual definition of creativity, conceived as "the capacity to realize something new and adapted to the context" (MacKinnon, 1962; Barron, 1988; Sternberg, 1988; Ochse, 1990; Sternberg and Lubart, 1991, 1995, 1996; Lubart et al., 2003). According to this definition, the creative product must be adapted to the context's criterion and be new enough to stand out from common products. Creativity lies upon three indicators (Mouchiroud and Lubart, 2001) which are the most used for evaluating creativity: fluency (number of appropriate responses), flexibility (variability of themes in responses), and originality (novelty of responses). The latter has long been considered the main component of creativity, corresponding to the evaluation and selection of own's idea, based on prior knowledge (Mouchiroud and Lubart, 2001; Lubart and Georgsdottir, 2004). Lubart et al. (2003) consider creativity as resulting from an interactive combination of cognitive (including convergent and divergent thinking, intelligence, knowledge), conative (personality traits and motivation), emotional and environmental (family, school) factors including new technologies, leading to different levels of creativity. This multivariate approach to creativity is to date one of the most documented and developed in experimental research.

In addition to these factors, it has recently been proposed to focus on a forgotten part of creativity: sensorimotor components. As a matter of fact, some recent studies focused on the links shared by these two processes (e.g., Nikolaeva et al., 2018; Gaggioli et al., 2019; Fleury et al., 2020). Recently, neuroscience studies highlighted the involvement of motor regions, such as the premotor cortex, the supplementary and pre-supplementary motor areas, and the posterior inferior frontal gyrus, in musical creativity (Bashwiner, 2018; Bashwiner and Bacon, 2019). The present exploratory study aims to extend this embodied perspective of creativity by focusing on sensorimotor afferences during creative drawings.

To what extent sensory afferences can modulate creative drawing executed on a digital screen? Although this question has not been investigated to our knowledge in the case of creativity, the multimedia learning theory (Mayer, 2005) usually applied to the context of digital learning could help in bringing some answers. Multimedia learning refers to any situation in which information processing occurs through two different channels, for instance, audio and visual, which is common in the digital context. Multimedia learning can be explained according to cognitive load theory (Mayer and Moreno, 2003; Sweller et al., 2011). Cognitive load theory proposes that there is a delimited amount of cognitive resources available for a specific task. In a learning task, cognitive resources can be allocated to the *essential processing* for the comprehension of the task; to the *intrinsic processing* devoted to the realization of the task, such as controlling gestures in a physical interaction learning task; or to the *extrinsic processing* to manage information primed by the design of the task but which are not essential in its comprehension, such as adding background music in a narrative learning (Mayer and Moreno, 2003; Sweller et al., 2011). Applied to a sensorimotor task in the multimedia learning context, increasing the processing demands inherent to the sensory modalities used in the task may exceed the processing capacity of the cognitive system, leading to a cognitive overload (Brown et al., 2009). As stated in Bara and Tricot (2017), this overload could be due to an increased intrinsic load, according to which the kind of movement performed during a task would make the task harder to accomplish. For example, while comparing tactile exploration of concave *versus* raised letters, Bara and Gentaz (2011) found that the strategy used by children in concave letter exploration led to higher cognitive costs than the one used in raised letter condition, resulting in less memorized concave letters than raised ones.

The rationale of the present study is the following: in order to investigate if, as suggested by recent findings, sensorimotor components are involved in creativity, then varying sensory afferences during a drawing creative task that reduce or increase the intrinsic cognitive load devoted to the motor control of the task would have consequences on originality performances. Tablets are electronic devices presenting a visual screen controlled by gestures and tactile actions executed with fingers or stylus (Dubé and McEwen, 2015). This tactile feature allows interacting very easily with the device, even at a young age, using a stylus or directly with finger (Cooper, 2005; Geist, 2012; McManis and Gunnewig, 2012; Dubé and McEwen, 2015; Piotrowski and Meester, 2018; Sakr, 2018). However, acting with fingers or with stylus on a tablet modifies afferent and efferent kinesthetic and proprioceptive information. Acting with fingers could reduce the cognitive load devoted to the motor control during the writing digital task whereas acting with a stylus would increase it. Alamargot and Morin (2015) demonstrated that compared to the roughness of a paper sheet, the smoothness of a tablet lessens the friction between the pen and the surface of the tablet which reduces proprioceptive feedback when used with a stylus. This reduction of proprioceptive feedback implies a compensating strategy to control the executing writing movement (Alamargot and Morin, 2015; Gerth et al., 2016; Guilbert et al., 2019), thus increasing the cognitive cost allocated to the motor control to complete the task – the intrinsic load. According to cognitive load theory, using a stylus during a learning task should increase the intrinsic load by requiring supplementary cognitive resources allocated to the movement control, resulting in poorer creative performances. On the contrary, acting with finger enhances the friction with the screen and therefore increases proprioceptive feedback, which should lead to a reduced intrinsic load in a learning task by facilitating motor control strategy. These variations in proprioceptive feedback consecutive to the tool used for acting on the tablet may affect differently children and adolescents as a consequence of their ability to process proprioceptive afferences. As a matter of fact, proprioception develops during childhood and adolescence as a consequence of precision improvement such that proprioception becomes more reliable over time (Holst-Wolf et al., 2016). More precisely, children younger than 8 years have some difficulty in processing proprioceptive information, leading them to rely more heavily on visual rather than proprioceptive information during an action to control that the performed movement is consistent with the desired one (Bard and Hay, 1983; Contreras-Vidal, 2006). Studies have demonstrated that variation of afferent feedback modifies writing learning abilities in preschoolers (Patchan and Puranik, 2016) as well as motor control of action (Alamargot and Morin, 2015).

Patchan and Puranik (2016) showed that using fingers on the tablet was the most efficient set thanks to an enhanced direct proprioceptive feedback on the tablet for handwriting learning. They trained 3–6 years old preschoolers to write on the tablet with fingers or with stylus, or on paper with a pen. After practicing 3 times every week for a total of 8 weeks, children wrote letters more correctly with fingers on the tablet than with stylus on the tablet or pen on paper. From these results, it could be suggested that direct contact of fingers on tablet fostered sensorimotor processing in children through the richness of proprioceptive information and favored motor learning. Results from Patchan and Puranik (2016) also showed that acting on the tablet with a stylus did not bring the same benefits as when using fingers: the sliding effect of the stylus lowered proprioceptive feedback which led to a less efficient strategy of motor control. In the same vein, Alamargot and Morin (2015) revealed that using a stylus on the smooth screen of a tablet during a writing task, compared to using a pen on paper, diminished the legibility of letters at second (6–7 years) and nine grades (14–15 years). Gerth et al. (2016) observed that in addition to preschoolers and second graders, adults were also affected by the sliding effect of the tablet's surface when using a stylus.

Contrary to young children, older ones are able to compensate for the loss of haptic information needed for movement control when writing with a stylus on a tablet (Alamargot and Morin, 2015; Guilbert et al., 2019), by producing bigger letters (amplification of movement) and writing faster (velocity increase). This compensating strategy differs according to the age of the participants (Alamargot and Morin, 2015; Gerth et al., 2016). While 7/8 years old children increase pauses, leading to a longer duration of movements, 10/11 years old children favor increasing pen pressure and speed to maximize proprioceptive sensory input (Alamargot and Morin, 2015). Compensating strategies are tied to the developmental trend in proprioceptive processing which also induces changes in the ability of children and adolescents to use internal models to control actions.

Internal models of actions allow the brain to mimic the transformation between the motor command and sensory signals (Kawato, 1999). In order to control ongoing actions through accurate and fluent movements, we make use of feedforward processes (Miall and Wolpert, 1996) allowing a prediction of the movements to be executed as well as a prediction of their related sensory consequences. In this case, two types of internal models are used. First, inverse models are used as a controller to select an adapted motor program. Second, direct models are used as a predictor of sensory consequences of the ongoing movement, informed by a copy of the motor program issued from inverse models – the efference copy (Wolpert and Kawato, 1998). Thus, feedforward internal models are predictive representations transforming action commands into their sensory consequences. This prediction process supported by internal models of action can also be used to control off-line actions, in a simulation process, like imagined action such as motor imagery (Jeannerod, 1994) or creativity. In the case of motor imagery in which covert actions are decoupled from any actual sensorimotor interaction (Jeannerod, 1994), internal models allow simulating future or potential motor actions without realizing it (Guilbert et al., 2018).

A significant increase in the use of predictive motor control is commonly observed around 8 years (Hay, 1978, 1979; Bard and Hay, 1983; Molina et al., 2008) followed by a second one around 11–12 years (Caeyenberghs et al., 2009; Smits-Engelsman and Wilson, 2013). This means that the loss of proprioceptive feedback could be compensated by maximizing proprioceptive afferences through the use of predictive motor control from the age of 8. Consequently, Guilbert et al. (2019) observed that reducing sensory feedback affected handwriting performances more in young than older children. Older children exaggerated letter size and pen pressure to maximize the amount of proprioceptive information and made shorter pauses suggesting an increased use of feedforward motor control. After 8 years, the sliding effect of the stylus on a tablet could thus be compensated by the use of feedforward motor control of action (Kandel and Perret, 2015; Guilbert et al., 2019). These results suggest that children older than 8 May exaggerate their movement to maximize the use of proprioceptive information of the ongoing action to compensate for the loss in haptic feedback. Using a stylus reduces haptic feedback, requiring feedforward sensory prediction to compensate for the unavailable information. However, it is only after 8 years that children would be able to compensate for the loss of tactile afferences as a consequence of an improvement in predictive processes.

These results obtained for the writing task offer some interesting issues concerning creative drawings. Creativity, as a cognitive process, could be modulated by the cognitive load processes involved in learning tasks. The ability to produce original drawings could depend on the cognitive load induced by the modification of sensory afferences primed by the task. The different proprioceptive feedback induced by the use of tablets offers a particularly well-adapted situation to vary this cognitive load.

To our knowledge, only one study dealt with the question of a possible link between direct fingertip feedback on tablets and originality (Bitu et al., 2019). They invited children aged 6–7 and 8–9 to draw, based on Karmiloff-Smith (1990), "a man that exists" (unoriginal drawing) and "a man that doesn't exist" (original drawing) with fingers on the tablet and with a pen on paper. Originality in drawings was assessed by means of a graphical scale developed by the authors, inspired by the study of Karmiloff-Smith (1990), which allowed to compute an originality score by comparing original drawings to unoriginal drawings of each child. Results showed that both 6–7- and 8–9-years old children were more original with fingers on the tablet than with pens on paper. From this study, Bitu et al. (2019) concluded a facilitator effect of finger use on a tablet for original drawing in comparison to a pen on paper. Yet, by comparing fingers on a touchscreen versus pen on paper, no conclusions could be stated about the nature of the observed benefits, which could be due to the tool (finger versus pen) as well as to the surface (touchscreen versus paper) on which drawings were produced. The use of a stylus on touchscreen was thus essential to disentangle the link between sensorimotor components and creativity. Indeed, as stated previously, while using fingers on the tablet is known to enhance tactile feedback on fingertips (Patchan and Puranik, 2016), using a stylus on a tablet leads to poorer proprioceptive feedback (Alamargot and Morin, 2015). The present study was aimed at extending this previous work by evaluating whether varying sensory afferences will impact the originality of children and adolescents drawing with finger and stylus on a tablet, and with pen on paper. To assess originality, we adapted our experiment from Karmiloff-Smith (1990) and Bitu et al. (2019) in which children operated several types of change between the drawing of a man who exists and a man who doesn't exist. We compared originality performance of four different age groups related to the development of the capacity to generate and use internal models for the control of on-line and off-line action, at 6–7 (1st graders), 8–9 (3rd graders), 11–12 (6th graders), and 13–14 years old (8th graders). These ages correspond to periods before (1st grade), during (3rd and 6th grade), and after (8th grade) the transition commonly reported in the literature concerning the use of internal models of action.

This study addresses the effect of modifying proprioceptive feedback to modulate originality in drawings. In the light of empirical studies, two hypotheses can be drawn. If creativity is related to sensory afferences, then it was expected that drawing with fingers on the tablet would increase originality at all ages, compared to drawings made with pen on paper. Increasing proprioceptive feedback on fingertips may enhance motor control strategy and, as a consequence, would reduce the cognitive load allocated to motor control in favor of the drawing task. Second, it was expected that, compared to drawings made with pen on paper, using a stylus would decrease originality in drawings at all ages as a consequence of an increased cognitive load this situation induces to compensate for the reduced sensory feedback available to the motor control execution of the drawing task.

MATERIALS AND METHODS

Participants

An initial sample of 70 participants was recruited in two schools located in Normandy, France. Participants were enrolled on a voluntary basis, following a convention that defined data collection, with the consent of Normandy education academy and schools, along with written authorization from their parents or legal tutor. Inclusion criteria were not met for 1 adolescent. None expressed their wish to abort the experiment before the end of the procedure, leading to a total of 69 children and adolescents aged from 6 to 14 years old (mean age = 10 years, 0 month) retained for this study. Children and adolescents in school were recruited in 1st grade (*n* = 15; mean age = 6 years, 11 months; min = 6 years, 6 months; max = 7 years, 4 months), 3rd grade (*n* = 22; mean age = 8 years, 10 months; min = 8 years, 2 month; max = 9 years, 3 months), 6th grade (*n* = 18; mean age = 11 years, 4 months; min = 10 years, 5 months; max = 12 years, 0 month) and 8th grade (*n* = 14; mean age = 13 years, 5 months; min = 12 years, 9 months; max = 14 years, 1 month).

Materials

Digital drawings were made on a Microsoft Surface Pro 4 tablet, 12.3″ screen and 2736 × 1824 px resolution, with a Microsoft Surface Pen measuring 144 mm × 9.5 mm × 10.2 mm. Drawings were made on an app designed in our laboratory, presenting a white surface on which it can be drawn with stylus or fingers by tracing only black traits. Erasing function was not enabled on the app. Paper sheets measuring the same size as the screen (260 mm × 175 mm), and a black pencil were used. Eraser was not allowed when drawing on paper.

Inclusion and Exclusion Criteria

For each participant, exclusion criteria were related to the task feasibility such as identified severe visual impairment (e.g., low vision, blindness) or severe motor impairment (e.g., excessive weakness), that would make the participant unable to perform the task correctly. To this end, a NEPSY-II visuomotor precision task (Korkman et al., 2014) was performed to measure graphomotor speed and accuracy during a graphical task. In this task, participants were asked to draw lines inside of tracks as quickly and accurately as possible. The number of errors and completion time are recorded in this test to compute a visuomotor precision score. The lower the score's value, the higher the visuomotor precision level. This test allowed to control for visuomotor precision impairment that could affect a drawing task, by excluding each participant under 2*SD* from the standardized mean of its age group.

Control Tasks

Prior to the experimentation, and after completing the NEPSY-II visuomotor task, each participant was asked to perform a functioning control task. The functioning task allowed control of children and adolescents' knowledge of the core movements to be applied on a tablet. Participants were asked to perform eight subtasks with Instagram used offline, referring to eight core movements from the touch gesture reference guide of Villamor et al. (2010). This task's procedure was sum up in **Table 1**. If a child or an adolescent was in difficulty with one of the instructions, the experimenter could help him/her by verbalizing the movement required to accomplish the given instruction (for example, "to open the app, you have to double tap").

Experimental Procedure

Children and adolescents were observed individually in a quiet room of the school and high school. They sat down in front of a table (700 mm × 500 mm × 750 mm) where the touchscreen or the sheet of paper was placed. Each participant had to produce three original and three unoriginal drawings of a man with fingers on the tablet, with stylus on the tablet and with a pen on paper following specific instructions. Following the procedure used by Karmiloff-Smith (1990) with pen and paper to bring creativity to drawings, children were asked to draw "a man who exists" for unoriginal drawings and "a man who doesn't exist" for original drawings. Tool order (pen on paper, finger on tablet, and stylus on tablet) and instruction order (original and unoriginal) were systematically counterbalanced across each participant. Once a

TABLE 1 | Description of the functioning task procedure according to the touch gesture reference guide (Villamor et al., 2010).

Instruction	Targeted movement from Villamor et al. (2010)
Move Instagram icon on desktop	Press and drag
Start Instagram app	Double tap
Take a picture and add a smiley	Single tap
Enlarge the smiley	Spread
Shrink it	Pinch
Move it	Tap and drag
Rotate it	Rotate
Turn smiley page to see recent smileys	Flick

first drawing was produced, it was removed before starting the execution of a new one. During the whole experimental session, the experimenter stood near the participant to help him/her the understand the task if he/she needed to.

Coding

A total of 414 individual electronic (276) and paper (138) drawings were collected for analysis: one half being original drawings and the other half unoriginal drawings. Unoriginal drawings were used as a baseline to assess originality considering interindividual differences in the representation of a man in a drawing. The logic was to consider as being original in the drawings, any modification operated in comparison to the unoriginal man drawn with the same tool. Originality in drawing was thus rated by comparing an original drawing to an unoriginal one made with the same tool by the same participant, allowing to score any modification operated from one condition to another with the same tool.

Several types of graphical changes were considered for the coding. As reported by Karmiloff-Smith (1990), seven categories of changes can be observed for the drawing of something (house, man, or animal) that does not exist in comparison to the drawing of something that exists as follows:

- Addition and deletion of elements (e.g., 2 heads instead of 1, or deletion of the arms), coded in our scale as the addition or deletion of elements that were already drawn on the drawing of a man who exists;
- Modification operated on the shape of elements (e.g., a triangular head instead of a circle one);
- Modification operated on the size of elements (e.g., a head two times bigger than the trunk);
- Insertion of new elements (e.g., wings added to the trunk), referring to the insertion of something that was not drawn on the drawing of a man who exist;
- Position, orientation, and perspective modifications of elements or the whole drawing (e.g., eyes and mouth inverted);
- Cross-conceptual categories modifications of the whole drawing (e.g., an animal-shaped man);
- Modifications operated on the shape of the whole drawing (e.g., a circle-shaped man).

These seven categories considered global modifications operated on the whole drawing and local modifications operated on elements of the drawing. Local modifications concerned 15 elements defined as parts constituting a man, whose presence is rated in Goodenough (1926) scale of a man drawing. Goodenough (1926) considered the head, eyes, nose, mouth, nostrils, ears, hairs, neck, trunk, arms, hands, fingers, legs, and feet. To this list, we added teeth as an element of the drawing since children could use this part of the face in an original way.

For each modification operated on elements (local modifications), 0.5 point was assigned. Each global modification was awarded the double, i.e., 1 point, since it concerned larger modifications than those consecutive to a local element modification. Only one exception to this was for "position, orientation, and perspective" category which could be rated up to 1.5 points since it could concern simultaneously local and global modifications. As an example, if a participant exchanged the mouth with an eye, it would concern at least two different elements of the "position change" category, which would result in a strongly increased originality score. This category was thus awarded 0.5 point for a local modification, no matter the number of concerned elements. One mouth and one eye exchanged would thus be awarded the same point (0.5) as if it was exchanged with two eyes. Concerning a modification of the whole drawing in position change (for example, a reversed man), 1 point could be awarded for a global modification.

To sum up, with the rating scale, the originality score calculated for each original drawing was defined by the sum of points awarded to local and global modifications made on the original drawing in comparison to the unoriginal drawing made by the same participant with the same tool. Three originality scores were thus calculated using the rating scale (**Table 2**), one for each original drawing executed on the touchscreen with fingers, with stylus, and on paper with a pen. The calculation of originality score in drawings was thus consistent with the definition of originality, that is, the capacity to assess and select a novel idea among all other own ideas, depending on the acquired knowledge, i.e., what children and adolescents know about the drawing of a man (Runco, 1991; Lubart and Georgsdottir, 2004).

The originality score could be rated up to 33.5 points for each original drawing but since the originality scale aims to cover a large range of modification possibilities in order to capture each modification operated on the original drawing, a single original drawing will most probably use only a few of this range of possibilities, leading to relatively low scores on the 33.5 total available points.

Data Analysis

To ensure that instructions were well understood, an interrater agreement score was calculated with a random sample of 100 unoriginal and original drawings produced in the three conditions (finger on tablet, stylus on tablet, and pen on paper). Two naive observers were asked to categorize the drawings as being original or unoriginal. Cohen's Kappa showed

TABLE 2 | Rating scale used to assess originality in drawings.

Graphical change	Definition	Type of change	Awarded points (min–max)
Deletion/addition (**Figure 1**)	Element in the drawing of a man who exist (head, eyes, nose, mouth, nostrils, teeth, ears, hairs, neck, trunk, arms, hands, fingers, legs, feet) replicated or deleted	Core element	0–7.5
Shape of elements (**Figure 2**)	Element (head, eyes, nose, mouth, nostrils, teeth, ears, hairs, neck, trunk, arms, hands, fingers, legs, feet) whose shape was different from the drawing of the man who exist	Core element	0–7.5
Size of elements (**Figure 3**)	Element (head, eyes, nose, mouth, nostrils, teeth, ears hairs, neck, trunk, arms, hands, fingers, legs, feet) whose size was different from the drawing of a man who exist	Core element	0–7.5
Insertion of new elements (**Figure 4**)	New element (head, eyes, nose, mouth, nostrils, teeth, ears hairs, neck, trunk, arms, hands, fingers, legs, feet) that were not present on the drawing of a man who exist	Core element	0–7.5
Position, orientation, and perspective (**Figure 3**)	Elements (head, eyes, nose, mouth, nostrils, teeth, ears, hairs, neck, trunk, arms, hands, fingers, legs, feet) and/or the whole drawing different in position, orientation, and perspective from the drawing of a man who exist	Core element and on the whole	0–1.5
Cross-conceptual category (**Figure 5**)	Whole drawing presenting insertions crossed with other conceptual categories (trees, animals, technology, . . .)	On the whole	0–1
Form of the whole (**Figure 5**)	Drawing whose whole form was differently shaped in comparison of the drawing of a man who exist	On the whole	0–1
		TOTAL	0–33.5

a strong inter-reliability agreement between the two raters ($K = 0.91$), showing that even if originality can be quite subjective, the instruction given to adolescents led them to produce original drawings.

Prior to any statistical analysis, a Mauchly's sphericity test ($p = 0.237$) allowed to perform a repeated measures ANOVA, with originality score as a dependent variable, tool (pen on paper, stylus on tablet, finger on tablet) as repeated factors, and grade (1st, 3rd, 6th, or 8th grade) as between-subject factor. In addition, we performed Cochran analyses to investigate whether the occurrence of graphical transformations varied according to the drawing conditions.

RESULTS

A first ANOVA analysis was conducted in order to control for the potential exhausting effect of the procedure, by testing the effect

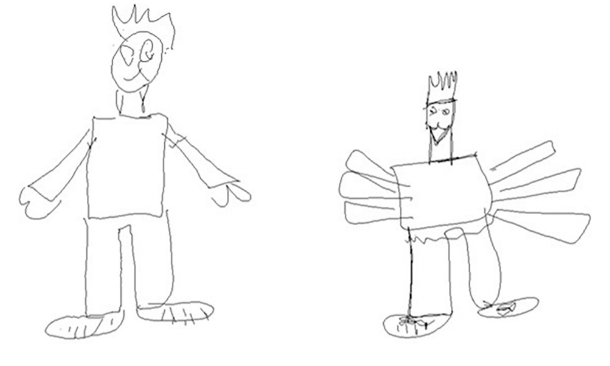


FIGURE 1 | Example of addition and deletion of elements in stylus drawing on tablet at 6th grade: compared to the unoriginal drawing (on the left), arms have been replicated and hands have been deleted in the original drawing (on the right).

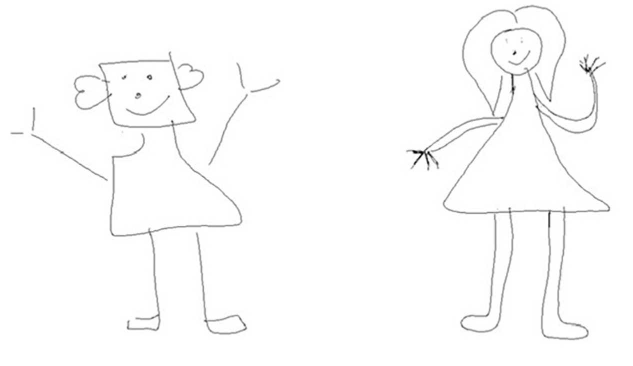


FIGURE 2 | Example of shape modification of elements in stylus drawing on tablet at 6th grade: compared to the unoriginal drawing (on the left), the shape of the head and of the trunk has been modified in the original drawing (on the right).

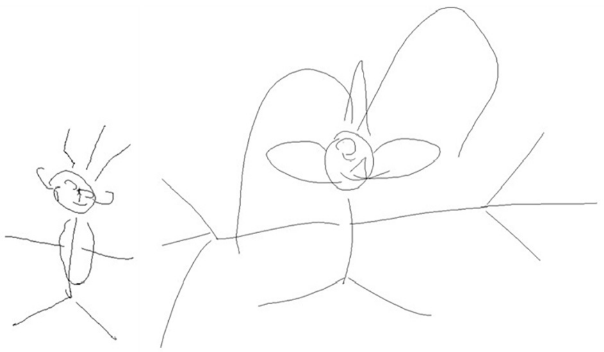


FIGURE 3 | Example of size modifications on elements and orientation modification in finger drawing on tablet at 3rd grade: compared to the unoriginal drawing (on the left), the size of the ears and hairs have been modified, and eyes orientation changed in the original drawing (on the right).

FIGURE 4 | Example of insertion of novel elements in stylus drawing on tablet at 6th grade: compared to the unoriginal drawing (on the left), ears have been added to the original drawing (on the right).

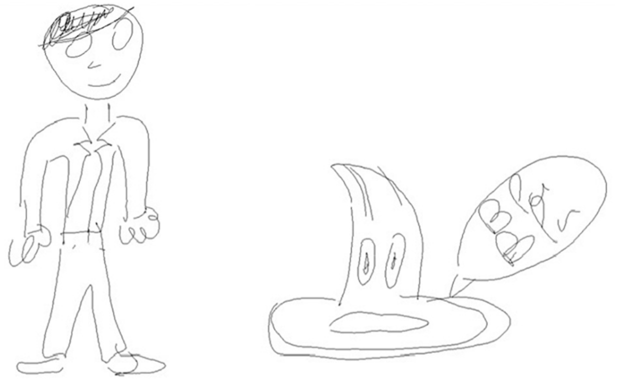


FIGURE 5 | Example of modifications operated on the form of the whole and of cross-conceptual categories insertion in finger drawing on tablet at 6th grade: compared to the unoriginal drawing (on the left), the whole original drawing (on the right) was differently shaped and included cross-category insertion such that the man was represented as a hat.

of the instruction order (original then unoriginal; unoriginal then original) and of the tool order (pen-finger-stylus; pen-stylus-finger; stylus-pen-finger; stylus-finger-pen; finger-stylus-pen; finger-pen-stylus) on originality scores. Results showed no significant effect of instruction order on the originality scores, $F(2,114) = 0.854$, $p = 0.42$ $\eta_p^2 = 0.15$, neither of tool order on originality scores $F(10,114) = 0.223$, $p = 0.98$. These results indicate that systematic counterbalancing was enough to thwart the exhausting factor of producing six drawings in raw. Therefore, instruction order and tool order were then excluded from the main analysis.

We then performed an ANOVA to analyze the effect of varying proprioceptive information with different tools (finger on tablet, stylus on tablet, pen on paper) on originality scores in the 4 groups of children (1st, 3rd, 6th, and 8th grade). Results showed a significant effect of the grade, $F(3,65) = 3.83$, $p = 0.01$ $\eta_p^2 = 0.15$, such that, as shown on **Figure 6**, originality scores increased from the 1st ($M = 2.59$, SD = 1.41), 3rd ($M = 3.23$, SD = 1.55), 6th ($M = 3.65$, SD = 1.30) to the 8th grade ($M = 3.87$, SD = 1.71). *Post hoc* Bonferroni comparisons validated this increasing trend, 8th graders scoring significantly higher than 1st graders ($p = 0.018$).

A significant effect of the tool used on originality scores was also observed, $F(2,130) = 85.43$, $p < 0.001$, $\eta_p^2 = 0.57$. As shown in **Figure 6**, originality scores were higher for drawings produced on tablet with finger ($M = 4.00$, SD = 1.41) and on tablet with stylus ($M = 3.69$, SD = 1.45), compared to drawings produced with pen on paper ($M = 2.30$; SD = 1.24). *Post hoc* Bonferroni comparisons confirmed that finger and stylus were not significant ($p = 0.148$), while significant differences were found between finger and pen ($p < 0.001$) and stylus and pen ($p < 0.001$).

Finally, the ANOVA also revealed an interaction between the tool used and the grade factor, $F(6,130) = 4.11$, $p < 0.001$, $\eta_p^2 = 0.16$, such that, although originality scores increased with the school grade and were higher for drawings produced with finger than with pen, an exception was observed for the 1st grade. As shown in **Figure 6** and **Table 3**, while 3rd, 6th, and 8th graders performed better with fingers ($p < 0.001$) and with stylus ($p < 0.001$) compared to pen on paper, 1st graders performed better only with fingers compared to pen on paper ($p < 0.001$), stylus compared to the pen being non-significant ($p = 0.111$).

In order to analyze the degree to which each item was used to modify the drawn man in each condition (finger, stylus, and pen), a second analysis was conducted using Cochran's Q test. This analysis was aimed at determining whether a drawing condition favored more specifically a category of transformation. First, **Table 4** shows that all elements of the man being confounded, each category of transformation was used in each condition, meaning that each tool is likely to elicit each type of modification. Results also revealed that each type of modification was not equally used. Participants made more addition or deletion [$Q(2) = 35.18$; $p < 0.001$], insertion [$Q(2) = 9.91$; $p = 0.007$], and

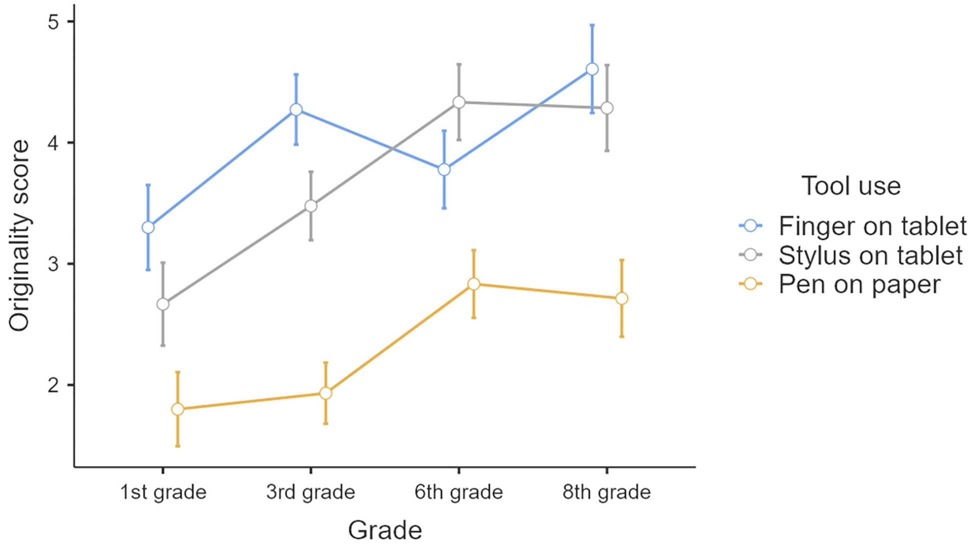


FIGURE 6 | Originality scores (mean and standard errors of mean) according to the tool used and grade.

modified more frequently the shape [$Q(2) = 40.71$; $p < 0.001$] and size [$Q(2) = 17.31$; $p < 0.001$] of elements when drawing with finger or with stylus on tablet, than with a pen on paper. Thus, although each category of transformation was observed whatever the support used, transformations were more used when drawing on a tablet with fingers and stylus than on paper with a pen, all elements confounded.

TABLE 3 | Descriptive analysis of originality scores according to the grade and tool used.

		Finger on tablet	Stylus on tablet	Pen on paper
1st grade	N	15	15	15
	Mean	3.30	2.67	1.80
	Median	3.00	3.00	1.50
	SD	1.37	1.32	1.18
	Range	4.50	4.50	4.50
	Min–Max	1.50–6	0.50–5.00	0.50–5.00
3rd grade	N	22	22	22
	Mean	4.27	3.48	1.93
	Median	4.25	3.50	2.00
	SD	1.34	1.29	1.03
	Range	5.50	5.50	4.50
	Min–Max	2.00–7.50	1.00–6.50	0.00–4.50
6th grade	N	18	18	18
	Mean	3.78	4.33	2.83
	Median	3.75	4.50	2.50
	SD	1.17	1.10	1.21
	Range	4.50	3.50	4.50
	Min–Max	2.00–6.50	2.50–6.00	1.00–5.50
8th grade	N	14	14	14
	Mean	4.61	4.29	2.71
	Median	4.50	4.25	2.50
	SD	1.57	1.63	1.37
	Range	6.00	6.00	4.50
	Min–Max	1.50–7.50	1.00–7.00	0.50–5.00

Then, we performed Cochran's Q test on each category of transformation for each element considered for the drawing of a man (see **Table 4**) according to drawing conditions (finger on tablet, stylus on tablet, and pen on paper). Among the 15 elements constituting the man, shoulders were never modified, and nostrils were very few modified (< 5% in each category). All other elements were modified in the three drawing conditions. However, Cochran's Q test revealed some differences in the degree to which each element was modified according to the tool used. Results showed that nose addition or deletion [$Q(2) = 7.52$; $p = 0.02$], neck addition or deletion [$Q(2) = 12.87$; $p = 0.002$], and finger addition or deletion [$Q(2) = 16.41$; $p < 0.001$] were more used with finger and stylus on tablet than with pen on paper, as well as eyes shape [$Q(2) = 12.87$; $p = 0.002$], trunk shape [$Q(2) = 8.66$; $p = 0.01$], and arms shape [$Q(2) = 16.06$; $p < 0.001$]. In addition, head shape [$Q(2) = 9.66$; $p = 0.008$], neck shape [$Q(2) = 6.72$; $p = 0.03$], and nose insertion [$Q(2) = 7.09$; $p = 0.03$] were more used only by finger on tablet than with pen on paper, and ears addition or deletion was more used with stylus on tablet than with finger on tablet or pen on paper, $Q(2) = 6.20$; $p = 0.04$. The form of the whole was more modified with finger on tablet than with stylus on tablet or pen on paper, $Q(2) = 16.13$; $p < 0.001$.

DISCUSSION

The present study was aimed at determining if the modification of sensory afferences could affect originality in drawings produced by children and adolescents aged 6–14.

First, control analyses proceeded on the originality scale revealing that all categories of modification were used by participants with each tool, all elements confounded. This result

TABLE 4 | Occurrence (percentage) of each item of originality scale as a function of drawing condition (significant differences on Cochran's Q test are in bold).

	Finger	Stylus	Pen
I – Deletion or addition of elements			
0 - All elements confounded	**15.49**	**13.86**	**8.6**
1- Head	5.8	4.35	0
2- Eyes	15.94	14.49	13.04
3- Nose	**21.74**	**28.98**	**11.59**
4- Mouth	20.29	10.14	11.59
5- Nostrils	0	1.45	0
6- Teeth	0	4.35	2.9
7- Ears	**4.35**	**11.6**	**2.9**
8- Hairs	26.09	27.54	23.2
9- Neck	**24.64**	**15.94**	**4.34**
10- Tronk	5.8	4.35	0
11- Shoulders	0	0	0
12- Arms	18.85	18.85	10.14
13- Hands	18.84	15.94	11.59
14- Fingers	**37.68**	**33.33**	**11.59**
15- Legs	23.19	13.04	15.94
16- Feet	25.64	17.4	18.84
II – Shape of elements			
0 – All elements confounded	**22.28**	**21.37**	**14.49**
1- Head	**3.68**	**26.09**	**15.94**
2- Eyes	**59.42**	**50.72**	**31.88**
3- Nose	8.69	14.49	5.8
4- Mouth	30.43	40.58	33.33
5- Nostrils	0	0	0
6- Teeth	0	1.45	1.45
7- Ears	0	2.9	1.45
8- Hairs	43.48	43.48	34.8
9- Neck	**15.94**	**10.14**	**5.8**
10- Trunk	**53.62**	**47.83**	**33.33**
11- Shoulders	0	0	0
12- Arms	**34.8**	**36.23**	**11.6**
13- Hands	4.35	5.8	1.45
14- Fingers	10.14	10.14	11.6
15- Legs	39.13	33.33	23.18
16- Feet	18.85	18.85	20.29
III – Size of elements			
0 – All elements confounded	**4.35**	**3.89**	**1.53**
1- Head	13.04	13.04	5.8
2- Eyes	1.45	2.9	0
3- Nose	4.35	1.45	0
4- Mouth	1.45	1.45	0
5- Nostrils	0	0	0
6- Teeth	0	0	0
7- Ears	1.45	1.45	1.45
8- Hairs	1.45	0	0
9- Neck	4.35	8.7	2.9
10- Tronk	11.59	4.35	4.35
11- Shoulders	0	0	0
12- Arms	13.04	10.14	4.35
13- Hands	1.45	1.45	1.45
14- Fingers	2.9	1.45	1.45
15- Legs	11.6	11.6	2.9
16- Feet	1.45	4.35	0
IV – Insertion of new elements			
0 – All elements confounded	**5.34**	**5.8**	**3.26**
1- Head	7.25	10.14	5.8
2- Eyes	1.45	1.45	1.45
3- Nose	**11.59**	**4.35**	**1.45**
4- Mouth	1.45	7.25	4.35
5- Nostrils	1.45	1.45	0
6- Teeth	5.8	1.45	5.8
7- Ears	**13.04**	**10.14**	**2.9**
8- Hairs	14.5	5.8	8.7
9- Neck	4.35	7.25	0
10- Tronk	**0**	**8.7**	**4.35**
11- Shoulders	0	0	0
12- Arms	1.45	2.9	5.8
13- Hands	5.8	8.7	2.9
14- Fingers	5.8	8.7	4.35
15- Legs	7.25	7.25	1.45
16- Feet	4.35	7.25	2.9
V – Position, orientation, and perspective modifications			
1- On the whole	0	0	0
2- On elements	2,9	0	1,45
VI – Cross-conceptual categories modification			
On the whole	4,35	4,35	2,9
VII – Form of the whole			
On the whole	**18,84**	**2,9**	**2,9**

means that the scale allowed for assessment modifications operated with each tool used, and thus didn't favor the scoring of originality in a drawing condition over another. Therefore, items used to assess originality were modified as well on the tablet with fingers or stylus as on paper with a pen. This analysis also revealed that modifications were higher on the tablet with fingers and stylus than with pen on paper. These higher originality scores on the tablet with finger or stylus are not tied to one specific category or element but reflect the overall occurrence of original modifications that were generally lower in pen on paper condition than with finger on the tablet or with stylus on the tablet. Thus, varying sensory afferences with the use of tablets seem to have led to globally more modifications.

If creativity is related to sensory afferences, then it was expected that drawing with fingers on the tablet would increase originality at all ages, compared to drawings made with pen on paper. Our first hypothesis stated that increasing proprioceptive feedback on fingertips could enhance motor control, resulting in a reduced cognitive load allocated to motor control in favor of the drawing task. Thus, we expected that enhancing proprioceptive feedback on fingertips when drawing on the tablet with fingers would increase originality at all ages in comparison to the use of a pen on paper.

The results from the present study confirm our first hypothesis, such that increasing haptic afferences to fingertips increased originality. As a matter of fact, children and adolescents produced more original drawings when drawing with fingers on the tablet rather than with pen on paper as revealed by higher originality scores obtained from drawing with fingers on the tablet. This result observed in adolescents up to 14 years extends the previous ones obtained in children aged 5–6 and 7–8 years demonstrating that drawing with fingers on the tablet led to more originality in drawings compared to the use of pen on paper (Bitu et al., 2019). Using fingers to draw on the tablet increased the friction between finger and screen leading in turn to an increase in the amount of available proprioceptive information. Drawing with fingers leads to an enrichment of sensory feedback to fingertips that, in turn, enhances the creative process during a drawing task from 1st to 8th grade. This effect was observed both before and after 7 years old: increasing proprioceptive afferences enhanced the originality of children aged 7 years and more, who are able to make appropriate use of proprioceptive feedback to correct their ongoing movement (Holst-Wolf et al., 2016), but also of younger ones who do not accurately use proprioceptive feedback for online control of movements (Bairstow and Laszlo, 1981; Laszlo and Bairstow, 1984). Consequently, it could be suggested that amplifying proprioceptive information may have facilitated children's creative process by reducing the cognitive load allocated to the motor control (intrinsic load), in favor of more cognitive resources for the creative process (essential load).

Conversely, using a stylus on the tablet is known to induce a sliding effect which decreases the friction between the stylus and the screen, leading in turn to a decrease in available proprioceptive afferences (Alamargot and Morin, 2015). This decrease in available proprioceptive afference could impair the cognitive process involved in the task, by increasing the intrinsic load with resources that can't be allocated for the essential processing, i.e., producing original drawing. Our second hypothesis stated that using a stylus on tablet would decrease originality in drawings at all ages as a consequence of the higher cognitive load induced by this situation. Thus, we expected that reducing proprioceptive feedback when drawing with a stylus on tablet would decrease originality at all ages in comparison to the use of a pen on paper. This second hypothesis was not validated. We observed that 1st graders performance when drawing with a stylus on tablet were similar to pen on paper. Lowering the available proprioceptive feedback in a drawing task did not affect 1st graders, who are known to struggle with the use of accurate proprioceptive information from the use of stylus on the tablet. It could be suggested that these young children may have used compensating strategies based on a greater mobilization of the visual component in accordance with empirical data reporting that young children are known to rely more heavily on visual rather than proprioceptive information during an action (Bard and Hay, 1983; Contreras-Vidal, 2006). In accordance with results reported by Guilbert et al. (2019) with 2nd graders, using a tablet with stylus may have enhanced visual information compared to a pen on paper, leading young children to pay more attention to the tracing that could, in turn, lead to a more intense analysis of the traced shapes and increase creativity. However, as suggested by cognitive load theory, paying more attention to the gesture should have deteriorated the cognitive resources allowed for the original task.

On the contrary, from 3rd to 8th grade, children and adolescents obtained higher originality scores when drawing with a stylus than when drawing with a pen. By 8/9 years, children were affected by the use of a stylus on a tablet, but in a positive way, such that it strengthened their originality performance. This unexpected effect could be explained as a consequence of more accurate use of afferent proprioceptive information by 8/9 years, which, along with visual information, leads to a significant improvement in predictive motor control (von Hofsten and Rösblad, 1988; Chicoine et al., 1992). As reminded in the introductive section, this improvement leads to a switch from a feedback motor control where children make pauses to control their movement with visual verification, to a feedforward motor control relying more accurately on the movement prediction. Consequently, by 8/9 years, the sensory prediction could be used to compensate for the loss of proprioceptive afferences. However, if 3rd to 8th grade children and adolescents had compensated for the initial loss of proprioceptive information with the use of feedforward motor control to make sensory predictions, this compensating strategy should have, according to the cognitive load theory (Sweller et al., 2011), increased the intrinsic load leading to a reduced cognitive resource allocated to the original task. The opposite was observed.

Altogether, the obtained results revealed that originality scores varied according to the proprioceptive afferences children and adolescents had to deal with during the drawing task. Results obtained with the stylus revealed a modulation of the tool on originality according to the age of the participants that do not fit in the hypotheses formulated with the cognitive load framework. Why did children and adolescents from 3rd grade and older draw more original drawings with stylus on tablet compared to pen on paper?

The embodied perspective of creativity could bring some arguments in the understanding of this increased performance induced by the use of a stylus in a creative drawing task. By enhancing proprioceptive information used in the prediction process, the feedforward motor control may have supported prediction possibilities involved in creativity (Dietrich and Haider, 2015). Indeed, the present results could be interpreted according to the theoretical model of creativity recently proposed by Dietrich and Haider (2015) who conceived creativity as an abstract form of sensorimotor prediction run by internal models of motor control of action (Wolpert et al., 1995; Miall and Wolpert, 1996; Kawato, 1999; Wolpert and Ghahramani, 2004). In the case of creativity, internal models would allow to generation and evaluate new and original ideas, on a trial-and-error basis, by chaining multiple internal model loops. In each of these loops, direct models would allow the generation of various ideas to reach a creative goal sent by the efference copy from inverse models. Generated ideas would be compared to the creative goal and, when they are not enough new or adapted to the context, are used as a source of information for a new loop iteration until the comparator selects the most original and adapted idea. The creative process would thus be a predictive mechanism rooted in sensorimotor activity in which sensory consequences of motor commands play a fundamental role. As a matter of fact, sensory inputs feed internal models to validate or invalidate the selected motor programs, by comparing sensory afferences available in the task to the sensory prediction. Applied to creativity, sensory afferences used in internal models would allow for validation or invalidation of each variation of ideas, by comparing sensory afferences available in the task, to the sensory prediction. Sensory afferences would thus feed multiple iterations of the internal model loops, allowing to generate of idea variations, and selecting the one the more adapted to the context. In this way, creativity would be rooted, as suggested by Dietrich and Haider (2015), in sensorimotor control of the action. Consequently, modifying sensory feedback with finger and stylus may have impacted the mobilization of the originality process: the sensory afferences specifically generated in a given action space (stylus on tablet or finger on the tablet) may have modulated originality by delimiting the range of possible outcomes, as well in generating as in selecting ideas. As reminded in the section "Introduction," up to 7/8 years, children felt some difficulties in using accurate proprioceptive afferences. By 8/9 years, they demonstrated an improvement in processing proprioceptive information which led to an increase in the use of feedforward motor control strategy to maximize sensory input (Guilbert et al., 2019). This improvement could be the reason why children's originality in stylus on tablet drawing was significantly higher than with pen on paper.

The present research revealed the last result that was not expected and could be in line with the previous idea. Whether drawings were made on the tablet or paper, originality scores among age groups showed an increasingly linear trend. This result partially confirms the first observation of Torrance (1968) who described creativity development as following a linear trend. However, Torrance (1968) added that 3 slump periods accompanied this positive progression, identified at 5 years, 9–10 years, and 13–14 years. He suggested that these three slump periods would be linked to changes occurring during the scholar course, such that children have to comply with the new scholar normative environment in which they are enrolled (Torrance, 1968). Contrary to Torrance, we did not observe these slumps. This could be explained according to the age of the participants (6–7 years, 8–9 years, 11–12 years, and 13–14 years) which are critical periods identified in the development of predictive motor control. As mentioned in the section "Introduction," these age groups correspond to periods before (1st grade), during (3rd and 6th grade), and after (8th grade) the transition concerning the use of internal models of action. From the present result, we reported and in regard to the several theories discussed, it could be suggested that the development of creativity overlaps the development of predictive motor control. Future research should therefore focus on creative development through the lens of predictive motor control, for example by tracking kinematics changes in comparison to creative assessment throughout childhood.

Several limitations can be noted in our study. First, we did not consider participants' daily use of technologies, which could have favored some children over others in the use of tablets in comparison to paper. However, Bitu et al. (2019) addressed a survey of parents with the aim of controlling this variable and found no effect of technology uses on originality gain in a drawing task on the tablet. Moreover, interacting with tablets involves only a few formal knowledge and simple gestures (Dubé and McEwen, 2015). For this reason, we used a functional task to make sure participants were able to use the basic gestures needed for the interaction on the tablet. Second, several factors pointed out by the multivariate approach to creativity (Lubart et al., 2003) could have played a role in explaining our results. More specifically, motivation may have been a confounding factor in the present study which could explain part of the difference obtained in originality scores between tablet and paper. Indeed, for children and adolescents, using tablets induce better motivation and engagement in the task (Clark and Luckin, 2013; Dubé and McEwen, 2015). More specifically, the easiness of use of tablet screens and the physical interaction consecutive to the tactile feature intensely engage children and adolescents in a multimodal task, leading to a better motivation (Amadieu and Tricot, 2014; Crescenzi et al., 2014; Geist, 2014; Dubé and McEwen, 2015). Further studies should isolate this component to determine the part of benefits due to motivation and engagement induced by the tools used.

To conclude, this study contributes to a better understanding of the process of creativity. This first empirical study highlights the crucial role played by sensory afferences in creative thinking. Moreover, it opens new avenues on the understanding of creativity that could be investigated as a predictive process rooted in predictive motor control as suggested by Dietrich and Haider (2015). Future studies should investigate this new issue.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

FB: conception and design of the work, acquisition, analysis and interpretation of data for the work, and drafting and revising the manuscript. BG-M: conception and design of the work and critical advices. MM: conception and design of the work, interpretation of data for the work, and drafting and revising the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

Alamargot, D., and Morin, M. F. (2015). Does handwriting on a tablet screen affect students' graphomotor execution? A comparison between grades two and nine. *Human Movem. Sci.* 44, 32–41. doi: 10.1016/j.humov.2015.08.011

Amadieu, F., and Tricot, A. (2014). *Apprendre avec le numérique: mythes et réalités*. Paris: Retz. doi: 10.14375/NP.9782725633206

Bairstow, P. J., and Laszlo, J. I. (1981). Kinaesthetic sensitivity to passive movements and its relationship to motor development and motor control. *Dev. Med. Child Neurol.* 23, 606–616. doi: 10.1111/j.1469-8749.1981.tb02042.x

Bara, F., and Gentaz, E. (2011). Haptics in teaching handwriting: the role of perceptual and visuo-motor skills. *Human Movem Sci.* 30, 745–759. doi: 10.1016/j.humov.2010.05.015

Bara, F., and Tricot, A. (2017). Le rôle du corps dans les apprentissages symboliques: apports des théories de la cognition incarnée et de la charge cognitive. *Recherches sur la Philosophie et le Langage* 33, 219–249.

Bard, C., and Hay, L. (1983). Etude ontogénétique de la coordination visuo-manuelle. *Can. J. Psychol.* 37:390. doi: 10.1037/h0080742

Barron, F. (1988). Putting creativity to work. *Nat. Creat.* 1988, 76–98.

Bashwiner, D. (2018). *The neuroscience of musical creativity. The Cambridge Handbook of the Neuroscience of Creativity*. Cambridge: Cambridge University Press. 495–516. doi: 10.1017/9781316556238.029

Bashwiner, D., and Bacon, D. (2019). Musical creativity and the motor system. *Curr. Opin. Behav. Sci.* 27, 146–153. doi: 10.1016/j.cobeha.2018.12.005

Bitu, F., Jouen, F., and Molina, M. (2019). Les apports de l'interface tactile numérique dans l'expression créative graphique : une étude exploratoire chez des enfants d'âge scolaire. *Enfance* 3, 357–373. doi: 10.3917/enf2.193.0357

Brown, M. C., McNeil, N. M., and Glenberg, A. M. (2009). Using concreteness in education: real problems, potential solutions. *Child Dev. Persp.* 3, 160–164. doi: 10.1111/j.1750-8606.2009.00098.x

Caeyenberghs, K., Wilson, P. H., VanRoon, D., Swinnen, S. P., and SmitsEngelsman, B. C. (2009). Increasing convergence between imagined and executed movement across development: evidence for the emergence of movement representations. *Dev. Sci.* 12, 474–483. doi: 10.1080/87565640802499183

Chicoine, A. J., Lassonde, M., and Proteau, L. (1992). Developmental aspects of sensorimotor integration. *Dev. Neuropsychol.* 8, 381–394. doi: 10.1080/87565649209540533

Clark, W., and Luckin, R. (2013). *iPads in the Classroom. What The Research Says*. London: London Knowledge Lab

Contreras-Vidal, J. L. (2006). Development of forward models for hand localization and movement control in 6-to 10-year-old children. *Human Movem. Sci.* 25, 634–645. doi: 10.1016/j.humov.2006.07.006

Cooper, L. Z. (2005). Developmentally appropriate digital environments for young children. *Lib. Trends* 54, 286–302. doi: 10.1353/lib.2006.0014

Crescenzi, L., Jewitt, C., and Price, S. (2014). The role of touch in preschool children's learning using iPad versus paper interaction. *Aust. J. Lang. Liter.* 37, 86–95.

Dietrich, A., and Haider, H. (2015). Human creativity, evolutionary algorithms, and predictive representations: the mechanics of thought trials. *Psychonomic. Bull. Rev.* 22, 897–915. doi: 10.3758/s13423-014-0743-x

Dubé, A. K., and McEwen, R. N. (2015). Do gestures matter? The implications of using touchscreen devices in mathematics instruction. *Learn. Instruct.* 40, 89–98. doi: 10.1016/j.learninstruc.2015.09.002

Duff, W. (1767). *An essay on original genius: and its various modes of exertion in philosophy and the fine arts, particularly in poetry*. Southill: Edward and Charles Dilly.

Fleury, S., Agnes, A., Vanukuru, R., Goumillout, E., Delcombel, N., and Richir, S. (2020). Studying the effects of visual movement on creativity. *Think. Skills Creat.* 36:100661. doi: 10.1016/j.tsc.2020.100661

Gaggioli, A., Falletta, E. M., Ferrise, F., Graziosi, S., Gallace, A., DAusilio, A., et al. (2019). Effects of interpersonal sensorimotor synchronization on dyadic creativity: gender matters. *Front. Psychol.* 9:2604. doi: 10.3389/fpsyg.2018.02604

Galton, F. (1879). Psychometric experiments. *Brain* 2, 149–162. doi: 10.1093/brain/2.2.149

Galton, F. (1883). *Inquiries into human faculty and its development*. Basingstoke: Macmillan. doi: 10.1037/14178-000

Geist, E. (2014). Using tablet computers with toddlers and young preschoolers. *Young Child*. 69, 58–62.

Geist, E. A. (2012). A qualitative examination of two year-olds interaction with tablet based interactive technology. *J. Instruct. Psychol.* 39, 26–35.

Gerth, S., Klassert, A., Dolk, T., Fliesser, M., Fischer, M. H., Nottbusch, G., et al. (2016). Is handwriting performance affected by the writing surface? Comparing preschoolers', second graders', and adults' writing performance on a tablet vs. Paper. *Front. Psychol.* 7:1308. doi: 10.3389/fpsyg.2016.01308

Goodenough, F. L. (1926). *Measurement of intelligence by drawings*. New York, NY: Harcourt.

Guilbert, J., Alamargot, D., and Morin, M. F. (2019). Handwriting on a tablet screen: role of visual and proprioceptive feedback in the control of movement by children and adults. *Human Movem. Sci.* 65, 30–41. doi: 10.1016/j.humov.2018.09.001

Guilbert, J., Jouen, F., and Molina, M. (2018). Motor imagery development and proprioceptive integration: which sensory reweighting during childhood? *J. Exp. Child Psychol.* 166, 621–634. doi: 10.1016/j.jecp.2017.09.023

Guilford, J. P. (1956). The structure of intellect. *Psychol. Bull.* 53:267. doi: 10.1037/h0040755

Guilford, J. P. (1967). Creativity: yesterday, today and tomorrow. *J. Creat. Behav.* 1, 3–14. doi: 10.1002/j.2162-6057.1967.tb00002.x

Hay, L. (1978). Accuracy of children on an open-loop pointing task. *Percep. Motor Skills* 47, 1079–1082. doi: 10.2466/pms.1978.47.3f.1079

Hay, L. (1979). Spatial-temporal analysis of movements in children: motor programs versus feedback in the development of reaching. *J. Motor Behav.* 11, 189–200. doi: 10.1080/00222895.1979.10735187

Holst-Wolf, J. M., Yeh, I., and Konczak, J. (2016). Development of proprioceptive acuity in typically developing children: normative data on forearm position sense. *Front. Human Neurosci.* 10:436. doi: 10.3389/fnhum.2016.00436

James, W. (1880). *The feeling of effort*. Whitefish: Kessinger Publishing.

Jeannerod, M. (1994). The representing brain: neural correlates of motor intention and imagery. *Behav. Brain Sci.* 17, 187–201. doi: 10.1017/S0140525X00034026

Kandel, S., and Perret, C. (2015). How do movements to produce letters become automatic during writing acquisition? Investigating the development of motor anticipation. *Int. J. Behav. Dev.* 39, 113–120. doi: 10.1177/0165025414557532

Karmiloff-Smith, A. (1990). Constraints on representational change: evidence from children's drawing. *Cognition* 34, 57-83. doi: 10.1016/0010-0277(90)90031-E

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* 9, 718–727. doi: 10.1016/S0959-4388(99)00028-8

Korkman, M., Kirk, U., and Kemp, S. (2014). *NEPSY-II*. Madrid: Pearson.

Laszlo, J. I., and Bairstow, P. J. (1984). Handwriting: difficulties and possible solutions. *Sch. Psychol. Int.* 5, 207–213. doi: 10.1177/0143034384054004

Lubart, T. I., and Georgsdottir, A. S. (2004). Créativité, haut potentiel et talent. *Psychologie Française* 49, 277–291. doi: 10.1016/S0033-2984(04)00047-0

Lubart, T., Mouchiroud, C., Tordjam, S., and Zenasni, F. (2003). *Psychologie de la créativité*. Paris: Armand Colin.

MacKinnon, D. W. (1962). The nature and nurture of creative talent. *Am. Psychol.* 17:484. doi: 10.1037/h0046541

Mayer, R. E. (2005). *Cognitive theory of multimedia learning. The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511816819.004

Mayer, R. E., and Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Edu. Psychol.* 38, 43–52. doi: 10.1207/S15326985EP3801_6

McManis, L. D., and Gunnewig, S. B. (2012). Finding the education in educational technology with early learners. *Young Child.* 67, 14–24.

Miall, R. C., and Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networ.* 9, 1265–1279. doi: 10.1016/S0893-6080(96)00035-4

Molina, M., Tijus, C., and Jouen, F. (2008). The emergence of motor imagery in children. *J. Exp. Child Psychol.* 99, 196–209. doi: 10.1016/j.jecp.2007.10.001

Mouchiroud, C., and Lubart, T. I. (2001). Children's original thinking : An empirical examination of alternative measures derived from divergent thinking tasks. *J. Genet. Psychol.* 162, 382–401. doi: 10.1080/00221320109597491

Nikolaeva, E. I., Novikova, A. V., and Vergunov, E. G. (2018). The correlation between intelligence, creativity and the parameters of sensorimotor integration in children of different ages. *Psychol. Russia State Art* 11, 68–80. doi: 10.11621/pir.2018.0206

Ochse, R. (1990). *Before the gates of excellence: The determinants of creative genius*. Cambridge: Cambridge University Press.

Patchan, M. M., and Puranik, C. S. (2016). Using tablet computers to teach preschool children to write letters: exploring the impact of extrinsic and intrinsic feedback. *Comp. Edu.* 102, 128–137. doi: 10.1016/j.compedu.2016.07.007

Piotrowski, J. T., and Meester, L. (2018). Can apps support creativity in middle childhood? *Comp. Human Behav.* 85, 23–33. doi: 10.1016/j.chb.2018.03.030

Runco, M. A. (1991). The evaluative, valuative, and divergent thinking of children. *J. Creat. Behav.* 25, 311–319. doi: 10.1002/j.2162-6057.1991.tb01143.x

Sakr, M. (2018). Multimodal participation frameworks during young children's collaborative drawing on paper and on the iPad. *Think. Skills Creat.* 29, 1–11. doi: 10.1016/j.tsc.2018.05.004

Smits-Engelsman, B. C., and Wilson, P. H. (2013). Age-related changes in motor imagery from early childhood to adulthood: probing the internal representation of speed-accuracy trade-offs. *Human Movem. Sci.* 32, 1151–1162. doi: 10.1016/j.humov.2012.06.006

Sternberg, R. J. (1988). A three-facet model of creativity. *Nat. Creat.* 1988, 125–147.

Sternberg, R. J., and Lubart, T. I. (1991). An investment theory of creativity and its development. *Human Dev.* 34, 1–31. doi: 10.1159/000277029

Sternberg, R. J., and Lubart, T. I. (1995). *Defying the crowd: Cultivating creativity in a culture of conformity*. New York, NY: Free Press.

Sternberg, R. J., and Lubart, T. I. (1996). Investing in creativity. *Am. Psychol.* 51:677. doi: 10.1037/0003-066X.51.7.677

Sweller, J., Ayres, P., and Kalyuga, S. (2011). *Cognitive load theory*. New-York, NY: Springer. doi: 10.1007/978-1-4419-8126-4

Torrance, E. P. (1968). A longitudinal examination of the fourth grade slump in creativity. *Gifted Child Quart.* 12, 195–199. doi: 10.1177/001698626801200401

Villamor, C., Willis, D., and Wroblewski, L. (2010). *Touch gesture reference guide*. Silicon Valley: LukeW.

von Hofsten, C., and Rösblad, B. (1988). The integration of sensory information in the development of precise manual pointing. *Neuropsychologia* 26, 805–821. doi: 10.1016/0028-3932(88)90051-6

Wolpert, D. M., and Ghahramani, Z. (2004). Computational motor control. *Science* 269, 1880–1882.

Wolpert, D. M., and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Network.* 11, 1317–1329. doi: 10.1016/S0893-6080(98)00066-5

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882. doi: 10.1126/science.7569931

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

REVIEW
published: 26 July 2022
doi: 10.3389/fpsyg.2022.899718

A Synthetic Review of Cognitive Load in Distance Interpreting: Toward an Explanatory Model

Xuelian Zhu[1,2] and Vahid Aryadoust[2]*

[1]Sichuan International Studies University, Chongqing, China, [2]National Institute of Education, Nanyang Technological University, Singapore, Singapore

Distance Interpreting (DI) is a form of technology-mediated interpreting which has gained traction due to the high demand for multilingual conferences, live-streaming programs, and public service sectors. The current study synthesized the DI literature to build a framework that represents the construct and measurement of cognitive load in DI. Two major areas of research were identified, i.e., causal factors and methods of measuring cognitive load. A number of causal factors that can induce change in cognitive load in DI were identified and reviewed. These included factors derived from tasks (e.g., mode of presentation), environment (e.g., booth type), and interpreters (e.g., technology awareness). In addition, four methods for measuring cognitive load in DI were identified and surveyed: subjective methods, performance methods, analytical methods, and psycho-physiological methods. Together, the causal factors and measurement methods provide a multifarious approach to delineating and quantifying cognitive load in DI. This multidimensional framework can be applied as a tool for pedagogical design in interpreting programs at both the undergraduate and graduate levels. It can also provide implications for other fields of educational psychology and language learning and assessment.

Keywords: cognitive load, causal factors, distance interpreting, measurement methods, pedagogical design

Edited by:
Hong Xu,
Nanyang Technological University,
Singapore

Reviewed by:
Debatri Chatterjee,
Tata Consultancy Services, India
Dechao Li,
Hong Kong Polytechnic University,
Hong Kong SAR, China
Campbell McDermid,
Gallaudet University, United States

***Correspondence:**
Xuelian Zhu
zhu_xuelian@qq.com

Specialty section:
This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: *19 March 2022*
Accepted: *30 May 2022*
Published: *26 July 2022*

Citation:
Zhu X and Aryadoust V (2022) A Synthetic Review of Cognitive Load in Distance Interpreting: Toward an Explanatory Model.
Front. Psychol. 13:899718.
doi: 10.3389/fpsyg.2022.899718

INTRODUCTION

Distance Interpreting (DI) refers to interpreting services provided by the interpreters who are geographically separate from clients and can only communicate through telephone calls or video links (Braun, 2020; AIIC, 2021). DI and onsite interpreting might share the same working mode, meaning that the interpreters listen and comprehend the source language and produce the target language either consecutively or simultaneously, but distance interpreters do so at different locations relative to the participants and with different technology requirements (Azarmina and Wallace, 2005; Braun and Taylor, 2012). The idea of distance interpreting meets with considerable support from the interpreting industry, represented by the AIIC's endorsement. In its position document, AIIC states:

> AIIC recognises that ICTs enable new interpreting modalities. These include setups whereby interpreters have no direct view of speakers/signers, but rather an indirect, ICT-enabled audio/audiovisual feed of speakers/signers who are not in the same physical location as

> interpreters, as well as setups where interpreters within the same team and even booth may be at different locations (AIIC, 2022).

Although the industry has recognized the position of interpreting in distance mode, studies have yet to comprehensively investigate this field. This is partly because distance interpreting was first applied in hospital and courtroom settings, in which consecutive mode is more widely applied; since DI requires constant turn taking to confirm information, the scope of research on DI is limited (Gracia-García, 2002; Jones et al., 2003; Braun, 2013). Another reason for the limitation of DI research might be a deficiency in technical support in the previous generation of DI, like limited bandwidth and video feed definition, which thoroughly hindered the wider spread of DI (Moser-Mercer, 2003; Ozolins, 2011). Currently, DI in simultaneous mode is commonly used with the support of technology and extended to live-streaming of videogames, sports narrations, and multilingual conferences that demand timely interpretation of different languages (Braun, 2020). These technologies have been particularly helpful during the COVID-19 pandemic, in which the need for DI services in both public service and conference settings was significantly heightened (Runcieman, 2020; Ait Ammour, 2021).

One of the recognized issues in DI is the role of cognitive load in performance (Moser-Mercer, 2005a,b; Mouzourakis, 2006), whose underlying mechanisms and effects have been examined in the published literature. Notably, Moser-Mercer (2005a) mentioned that fatigue during distance interpreting may be a "consequence of allocating additional cognitive resources" (p. 1). In line with this postulation, later studies investigated visual ecologies in DI to understand how and what visuals should be presented to interpreters during DI, as the visuals are additional input that may ensue change in cognitive load (Mouzourakis, 2006; Licoppe and Veyrier, 2017; Plevoets and Defrancq, 2018). In addition, Wessling and Shaw (2014) pointed out that the emotional state of distance interpreters, partly occasioned by cognitive load, might influence "longevity" in the field. However, these findings are tentative and inconclusive. Notably, there is no general framework to theorize and measure the extent to which different factors instigate change in cognitive load of DI interpreters. Understanding these factors, particularly the mode-specific nature of DI, is key to understanding whether the distance mode of working has a positive or adverse effect on the cognitive load of interpreters. To bridge this gap in knowledge, we review the causal factors inducing change in cognitive load in DI and methods of quantifying and measuring change in cognitive load.

In what follows, we will present a brief review of the scope of cognitive load theory, identify the causal factors inducing change in distant interpreters' cognitive load, align these factors with pertinent measurement methods, and finally discuss the implications of the study for research. It should be noted that the study is focused on spoken language interpreting and excludes sign language interpreting.

COGNITIVE LOAD THEORY AND INTERPRETING

Cognitive load (also known as mental workload, mental load, or mental effort) has drawn research interest from scholars in diverse disciplines (Sweller et al., 2011; Ayres et al., 2021; see also chapters in Zheng, 2018). Cognitive load is defined as the mental workload imposed on a performer when executing a particular task (Yin et al., 2008; Sweller et al., 2011; Sweller, 2018). Numerous studies have attempted to determine the constituents of cognitive load and how its components interact with each other (e.g., Moray, 1967; Hart and Staveland, 1988; Meshkati, 1983; unpublished Doctoral dissertation[1], 1988; Young and Stanton, 2005; Pretorius and Cilliers, 2007; Byrne, 2013; Young et al., 2015; Kalyuga and Plass, 2018; Schnaubert and Schneider, 2022). Generally, these studies consider cognitive load as a multidimensional construct comprising several fundamental aspects, like tasks, operators, and context (Martin, 2018; Paas and van Merriënboer, 2020). Among them, cognitive load theory (CLT; Sweller et al., 2011) has offered important insights on the role of working memory, types of cognitive load, and the role of individual characteristics in cognitive tasks. In this theory, cognitive load consists of the *mental load* engendered by the task and environment factors and the *mental effort* or the cognitive resources allocated by the task performer to deal with task demands (Meshkati, 1988; Paas and Van Merriënboer, 1994; Sweller et al., 1998, 2019; Yin et al., 2008).

CLT was first introduced in the field of learning and instruction in the 1980s. The theory posits that new information (perceived stimulus) is first processed by working memory (WM) and then stored in long-term memory for future use (Sweller et al., 2011). In addition, in CLT, WM is postulated to have limited capacity, as visual and auditory channels compete for resources, while long-term memory is arguably limitless (Sweller, 1988). Due to the limited capacity of WM, it is crucial to maintain cognitive load at a manageable level to sustain productivity. Since WM is integral to the process of interpreting, it has been extensively discussed in the interpreting literature by many scholars (e.g., Moser-Mercer et al., 2000; Christoffels and De Groot, 2009; Seeber, 2011, 2013; Chmiel, 2018; Dong et al., 2018; Mellinger and Hanson, 2019; Wen and Dong, 2019; Bae and Jeong, 2021).

According to Sweller et al.'s (2011) tripartite model, there are three types of cognitive load: intrinsic, extraneous, and germane. Intrinsic cognitive load refers to the inherent difficulty of the information and the interactivity of the characteristics of the input; accordingly, task complexity depends on the nature and content of the information and the skills of the person who processes the information (Leppink et al., 2013). Extraneous cognitive load, on the other hand, is generated by the manner in which information is presented and whatever the learner (processor) is required to do and as such, it is under the control of task designers (Cierniak et al., 2009). In addition, germane cognitive load is required for learning, processing

---

[1] Meshkati, N. (1983). *A conceptual model for the assessment of mental workload based upon individual decision styles.* [Unpublished Doctoral dissertation]. University of Southern California.

and (re)constructing information; it can compete with and occupy the WM resources that help with processing the intrinsic cognitive load (Paas and van Merriënboer, 2020). In instruction designs, it is recommended to limit extraneous cognitive load while promoting germane load so as to direct the learner's attention to the cognitive processes that are relevant to the processing of key information (van Merriënboer et al., 2006). The same limit to extraneous cognitive load applies in interpreting. For example, a better booth design that blocks out environmental noise would provide a better venue for interpreters as it lowers the extraneous cognitive load caused by noise and allows interpreters to allocate their cognitive resources to the processing of intrinsic and germane cognitive loads.

Early models of cognitive load in interpreting studies were mainly drawn from conceptual discussions and, thus were backed up by little supporting data (Setton, 1999). In a novel theory for its time, Gerver (1975) argued that information is processed during simultaneous interpreting (SI) through "a buffer storage" (p. 127), which is separate for the source and target languages. Although there is yet no empirical support for this hypothesis, this view of storage aligns with the idea that informational sources can be processed in parallel (Timarová, 2008). Successively, Moser (1978) developed a process model of interpreting that placed generated abstract memory (GAM) at the center of discussion. She proposed that GAM is the equivalent of short-term memory, which was later reconceptualized as WM by Baddeley and Hitch (1974) and Baddeley (1986, 2000, 2012). Kirchhoff (1976) described cases in which the completion of tasks requires more processing capacity than is available to the interpreter.

In a similar vein, Gile's (1995, 2009) effort model, Gile's (1999) tightrope hypothesis, and Seeber's (2011) cognitive load model all underscored the role of various factors in successful SI and had a strong impact on the interpreting research (Pöchhacker, 2016). Gile's (1995, 2009) effort model is an operation model based on the theoretical assumption of limited attentional resources. The model assimilates interpreters to a tightrope walker who has to utilize nearly all their mental effort during interpreting, which is available only in limited supply. The interpreting process comprises three core efforts: comprehension, production and short-term memory. Comprehension is the process of perceiving and understanding the input, while production involves the articulation of the translated code. Short-term memory (STM) is an interpreter's capacity to tentatively store limited bits of information. STM capacity is indicated, among other things, by ear-voice-span (EVS) during interpreting, which refers to the time lag between comprehension and production during which interpreters make decisions about their interpreting (Christoffels and De Groot, 2004; Chen, 2018; Collard and Defrancq, 2019; Seeber et al., 2020). Gile's effort model provides a reliable representation of interpreting which may be useful in practice and pedagogical design in interpreting training programs. However, Seeber (2011) argues that Gile's effort model, which is based on Kahneman's (1973) single resource theory, assumes that interpreters draw resources from undifferentiated pools, so it is unable to identify interferences of subtasks. Built on this argument, Seeber's (2011) cognitive load model is founded upon multiple resource theory (Wickens, 1984, 2002); it recognized and accounts for the conflict and overlap between language comprehension and production during interpreting. Seeber (2011) used this analytical framework to illustrate how the overall demand of interpreting is affected by comprehension and production tasks and their interference with each other. This approach to understanding cognitive load in interpreting broadens the scope of research on cognitive load and sets the stage for the integration of the multimodality approach in interpreting research (Seeber, 2017; Chmiel et al., 2020).

In our discussion, the cognitive load in distance interpreting as well as the general and interpreting-specific cognitive models will remain crucial to identifying the causal factors and their measurement methods.

OVERVIEW OF THE THEORETICAL FRAMEWORK OF THE STUDY

As demonstrated in **Figure 1**, cognitive load in DI comprises mental load, mental effort, and pertinent measurement models (in the box on the right-hand side). Inspired by the CLT surveyed earlier, mental load refers to the cognitive demand of the DI tasks and environment, whereas mental effort refers to the total work done by the DI interpreter to complete the task. That is, task and environment factors impose mental load on the interpreter, who then devotes measurable mental efforts to perform the task (Meshkati, 1988; Yin et al., 2008; Chen, 2017).

Based on Paas et al. (2003) and Schultheis and Jameson (2004), the preceding factors in DI can be measured using four methods: subjective methods (Ivanova, 2000), performance methods (Han, 2015), analytical methods (Gile, 2009; Seeber, 2013) and psycho-physiological methods (Seeber and Kerzel, 2011). Subjective methods demand that participants rate the cognitive load they have experienced or are experiencing in a task; performance methods are used to measure cognitive load based on participants' overt performance and behavior in DI; analytical methods are also subjective and based on the prior knowledge of the investigator who estimates or predicts the cognitive load of the input; and psycho-physiological methods are used to evaluate the neurophysiological processes underlying DI to infer the cognitive load that participants experience (Paas et al., 2003; Schultheis and Jameson, 2004).

As discussed later in this paper, the use of measurement methods in quantifying cognitive load in DI offers several advantages. First, they allow the researcher to quantify and measure cognitive load in DI as a multidimensional mental and verbal activity. The causal factors, which regulate change in cognitive load in DI, have the role of surrogates or indicators of immediate or cumulative cognitive load, so measuring these factors would provide an estimation of cognitive load in DI. Second, the measurement methods can be used to predict the mental effort that the distance interpreter may encounter by either subjective reports from the interpreter or objective measurement or analysis of the interpreter's performance and/or their psycho-physiological reactions. As a result, stakeholders (such as conference organizers) might use measurement results to create more ergonomically effective working environments, thus allowing the

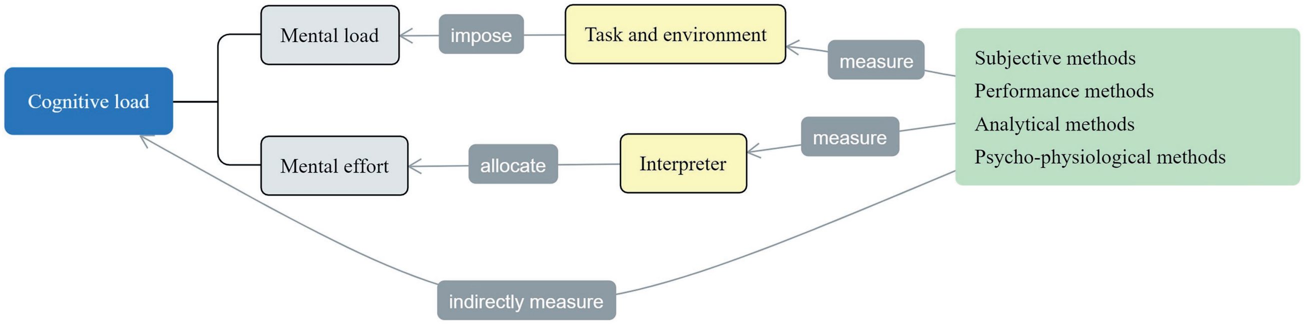


FIGURE 1 | Representation of cognitive load and its measurement in DI.

interpreter to direct their mental effort to improving their performance. Third, the DI interpreting trainers use these measurement methods to estimate the mental load that task and environment factors impose on the trainees. This way, interpreter trainers can select proper training materials and design proper training curriculum to prepare the future DI interpreters.

We will further discuss the advantages (and shortcomings) of each measurement method in the context of DI, and how they may be used to quantify the indicators of cognitive load in DI.

CAUSAL FACTORS

A General View of the Causal Factors

The cognitive capacity of DI interpreters is key to performing interpreting tasks. Cognitive capacity can be proxied by measurements of cognitive load consisting of mental load and mental effort. Understanding the task factors that induce change in mental load can help trainers to prepare the DI interpreters for complex interpreting tasks. Similarly, it is helpful to understand and measure interpreters' mental effort to discriminate interpreter levels such as novice versus master interpreters, which, in turn, can provide diagnostic information in training programs.

However, cognitive load is a latent construct, which cannot be directly measured. Accordingly, measurement of cognitive load is carried out through delineating and operationalizing observable surrogates (indicators) that proxy cognitive load (Chen, 2017). In our study, the surrogates are called causal factors of cognitive load, meaning that these factors can *cause* change in cognitive load in DI—i.e., they have a cause-effect relationship with cognitive load. In theory, in a well-designed study where extraneous factors are properly controlled, the amount of cognitive load of DI interpreters can be directly manipulated by changing the magnitude of causal factors. A recent study by Braun (2020) shows that the causal factors in DI are more diverse than those in the related fields such as SI (e.g., Kalina, 2002; Campbell and Hale, 2003).

In the absence of a validated model for the factors that cause cognitive load in DI, we synthesized the extant literature in interpreting studies (e.g., Kalina, 2002; AIIC, 2021) and the communicative language ability framework (Bachman and Palmer, 1996), to present an integrative framework of DI that specifies causal factors (as demonstrated in **Figure 2**). The proposed model for causal factors in interpreting consists of two dimensions: (1) task and environment factors, and (2) interpreter factors. The task and environment factors comprise DI task factors, linguistic and paralinguistic features of the input, as well as the environment factors. The DI interpreter factors, on the other hand, comprise linguistic knowledge, topic knowledge, personal traits, technology awareness, interpreting strategies and (meta-) cognitive processes of the interpreters. It is suggested that the completion of an interpreting task is the result of the interaction between task and environment and interpreter characteristics. Even though these characteristics are defined separately, they are related, since there is no interpreting without an interpreter and there is no interpreter if there is no interpreting task in a given environment. Nevertheless, in experimental designs, investigators usually manipulate one or a few factors and keep the other factors constant to examine the change of cognitive load ensued from the manipulated factors (Choi et al., 2014). For interpreters with similar proficiency levels, better performance can be viewed as an indication of a lower cognitive load due to facilitating environments (e.g., interpreting from a home studio would be more convenient than interpreting from a booth at a conference room.). In experimental designs to examine cognitive load in DI, the causal factors can be perceived as independent variables which exert an influence over cognitive load, thus resulting in the difference in measurement and analysis results (e.g., interpreters' performance quality and/or gaze behavior measured by eye-fixation indices).

The factors and sample studies in interpreting research are presented in **Table 1**. DI shares many features with onsite interpreting but differs from it in that the DI process is mediated by technology (Mouzourakis, 2006; Braun, 2020). These characteristics of DI add "another layer of complexity to communication" (Davitti and Braun, 2020, p. 281) and have the potential to cause significant change in interpreters' cognitive load. For this reason, we italicized the factors that can significantly influence the DI mode to differentiate their mechanism from interpreting. Therefore, the italicized factors distinguish DI from onsite interpreting, while the non-italicized features are shared between the two. It should further be noted that the

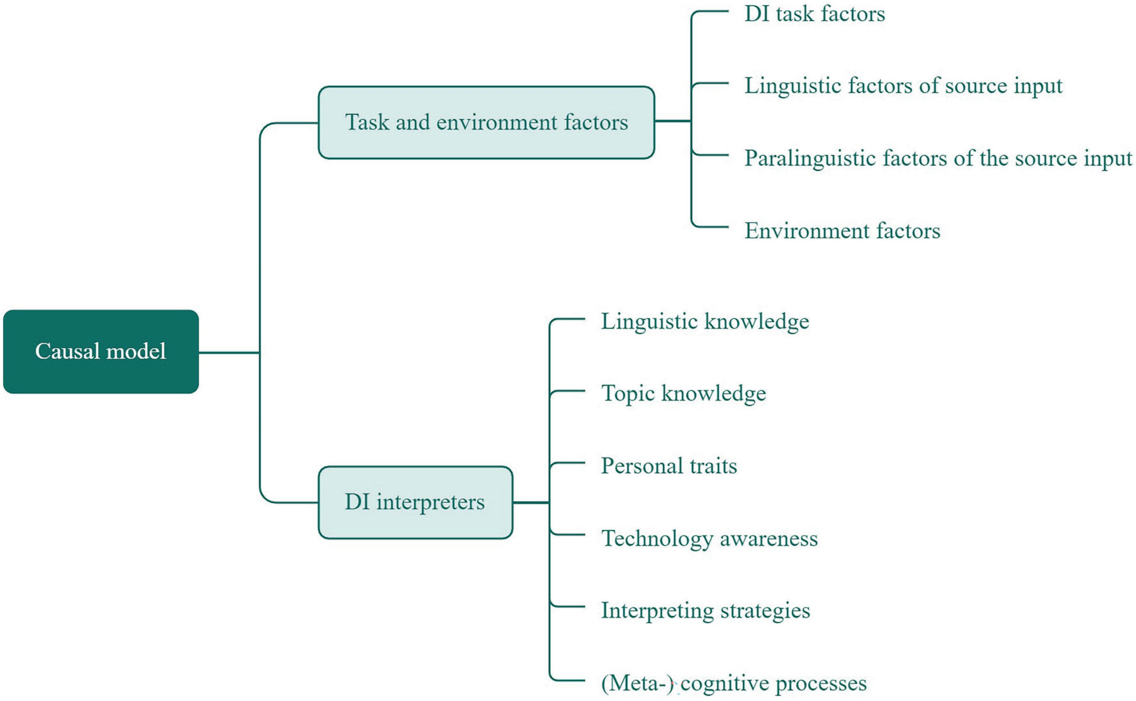


FIGURE 2 | Causal factors that induce change in interpreters' cognitive load during DI.

factors presented in this framework do not make an exhaustive and complete list but rather present a framework in-progress that is to be extended as further empirical and conceptual studies are conducted. From this perspective, the framework is reminiscent of the notion that "no model is meant to correspond exactly to the phenomena" (Moser-Mercer, 1997, p. 159) and as our knowledge about DI expands, so does the model. We will provide a detailed discussion of the causal factors and show how they can result in change in the cognitive load and performance of interpreters.

Task and Environment Factors

In a cognitive load model, task and environment factors include task criticality and psychological and environmental factors (Meshkati, 1988). There are four general factors that constitute the task and environment factors, namely DI task factors, linguistic factors of source input, paralinguistic factors of source input and environment factors. They are discussed below.

DI task factors are the mode of presentation, materials received, information specificity, DSI turn duration, participant interactivity, and directionality. With respect to the mode of presentation, DI interpreters may receive audiovisual input, audio only, or video (Braun, 2015). It remains inconclusive as to whether the use of one modality (audio only) or two modalities (audio and visual input) can reduce the cognitive load of DI interpreters (Moser-Mercer, 2005a; Sweller et al., 2011). Nevertheless, it has been argued that "[i]nterpreting *via* video link has come to be seen as a more effective way of providing spoken language interpreting services than telephone interpreting" (Skinner et al., 2018, p. 13). In video-relayed DI, the image typically consists of the speaker, the podium, a panoramic view of the meeting room, or a partial view of the meeting room (Mouzourakis, 2003). However, in cloud-based meetings and conferences, only the speaker can be captured by the computer-embedded camera and/or the image on the screen with PowerPoint (PPT) slides. The image provided is determined by the angle of the feeding cameras, which might be microphone-activated or handled by multiple cameramen. In onsite SI, interpreters can make decisions about their own viewing behaviors to search for useful information to assist with the comprehension and production of language. Thus, the gaze of interpreters is a problem-driven, selective, and active process (Mouzourakis, 2006). In addition, research in language comprehension has shown that dividing attention between spatially and temporally segregated stimuli results in high cognitive load measured by proxies of gaze behavior and brain activity (Aryadoust et al., 2020b). This suggests that the separation of different sources of input in multimodal DI should be considered and investigated.

Materials and information specificity are pre-interpreting factors, which refer to the materials given to the interpreters before the conferences or meetings to facilitate their preparation (Kalina, 2005). Interpreters may receive PPT slides, speech drafts, or other supplementary materials before the conference (Díaz-Galaz 2011). Under such circumstances, these factors are expected to have a facilitating role in DI, thus probably reducing cognitive load.

TABLE 1 | An overview of previous studies on the causal factors that change interpreters' cognitive load during interpreting.

General dimensions		**Factors**	**Examples and/or explanation**	**Sample relevant studies**
Task and environment factors	DI task factors	*Mode of presentation*	Audio only, video; Images: none, slides, online platform window (e.g., Zoom), stage view, head view, upper body view, spatial arrangements	Mouzourakis (2006); Braun and Taylor (2012)
		Materials and information specificity	Slides, speech draft, charts, reference list, website links	Kalina (2002); Díaz-Galaz et al. (2015)
		Turn duration	20 min per turn/ one speaker per turn	AIIC (2021)
		Participant interactivity	With an onsite audience, no onsite audience, platform audience (e.g., Zoom)	Napier et al. (2018)
		Directionality	L1-L2, L2-L1	Chen (2020)
	Linguistic factors of the source input	Lexical	Content word density, vocabulary difficulty	Korpal and Stachowiak-Szymczak (2018); Plevoets and Defrancq (2018)
		Syntactic	Simple sentences, compound sentences	Seeber and Kerzel (2011); Ma et al. (2021)
		Semantic	Culturally loaded expressions	Zheng and Zhou (2018)
		Textual organization	Cohesion, genre	Vianna (2005); Kuang and Zheng (2022)
		Pragmatic	Intention, implication	Setton (2002)
		Sociolinguistic	Language variety, register	Jenkins (2000)
	Paralinguistic factors of the source input	Delivery speed	Fast, normal, slow	Korpal and Stachowiak-Szymczak (2019)
		Accent	American English, Singlish	Gluszek and Dovidio (2010)
		Body language	Audio only (no body gestures at all) or video (with/without facial expressions, with/without upper body gestures)	Braun (2020); Vranjes and Oben (2022)
	Environment factors	*Booth type*	DSI hub, home studio, onsite booth, virtual booth	AIIC (2020a)
		Platform	Open or closed platforms	AIIC (2020b)
		Physical parameters	Acoustic condition, ventilation, lighting, air quality	Moser-Mercer (2003); Roziner and Shlesinger (2010)
		Transmission techniques	Onsite infrared/radio signals, ethernet, WiFi, mobile data; bandwidth	Mouzourakis (2006)
Interpreter's factors		Linguistic knowledge	Knowledge of the language DI interpreters work in: grammatical knowledge, textual knowledge, functional knowledge, and sociolinguistic knowledge	Kalina (2000); Napier (2002); Li (2018)
		Topic knowledge	The content knowledge of the cultures the interpreters work in	Kalina (2000)
		Interpreting strategies	The skills to comprehend and produce languages	Chang and Schallert (2007); Ma and Li (2021); Amos et al. (2022)
		Personal traits	Personal factors like motivation, anxiety and stress resistance	Chiang (2009); Rojo López et al. (2021); Kuang and Zheng (2022)
		Technology awareness	Being mindful of the use of technology and the ability to use technology in interpreting	Desmet et al. (2018); Prandi (2018)
		(meta-)cognitive processes	The ability to build up a mental representation; meta-cognitive processes to plan, monitor and evaluate	Cañada and Arumí (2012); Lin et al. (2018); Bravo (2019)

Another task-related factor is turn duration in DI in simultaneous mode, which is suggested to be 20–30 min to avoid interpreter fatigue and possible decreases in interpreting quality (Kalina, 2002; Chmiel, 2008). However, there is no consensus over the turn duration in the general field of DI or distance consecutive interpreting. Despite the inadequacy of research on this factor, it can be surmised that physical and mental fatigue is likely to affect the quality of DI, and should therefore be taken into account in designing studies and in real-life DI.

In contrast with onsite interpreting, DI places the participants (i.e., interpreters, speakers and audience) in physically remote locations in computer-based collaborative environments, which may lead to less seamless participant interactivity. A growing body of evidence suggests that online environments can facilitate participant interactivity during interpreting (Arbaugh and Benbunan-Fich, 2007). In DI, especially in telephone-relayed interpreting, participant interactivity mainly includes turn-taking management, stopping the primary speakers(s) by using proper techniques to cut in, and seeking clarification (Wang, 2018). In simultaneous DI, where interpreters work with a partner or a team (in multilingual relay interpreting), physical separation from partners, clients, and technicians may affect performance in the distance mode. For instance, while onsite SI interpreters can maintain close contact with clients in a conference hall to acquire first-hand updates on conference procedures, DI interpreters cannot have physical proximity with the client (Davitti and Braun, 2020). Therefore, the remote mode of DI is likely to impede or minimize interactivity and limits the efficacy of communication (Licoppe and Veyrier, 2017; Powell et al., 2017). Mouzourakis (2003) believes that better technology equipment and more ergonomic arrangements of screens and monitors can mitigate this obstacle.

With regards to directionality, it has been recommended that SI interpreters should work from their L2 language (second language) to their L1 language to ensure quality (Seleskovitch and Lederer, 1989). However, scholars are still debating whether working in both directions is feasible, especially in the Chinese and Arabic languages based on empirical performance tests (i.e., Al-Salman and Al-Khanji, 2002; Chang and Schallert, 2007).

Linguistic factors of the source input entail the lexical, syntactic, semantic, textual, pragmatic, and sociolinguistic levels, and are common between onsite interpreting and DI (Hymes, 1972, 1974; Canale and Swain, 1980; Leung, 2005). At the lexical level, some types of vocabulary can result in a heavier cognitive load, specifically technical terminologies, numbers, and acronyms, since these words require domain-specific knowledge to process (Kalina, 2005; Gile, 2008). Semantic units, such as culturally loaded expressions or counterfactuals, might cause cognitive saturation in DSI (Setton, 2002; Vianna, 2005). Syntactic organization and complexity can also affect cognitive load in SI (Chmiel and Lijewska, 2019). For example, Seeber and Kerzel (2011) used eye-tracking to investigate the differences in the cognitive load of verb-initial (syntactically symmetrical) and verb-final (syntactically asymmetrical) structures during German-English SI, finding higher cognitive load with asymmetrical structures. In another study, Setton (1999, 2001, 2002) suggested that the pragmatic dimension, which is pertinent to the underlying message, can help build a mental model of SI for analyzing and demonstrating attitudes, intentions, and implications.

Sociolinguistic factors are related to language variety and register. For example, with English as a *lingua franca*—meaning that English is used as a common communicative means across different cultures—unique varieties of English have emerged in different parts of the world (e.g., Singlish, Indian English, etc.; Firth, 1996; Jenkins, 2000). Thus, an unfamiliar variety of linguistic codes might increase interpreters' cognitive load. In fact, language policy research on how interpreters react to different varieties of language has gained much popularity among translation and interpreting researchers due to its strong influence in the 21st century (Zhu and Aryadoust, 2022).

Paralinguistic characteristics of the source input include delivery speed, accent, and body language of the speakers. Delivery speed is known to be a major factor affecting cognitive load in SI (Pio, 2003; Meuleman and Van Besien, 2009). Accent also contributes to cognitive load change in SI since it affects the comprehension of the input (Gile, 2009; Gluszek and Dovidio, 2010). Another factor is body language, which in DI is quite different from that in onsite SI where interpreters can see and thus utilize the body language and facial expression of speakers for useful information. However, due to the scale and arrangement of the interpreting booth, interpreters might not have a clear visual of the speakers. In DI, body language (if any, depending on the image, as discussed earlier) is delivered using digital technology. As a result, interpreters might have high-quality images of the speaker's body language to use for paralinguistic help. However, at the time of writing this paper, whether DI digital images or onsite images influence interpreters' cognitive load is unchartered waters requiring further study. Furthermore, the kind of digital images that may lower cognitive load, thus enhancing performance, is still open for discussion.

Environment factors include booth type, platform, physical parameters, transmission techniques, and contact with participants (Braun et al., 2018; AIIC, 2021). DSI interpreters might work in DSI hubs, home studios, or onsite booths that vary not only in their air quality, acoustics, ventilation, and temperature, but also differ in signal quality based on whether ethernet, WiFi, or mobile data is used for transmission (Moser-Mercer, 2003; Mouzourakis, 2006). For example, home-based studios may not have optimal acoustics since their location may be a home office or simply a desk in the corner of the interpreter's house. Research shows that background noise affects interpreters' stress levels, thus increasing cognitive load (Koskinen and Ruokonen, 2017). In a recent neuroimaging study, Lee et al. (2020) found that environmental and nature sounds evoke significantly different neurocognitive processes than long pieces of discourse. Similarly, it has been shown that noise could have adverse effects on language comprehension performance under experimental conditions (Fujita, 2021). This suggests that the concurrent presence of background noise and language input in DI can result in the additional activation of some brain regions and intensify the mental load due to interfering noise. In sum, physical environments can interact with task and interpreter characteristics, as predicted by instruction research on CLT (Choi et al., 2014).

Having described the task and environment variables in DI, we will now examine the operator's (in this case, the DI interpreter) characteristics and possible moderating variables (Meshkati, 1988).

Interpreter's Factors in DI

In DI, interpreter factors include linguistic knowledge, topic knowledge, interpreting strategies, personal traits, technology awareness, and (meta-)cognitive processes. Skinner et al. (2018) states that "any modifications to interpreters' working environments are likely to impact their performance and how they process information" (p. 19) since interpreting involves highly complex cognitive and metacognitive processes (Moser-Mercer, 2000; Gile, 2009). That is to say, the factors of the (meta-)cognitive processes are under the influence of all the other factors as indicated in **Figure 2**, which are explained in detail here.

The prerequisite for any DI interpreter to render professional services is *linguistic knowledge*, which is knowledge of the languages they work in (Kalina, 2000). The linguistic knowledge required in interpreting can be thought as "a domain of information in memory that is available for use" (Bachman and Palmer, 1996, p. 67). Linguistic knowledge is a component of the memory storage and consists of an array of competencies such as grammatical and discourse competence. It may be said that interpreters' linguistic knowledge contributes to their capacity to perform more complex cognitive tasks (Choi et al., 2014). Then, if the tasks and other conditions remain the same, the interpreter with better linguistic knowledge can perform better since they can allocate their memory resources to other cognitive tasks involved in interpreting such as auditory and/or visual processing, long-term memory extraction or speech production monitoring.

There are four areas of linguistic knowledge in DI, similar to those described by Bachman and Palmer (1996): grammatical knowledge, textual knowledge, functional knowledge, and sociolinguistic knowledge. Grammatical knowledge concerns vocabulary, syntax, and phonology or graphology. Textual knowledge concerns cohesion, rhetorical use, and conversational organization. Functional knowledge involves the use of ideational functions (i.e., descriptions of happiness, explanations of sadness), manipulative functions (i.e., compliments, commands), heuristic functions (i.e., teaching, problem-solving), and imaginative functions (i.e., jokes, novels). Sociolinguistic knowledge is knowledge of dialects, registers, idiomatic expressions, culture-loaded references, and figures of speech. Numerous studies have shown that linguistic knowledge is directly proportional to the cognitive load, thus affecting their performance (Gile, 1995; Angelelli and Degueldre, 2002).

Topic knowledge of cultures of various countries or regions is another key factor in DI. This includes knowledge of the administrative structures of both sides of participants, political, economic, social, and ethnic characteristics of the participants' areas of origin, and even literature and the arts (Kalina, 2000).

Topic knowledge has always been recognized as a powerful predictor of comprehension in both expert and novice groups since it interacts with text structure and verbal ability at the micro level while engaging in the metacognitive strategy use at the macro level (McNamara et al., 2007; Díaz-Galaz et al., 2015). That is, interpreters use the topic knowledge to integrate the information presented in the source speech in a continuous way to construct a mental representation, which they subsequently reformulate and articulate in the target language.

In recent years, *personal traits*, also termed personal psycho-affective factors, such as motivation, anxiety, and stress resistance, have also been recognized as important components of an interpreter's aptitude (Bontempo and Napier, 2011; Rosiers et al., 2011; Timarová and Salaets, 2011). For example, Korpal (2016) measured heart rate and blood pressure to determine whether the speed of a speaker's delivery influenced the interpreter's stress level. Heart rate, but not systolic or diastolic blood pressure indices, was significantly associated with speech rate, supporting the assumption that a faster speech rate can make interpreters experience higher levels of stress. In DI, several studies have examined stress and burnout, finding that the DI interpreters experience high levels of stress and burnout (Roziner and Shlesinger, 2010; Bower, 2015). A study by Seeber et al. (2019) suggested that providing appropriate visual input is important to alleviate the alienation of DI interpreters and a study by Ko (2006) concluded that longitudinal empirical studies are an essential methodology in DI research. The research methods used to investigate personal traits are becoming increasingly diverse, ranging from qualitative surveys to objective psycho-physical instruments (e.g., eye-tracking).

Interpreting strategies are derived from the understanding of the underlying processes of interpreting, and can help interpreters to apply the optimal interpreting solutions across communicative settings (Riccardi, 2005). These strategies refer to the skills needed to comprehend and produce language, which might be included among or in addition to the strategies used in monolingual language processing. Notably, interpreters (i) are not expected to alter or filter out information, and (ii) may not have sufficient domain knowledge (Riccardi, 1998, 2005; Kalina, 2000; Shlesinger, 2000; Chang and Schallert, 2007).

Given that DI is a technology-supported language mediation method, *technology awareness* is an important interpreter factor that affects cognitive load. Compared to translation, where computer-assisted services have become standard, interpreting has experienced only a modest impact due to advances in information and communication technology (Fantinuoli, 2018). This is because voice recognition technologies cannot fully recognize natural spoken language and its hesitations, repairs, hedges, and unfinished sequences (Desmet et al., 2018). Nevertheless, technology can be used to support DI in multiple ways.

Technology awareness in DI comprises two dimensions: (i) being mindful that technology is an inevitable part of the interpreting industry and being ready to accept it; and (ii) being able to recognize and utilize technology to perform DI (Parsons et al., 2014). For example, it is well known that numbers are difficult to interpret and consume tremendous cognitive resources. To support the translation of numbers, Desmet et al. (2018) used booth technology to automatically recognize numbers in source speech and present them on a screen, which significantly enhanced translation accuracy. Similarly, Prandi (2018) explored the use of computer-assisted interpreting tools to manage terminology, aiming to reduce local cognitive load during terminology search when delivering the target text. With new information and tools emerging due to technological advances, DI interpreters need to be fully prepared to reap the advantages and opportunities of technology while minimizing the risks and consequences that arise from their use.

Mellinger and Hanson (2018) pointed out that the intersection between technology and interpreting remains "under-theorized," particularly regarding the adaptation of technology in accomplishing the interpreting task. They conducted a survey-based investigation to examine the self-perception of the interpreters' role in technology use and adoption. They found that community interpreters in medical and court settings where DI first appeared and achieved acceptance are more likely to adopt new technologies than their counterparts in conference settings. Few empirical studies have hitherto examined interpreters' technology awareness, thus leaving a big gap in the understanding of the use of technology in DI.

Cognitive process is the ability to build up a mental representation of the verbal message through comprehension, parsing the information, and integrating it with one's own pre-existing knowledge (Setton, 1999). *Metacognitive processes*, on the other hand, refer to strategies for efficient management of processing resources and consist of planning, monitoring, and evaluating (Flavell, 1976). Bravo (2019) reviewed metacognition research in interpreting and concluded that "[m]onolingual communication requires a lesser degree of metacognitive awareness than interpreter-mediated communicative events do" (p. 148). This is due to the fact that the nature of the interpreting task, which demands the ability to quickly shift attention across many cognitive activities, needs a meta-level skill to perform quality control. Through applying proper cognitive processes and metacognitive strategies, interpreters interact with the task and environment factors, a process that influences interpreters' performance.

So far, we have looked into the factors which induce change in cognitive load during the interpreting process and categorized them as causal factors consisting of task and environment factors and interpreter factors. These factors act as the proxies representing dimensions of cognitive load. The methods for measuring these factors to assess the cognitive load that DI interpreters experience are discussed in the following section.

MEASUREMENT METHODS

Cognitive load in DI is conceptualized as a multidimensional construct representing the load that performing a particular interpreting task imposes on the distance interpreter's cognitive system (Paas and Van Merriënboer, 1994; Seeber, 2013). To measure this construct, "the most appropriate measurement techniques" should be used (Meshkati, 1988, p. 306). Scholars have attempted to model and measure cognitive load with various methods (DeLeeuw and Mayer, 2008; Wang et al., 2020; Ayres et al., 2021; Krell et al., 2022; Ouwehand et al., 2022). In this study, we adopt a comprehensive and fine-grained categorization of Paas et al. (2003) and Schultheis and Jameson (2004) who proposed four discrete methods for measuring cognitive load: subjective methods, analytical methods, performance methods, and psycho-physiological methods. This section provides a review of the pros and cons of each of these measurement methods and how they can be used to investigate specific DI factors. Some of the pioneering studies utilizing these measurement methods in DI are presented in **Table 2**. These measurement methods, despite having been used and verified in interpreting studies for quite some time, have not been widely used in DI research. Therefore, the following discussion of these methods will largely rely on their previous application in interpreting studies.

Subjective Methods

Subjective methods of measuring cognitive load in DI—such as Likert scales and verbal elicitations and reports—require participants to directly estimate or compare the cognitive load they experienced during a specific task at a given moment (Reid and Nygren, 1988). Subjective methods of cognitive load are based on the assumption that participants are able to recall their cognitive processes and report the amount of mental effort required, which is a limitation that researchers should be aware of (Ericsson and Simon, 1998). Subjective measures enjoyed popularity in early research because they are easy to use, non-intrusive, low-cost, and can discriminate between different load conditions (Luximon and Goonetilleke, 2001).

In interpreting studies, subjective methods can provide data on: (i) how interpreters allocate attention; (ii) problem-solving strategies used by interpreters; (iii) the effect of interpreting expertise; and (iv) general assessment of cognitive activities in interpreting (i.e., comprehension, translation, and production; Ivanova, 2000). In early studies of DI, subjective methods were used to investigate how transmission techniques are used by interpreters (Böcker and Anderson, 1993; Mouzourakis, 1996, 2006; Jones et al., 2003). Among them, Mouzourakis (2006) reviewed the large-scale DI experiments that were conducted at the United Nations and the European Union institutions in which the subjective data collected by questionnaire were

TABLE 2 | Measurement methods of cognitive load in DI.

Measurement methods	Pioneering studies in DI	The specific factors under investigation
Subjective methods	Mouzourakis (2006); Bower (2015)	Stress and burnout; Sound and image transmission in the environment
Analytical methods	Braun (2007)	Cognitive process of the interpreter
Performance methods	Jiménez-Ivars (2021); Braun (2013); Gany et al. (2007)	Psycho-affective factors of the interpreters; Distance working mode; Distance working mode
Psycho-physiological methods	Kuang and Zheng (2022); Roziner and Shlesinger (2010)	Source speech difficulty and interpreters' experience; Mode of presentation of the DI task

used to indicate how the technical setup for sound and image transmission would impact interpreters' perceptions of DI. Later, this method was also adopted in DI studies to evaluate interpreters' stress levels (Ko, 2006; Bower, 2015; Costa et al., 2020), the effect of interpreters' visibility on participants (Licoppe and Veyrier, 2017), turn-taking (Davitti and Braun, 2020; Havelka, 2020), and the effect of different presentation modalities (Braun, 2007).

However, Lamberger-Felber (2001) performed a comparative study of objective text presentation parameters and interpreters' subjective evaluations of texts, finding differences for almost all parameters investigated. This caution is important because subjective judgments usually serve as the method (or part of the method) used to assess the difficulty of instruments (e.g., Su and Li, 2019; Weng et al., 2022) together with more 'objective' measures (e.g., using textual analysis to measure the difficulty level of the source text). Seeber (2013) acknowledged that subjective methods might not be able to "reliably assess cognitive load" (p. 8) in SI.

Chandler (2018) cautioned that subjective methods are limited due to their dependence on participants' self-appraisal and personal judgment as well as being problematic with young children. For example, Low and Aryadoust (2021) found that listeners' self-reports of strategies had a significantly lower predictive power in accounting for oral comprehension performance compared with gaze behavior measures collected through an eye tracker. In the general field of cognitive load measurements, an instrument is usually developed and validated in one study, and further validated in other studies under different conditions. A typical example is that the cognitive load scale (CLS) developed by Leppink et al. (2013) was widely used as a subjective measurement tool of three types of cognitive load at large, and was further validated and expanded by Andersen and Makransky (2021) to measure the cognitive load for physical and online lectures. In DI, the lack of internal reliability of the instruments is a noteworthy concern, since the instruments are mostly formulated by the investigators to answer their own specific questions, and therefore require further validation. Recognizing these limitations, Ayres et al. (2021) proposed that the combination of subjective and physiological measures is most effective in investigating change in cognitive load. Therefore, we recommend that the results of the studies that use subjective methods for data collection be cross-validated with objective techniques such as eye-tracking and neuroimaging (see Aryadoust et al., 2020b, for an example of measuring cognitive load using eye-tracking and brain imaging in comprehension tasks). In addition, appropriate reliability checks should be applied to ensure the precision and replicability of subjective methods of measuring change in cognitive load in DI (Moser-Mercer, 2000; Riccardi, 2005; Fantinuoli, 2018).

Performance Methods

Performance methods of measuring cognitive load have long been used in interpreting studies to measure speed and accuracy. For example, calculation of the ear-to-voice span (EVS) by Oléron and Nanpon (1965) probed cognitive processing of interpreters' performance *via* quantitative measures. Later, Barik (1973) compared the performance of professionals, interpreting students, and bilinguals without any interpreting experience by counting the errors, omissions, and additions in their output. This tradition of comparing the performance of participants with different levels of expertise has carried on until the present day in investigating cognitive behavior. Given that participants' performances may have been evaluated by human raters with different leniency and severity, modern psychometric methods like the many-facet Rasch measurement (MFRM) have been used to validate rating scales and identify the degree of severity/leniency of raters (Han, 2015, unpublished Doctoral dissertation)[2].

Performance methods are also used to compare the different tasks concerning language comprehension and production. For example, Christoffels and De Groot (2004) designed an experiment comparing SI, paraphrasing, and shadow sentences (repeating), with the latter two considered to be delayed conditions. The authors assumed that participants would perform better in the delayed conditions than in SI, since simultaneous comprehension and production are a major cause of difficulty in SI. The results showed that the poorest performance was for SI, followed by paraphrasing and then repeating, which indicated the increased cognitive load of SI compared to the other two tasks.

In DI, performance methods have been widely used to compare the quality of onsite interpreting along with interpreting in different distance modes (Oviatt and Cohen, 1992; Wadensjö, 1999; Moser-Mercer, 2005a,b; Gany et al., 2007; Locatis et al., 2010; Braun, 2013, 2017; Jiménez-Ivars, 2021). For example, Gany et al. (2007) investigated how various interpreting methods affect medical interpreters' speed and errors through comparing their DI performance with onsite interpreting, finding that the DI mode resulted in fewer errors and was faster. Combined with interpreters' subjective perceptions of their performances, objective performance measures have helped to increase the acceptance of DI. However, previous studies are not directly comparable with each other since they were set up under different circumstances and interpreting modes, making the further evaluation of DI quality necessary (Braun, 2020).

For measuring cognitive load in DI, it is suggested that performance measures be used along with other methods to investigate factors like interpreters' overall performance, interpreters' linguistic knowledge, and topic knowledge (Mazza, 2001; Gile, 2008; Timarová and Salaets, 2011).

Analytical Methods

Analytical methods use expert opinions and mathematical models to estimate cognitive load in DI and interpreting at large (Paas et al., 2003). Psychologists proposed an analytical and empirical framework to accommodate the measurement of cognitive load (e.g., Linton et al., 1989; Xie and Salvendy, 2000). Following their lead, early interpreting scholars proposed

---

[2] Han, C. (2015). *Building the validity foundation for interpreter certification performance testing.* [Unpublished Doctoral dissertation]. Macquarie University.

several cognitive models to explain the cognitive processes involved in interpreting (Gerver, 1975, 1976; Moser, 1978; Gile, 1995, 2009; Mizuno, 2005; Seeber, 2011; Ketola, 2015).

Gerver's (1975, 1976) model focused on the short-term storage of the source text, which stays in a hypothetical input buffer in the mind from which the source text is sent out for further processing. The processing is performed in cooperation with long-term memory, which activates pertinent linguistic units for externalization *via* an output buffer. It is quite a modern idea that Gerver (1976) described information from several sources being processed in parallel. However, the separate input and output buffers still lack any theoretical or empirical support (Timarová, 2008).

Moser (1978) proposed another cognitive model of SI that assigns a crucial role to WM, which she instead called generated abstract memory (GAM). In this model, WM is conceptualized as a structural and functional unit that stores processed chunks (the STM storage function), performs comprehension tasks in cooperation with long-term memory (the executive functions), and plays a role in production as well (see Moser, 1978, p. 355 for details). According to Moser-Mercer (1997), the model also contains a prediction step for incoming content which, she argues, plays a crucial role in interpreting.

An alternative cognitive model of SI was proposed by Darò and Fabbro (1994). This model is based on the models of Baddeley and Hitch (1974) and Baddeley (1992) but only adopts the verbal slave system and central executive elements. Darò and Fabbro (1994) explicitly separate the cognitive process into two general processes: WM and long-term memory processes. Interestingly, in this model, translation in two directions is performed by two separate mental modules, which are the basis for investigation of directionality effects in future studies (See Darò and Fabbro, 1994, p. 381, for model details).

Gile's (1995, 2009) effort model and Seeber's (2011) cognitive load model are two milestone analytical models in measuring cognitive load. These models were extensively used as a means of understanding cognitive load in SI and are discussed in terms of the measurement of cognitive load in the current study. Both models provide useful insights for capturing the complex multi-tasking process in SI. However, Seeber (2013) acknowledges that both models are unable to account for individual differences, which is a major constraint for measuring cognitive load in interpreting. For example, due to individual differences in EVS in SI, it is difficult to establish a cause-and-effect relationship between cognitive load, performance quality, and performance speed (Seeber, 2013).

Braun and her colleagues attempted to establish several analytical frameworks to assess DI interpreters' cognitive load generated across different settings (Braun, 2007, 2017; Braun and Taylor, 2012). For example, Braun (2007) used a process-oriented model of communication in which linguistic and cognitive processes contribute to discourse comprehension and production. Braun managed to investigate the adaptation process of the DI interpreters. Using a microanalytical framework, Davitti and Braun (2020) drew on extracts from a corpus of DI encounters to identify the coping strategies in online collaborative contexts, which include managing turn-taking, spatial organization, and the opening and closing of a DI encounter. For example, by analyzing the spatial organization behavior of interpreters, they found that explicit instructions from DI interpreters can create more mutual visibility and awareness to ease their mental load and support their performance. Overall, the study provides substantial implications for interpreter education.

Analytical methods in interpreting can be mainly used to evaluate input dimensions of cognitive load, meaning that researchers can apply judgment and/or mathematical methods to measure the cognitive load induced by the task and environment factors (Chen, 2017; Ehrensberger-Dow et al., 2020). Although a fine-grained analytical model of the overall cognitive load in DI is still to be developed, we believe the current study would be a step toward specifying the key causal factors of cognitive load in DI and their measurement methods.

Psycho-Physiological Measures

Based on the assumption that psycho-physiological variables covary with cognitive load, psycho-physiological measures have been used in the investigation of cognitive load, such as changes in heart, brain, skin, or eye responses (Paas et al., 2003; Schultheis and Jameson, 2004). Psycho-physiological measures can provide direct and continuous data in DI, allowing online measurement with a high sampling rate and sensitivity without interference from additional task(s). These measures are particularly useful for probing the "black box" of interpreters, that is, their cognitive process (Seeber, 2013). Brain imagining, stress hormones, eye-tracking, and galvanic skin response (GSR), which are widely used methods in measuring cognitive load, are discussed here.

Brain Imaging

Brain imaging can provide a "window" to examine cognitive load in interpreting. Kurz (1994, 1995) used electroencephalography (EEG) to investigate the effect of directionality during shadowing and SI tasks. Price et al. (1999) and Rinne et al. (2000) both used positron emission tomography (PET) to examine brain activation during interpreting, finding pronounced bilateral activation of the cerebellum and temporal and frontal regions during the assigned tasks. Tymoczko (2012) pointed out that using technology and neuroscience in interpreting research is "one of the most important known unknowns of the discipline" (p. 98). Recent research (e.g., Elmer et al., 2014; Becker et al., 2016; Hervais-Adelman et al., 2018; Van de Putte et al., 2018; Zheng et al., 2020) has continued to use the functional magnetic resonance imaging (fMRI) method of brain imaging to better understand interpreting (See a comprehensive review of brain imaging in interpreting studies in Muñoz et al., 2018). Of course, a caveat concerning the use of fMRI is that it reduces the ecological validity of experiments. Recent neuroimaging techniques such as functional near-infrared spectroscopy (fNIRS), which allow for more mobility and maintain ecological validity, are recommended as alternative measurement methods (Aryadoust et al., 2020a).

The Stress Hormone Cortisol

The level of the stress hormone cortisol is a psycho-physiological measure that has been used in DI research. Roziner and Shlesinger (2010) sampled the interpreters' saliva four times a day for the cortisol levels, and compared these indices between the onsite and distance modes, finding that the mean cortisol levels of interpreters in distance modes were slightly higher than that in onsite modes in working hours. For future studies, different indicators of stress level (e.g., blood pressure, heart rate, GSR) can be measured together for a better understanding of the phenomenon.

Eye-Tracking

Among psycho-physiological techniques, eye-tracking has long attracted the interest of interpreting researchers given that our eye-movements, or gaze behavior, can reflect the continuous processes in our mind (Hyönä et al., 1995; Dragsted and Hansen, 2009; Chen, 2020; Tiselius and Sneed, 2020; Amos et al., 2022; Kuang and Zheng, 2022). Specifically, eye-tracking is noninvasive and has proved to be useful for describing "how the interpreter utilizes his or her processing resources, and what factors affect the real-time performance" (Hyönä et al., 1995, p. 8). In recent years, eye-tracking has been applied in various studies to investigate how cognitive load in the interpreting process is affected by factors including syntactic characteristics of source speech (Seeber et al., 2020), delivery rate (Korpal and Stachowiak-Szymczak, 2020), interpreting strategies (Vranjes et al., 2018a,b), and cognitive effort (Su and Li, 2019). A detailed review of the application of eye-tracking in investigating cognitive load in interpreting research is reported by Zhu and Aryadoust (under review).

Galvanic Skin Response

In recent years, GSR, a marker of emotional arousal, has also attracted attention in interpreting studies (Korpal and Jasielska, 2019). In research using GSR, it is assumed that physiological arousal activates the sympathetic nervous system (SNS), resulting in more pronounced skin conductance.

Even though they require a complex and refined experimental design, psycho-physiological measures appear to be a promising set of methods for measuring real-time or delayed cognitive load during the interpreting process (Seeber and Kerzel, 2011). The task and environment factors discussed earlier induce a certain level of cognitive load, thus leading to physiological changes in interpreters. As a result, interpreters are ideal subjects for physiological measures. Given their promising applications in DI, the question is not whether psycho-physiological measures should be used, but how to control the variables to investigate what is really being estimated (the construct itself; Seeber, 2013).

In summary, the four aforementioned methods of measuring cognitive load in DI each have their own unique strengths and weaknesses. Given that the empirical research on DI started two decades ago, these measurement methods have not been widely applied in DI research. The studies we surveyed above provide examples for future researchers to design their studies. Furthermore, the use of these methods in the broader interpreting field can provide a link between previous interpreting studies and future DI studies. These studies can be replicated in the field of DI to examine how technological challenges and remoteness may alter cognitive complexity and difficulty in interpreting (Braun, 2007), thus providing research directions for interpreting practitioners and trainers. Researchers should consider their research aims and variables under investigation to determine the most appropriate measure types for their study.

TOWARD A COHESIVE REPRESENTATION OF COGNITIVE LOAD IN DI

This paper has discussed cognitive load theory and the general cognitive models in interpreting. It has also presented a discussion of the causal factors inducing change in cognitive load in DI and their measurement methods. These discussions were intended to present a scientific representation to bridge the abstract concept of cognitive load and the world experienced in DI practice and research. As a first attempt to integrate representations of cognitive load and measurement methods in DI, this current discussion offers several important implications related to DI.

First, this synthesis review provides a multicomponential representation of cognitive load in an endeavor to concretize this abstract concept (Huddle et al., 2000). As previously discussed, the factors that change cognitive load play the role of surrogates for it, but they are scattered across previous studies on cognitive load in interpreting. This paper sought to synthesize them into two general categories (i.e., task and environment factors, and DI interpreter factors) and then unpack the categories to make them more accessible to interested researchers. Informed by CLT, the identified causal factors constitute our endeavor to assist the stakeholders (e.g., conference organizers and training program managers) in controlling for extraneous factors and directing mental resources of interpreters to intrinsic and germane sources of cognitive load. The study also provides a framework for future experimental designs to control confounding variables and optimize research design by identifying influential variables in DI and their relationship. For example, in terms of experimentation for identifying causal relationships between DI interpreters' performance and mode of presentation (e.g., video-relayed or telephone interpreting), researchers should control for confounding variables like source speech complexity and participants' proficiency level to be able to establish causality. Our discussion of the causal factors could be used as a checklist for the experiment designers in DI or even interpreting studies in general.

Second, we identified the factors that can affect cognitive load in DI and reviewed the relevant literature in interpreting studies in general and DI in particular, which makes the current study distinct from past discussions of interpreting which tend to be broad and general. Accordingly, the present synthesis may also serve as the basis for future replication studies in DI. This is in line with Rojo López and Martín (2022) who suggested that replication studies are required to foster the standardization of general research methodology in studying the cognition behavior of interpreters. Since technology has played an increasingly important role in the interpreting service industry, the factors specified for DI in this study would provide important implications for research and practice. For example, the identification of technology awareness factors would help curriculum designers and trainers in translation and interpreting programs to embed these factors into course contents (Ehrensberger-Dow et al., 2020).

Third, the framework presented in this paper serves not only as a descriptor of factors affecting cognitive load, but also as a predictive tool in which the pros and cons of the measurement methods were analyzed and presented (Chittleborough and Treagust, 2009). Thus, it can help researchers actively make plausible predictions and informed decisions in study design, e.g., by choosing appropriate measurement methods to conduct investigations and controlling for possible extraneous variables. As earlier noted, subjective measurement methods such as interviews or surveys can help provide *post hoc* evaluations of cognitive load in DI, while performance measures provide indications of interpreting proficiency based on objective and/or subjective ratings. Analytical measures, on the other hand, provide an estimate of cognitive load based on subjective and analytical data, thus relying on the prior knowledge of the investigators (Seeber, 2013). Finally, psycho-physiological measures register detailed real-time patterns of cognitive activity, and require a well-controlled experimental design which is not confounded by construct-irrelevant causal variables.

We note that the current discussion is based mainly on researchers' understanding of cognitive load in DI and interpreting studies in general, which makes it an expressed model. An expressed model needs to be tested by empirical studies and agreed upon by society, to become a consensus model (Gilbert, 2004). Therefore, we call for future empirical studies to validate this representation with recommended methods and expand the scope of these methods to examine the factors contributing to cognitive load in DI.

AUTHOR CONTRIBUTIONS

XZ and VA contributed equally to the conceptualization, writing, and revisions of the article and approved the submitted version.

FUNDING

This research was supported by a research fund from Sichuan International Studies University (SISU), China (Grant No: SISUWYJY201919).

ACKNOWLEDGMENTS

The first author wishes to thank Nanyang Technological University (NTU), Singapore, for supporting this publication through the Doctorate Scholarship of NTU-RSS.

REFERENCES

AIIC (2020a). Covid-19 Distance nterpreting Recommendations for Institutions and DI Hubs. Available at: https://aiic.org/document/4839/AIIC%20Recommendations%20for%20Institutions_27.03.2020.pdf (Accessed May 30, 2021).

AIIC (2020b). AIIC Guidelines for Distance Interpreting (Version 1.0). Available at: https://aiic.org/document/4418/AIIC%20Guidelines%20for%20Distance%20Interpreting%20Version%201.0 (Accessed May 30, 2021).

AIIC (2021). AIIC and distance interpreting. Available at: https://aiic.org/site/world/about/profession/distanceinterpreting (Accessed May 30, 2021).

AIIC (2022). AIIC position on distance interpreting. Available at: https://aiic.org/document/1031/AIIC-Position-on-Distance-Interpreting-05032018.pdf (Accessed February 15, 2022).

Ait Ammour, H. (2021). Onsite interpreting versus remote interpreting in the COVID 19 world. *Rev. Appl. Linguis.* 5, 339–344.

Al-Salman, S., and Al-Khanji, R. I. (2002). The native language factor in simultaneous interpretation in an Arabic/English context. *Meta* 47, 607–626. doi: 10.7202/008040ar

Amos, R. M., Seeber, K. G., and Pickering, M. J. (2022). Prediction during simultaneous interpreting: evidence from the visual-world paradigm. *Cognition* 220:104987. doi: 10.1016/j.cognition.2021.104987

Andersen, M. S., and Makransky, G. (2021). The validation and further development of the multidimensional cognitive load scale for physical and online lectures (MCLS-POL). *Front. Psychol.* 12:642084. doi: 10.3389/fpsyg.2021.642084

Angelelli, C., and Degueldre, C. (2002). "Bridging the gap between language for general purposes and language for work: An intensive superior level language/skill course for teachers, translators, and interpreters," in *Developing Professional-Level Language Proficiency*. eds. B. L. Leaver and B. Shekhtman (Cambridge: Cambridge University Press), 91–110.

Arbaugh, J. B., and Benbunan-Fich, R. (2007). The importance of participant interaction in online environments. *Decision Support Syst.* 43, 853–865. doi: 10.1016/j.dss.2006.12.013

Aryadoust, V., Foo, S., and Ng, L. Y. (2020a). What can gaze behaviors, neuroimaging data, and test scores tell us about test method effects and cognitive load in listening assessments? *Lang. Test.* 39, 56–89. doi: 10.1177/02655322211026876

Aryadoust, V., Ng, L. Y., Foo, S., and Esposito, G. (2020b). A neurocognitive investigation of test methods and gender effects in listening assessment. *Comput. Assist. Lang. Learn.* 35, 743–763. doi: 10.1080/09588221.2020.1744667

Ayres, P., Lee, J. Y., Paas, F., and van Merriënboer, J. J. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Front. Psychol.* 12:702538. doi: 10.3389/fpsyg.2021.702538

Azarmina, P., and Wallace, P. (2005). Remote interpretation in medical encounters: a systematic review. *J. Telemed. Telecare* 11, 140–145. doi: 10.1258/1357633053688679

Bachman, L. F., and Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (1992). Working memory. *Science* 255, 556–559. doi: 10.1126/science.1736359

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29. doi: 10.1146/annurev-psych-120710-100422

Baddeley, A. D., and Hitch, G. (1974). "Working memory," in *The Psychology of Learning and Motivation*. ed. G. A. Bower (New York, NY: Academic Press), 47–89.

Bae, M., and Jeong, C. J. (2021). The role of working memory capacity in interpreting performance: an exploratory study with student interpreters. *Transl. Cognit. Behav.* 4, 26–46. doi: 10.1075/tcb.00050.bae

Barik, H. C. (1973). Simultaneous interpretation: temporal and quantitative data. *Lang. Speech* 16, 237–270. doi: 10.1177/002383097301600307

Becker, M., Schubert, T., Strobach, T., Gallinat, J., and Kühn, S. (2016). Simultaneous interpreters vs. professional multilingual controls: group differences in cognitive control as well as brain structure and function. *NeuroImage* 134, 250–260. doi: 10.1016/j.neuroimage.2016.03.079

Böcker, M., and Anderson, D. (1993). "Remote conference interpreting using ISDN videotelephone: a requirements analysis and feasibility study," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 37* (Thousand Oaks, CA: SAGE Publications), 235–239.

Bontempo, K., and Napier, J. (2011). Evaluating emotional stability as a predictor of interpreter competence and aptitude for interpreting. *Interpreting* 13, 85–105. doi: 10.1075/intp.13.1.06bon

Bower, K. (2015). Stress and burnout in video relay service (VRS) interpreting. *J. Interpretation* 24:2

Braun, S. (2007). Interpreting in small-group bilingual videoconferences: challenges and adaptation processes. *Interpreting* 9, 21–46. doi: 10.1075/intp.9.1.03bra

Braun, S. (2013). Keep your distance? Remote interpreting in legal proceedings: a critical assessment of a growing practice1. *Interpreting* 15, 200–228. doi: 10.1075/intp.15.2.03bra

Braun, S. (2015). "Remote interpreting," in *Routledge Handbook of Interpreting*. eds. H. Mikkelson and R. Jourdenais (London: Routledge), 352–367.

Braun, S. (2017). What a micro-analytical investigation of additions and expansions in remote interpreting can tell us about interpreters' participation in a shared virtual space. *J. Pragmat.* 107, 165–177. doi: 10.1016/j.pragma.2016.09.011

Braun, S. (2020). "Technology, interpreting," in *Routledge Encyclopedia of Translation Studies*. eds. M. Baker and G. Saldanha (London: Routledge), 569–574.

Braun, S., Davitti, E., and Dicerto, S. (2018). "Video-mediated interpreting in legal settings: assessing the implementation," in *Here or There: Research on Interpreting via video link*. eds. J. Napier, R. Skinner and S. Braun (Washington, DC: Gallaudet University Press), 144–179.

Braun, S., and Taylor, J. (2012). "Video-mediated interpreting: an overview of current practice and research," in *Videoconference and Remote Interpreting in Criminal Proceedings*. eds. S. Braun and J. Taylor (Antwerp: Intersentia), 33–68.

Bravo, E. A. F. (2019). Metacognitive self-perception in interpreting. *Translation, Cognition & Behavior* 2, 147–164. doi: 10.1075/tcb.00025.fer

Byrne, A. (2013). Mental workload as a key factor in clinical decision making. *Adv. Health Sci. Educ.* 18, 537–545. doi: 10.1007/s10459-012-9360-5

Campbell, S., and Hale, S. (2003). "Translation and interpreting assessment in the context of educational measurement," in *Translation Today: Trends and Perspectives*. eds. G. M. Anderman and M. Rogers (Clevedon: Multilingual Matters), 205–224.

Cañada, M. D., and Arumí, M. (2012). Self-regulating activity: use of metacognitive guides in the interpreting classroom. *Educ. Res. Eval.* 18, 245–264. doi: 10.1080/13803611.2012.661934

Canale, M., and Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Appl. Linguis.* 1, 1–47. doi: 10.1093/applin/1.1.1

Chandler, P. (2018). "Forward," in *Cognitive load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice*. ed. R. Z. Zheng (London: Routledge), viii–x.

Chang, C., and Schallert, D. L. (2007). The impact of directionality on Chinese/English simultaneous interpreting. *Interpreting* 9, 137–176. doi: 10.1075/intp.9.2.02cha

Chen, S. (2017). The construct of cognitive load in interpreting and its measurement. *Perspectives* 25, 640–657. doi: 10.1080/0907676X.2016.1278026

Chen, S. (2018). A pen-eye-voice approach towards the process of note-taking and consecutive interpreting: An experimental design. *Int. J. Comp. Lit. Translat. Stud.* 6, 1–6. doi: 10.7575/aiac.ijclts.v.6n.2p.1

Chen, S. (2020). The impact of directionality on the process and product in consecutive interpreting between Chinese and English: evidence from pen recording and eye tracking. *J. Specialised Transl.* 34, 100–117.

Chiang, Y. N. (2009). Foreign language anxiety in Taiwanese student interpreters. *Meta* 54, 605–621. doi: 10.7202/038318ar

Chittleborough, G. D., and Treagust, D. F. (2009). Why models are advantageous to learning science. *Educación química* 20, 12–17. doi: 10.1016/S0187-893X(18)30003-X

Chmiel, A. (2008). Boothmates forever?—On teamwork in a simultaneous interpreting booth. *Across Lang. Cult.* 9, 261–276. doi: 10.1556/Acr.9.2008.2.6

Chmiel, A. (2018). In search of the working memory advantage in conference interpreting–training, experience and task effects. *Int. J. Biling.* 22, 371–384. doi: 10.1177/1367006916681082

Chmiel, A., Janikowski, P., and Lijewska, A. (2020). Multimodal processing in simultaneous interpreting with text: interpreters focus more on the visual than the auditory modality. *Targets* 32, 37–58. doi: 10.1075/target.18157.chm

Chmiel, A., and Lijewska, A. (2019). Syntactic processing in sight translation by professional and trainee interpreters: professionals are more time-efficient while trainees view the source text less. *Targets* 31, 378–397. doi: 10.1075/target.18091.chm

Choi, H. H., Van Merriënboer, J. J., and Paas, F. (2014). Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load. *Educ. Psychol. Rev.* 26, 225–244. doi: 10.1007/s10648-014-9262-6

Christoffels, I. K., and De Groot, A. M. (2004). Components of simultaneous interpreting: comparing interpreting with shadowing and paraphrasing. *Bilingualism* 7, 227–240. doi: 10.1017/S1366728904001609

Christoffels, I., and De Groot, A. (2009). *Simultaneous Interpreting. Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 454–479). Oxford: Oxford University Press.

Cierniak, G., Scheiter, K., and Gerjets, P. (2009). Explaining the split-attention effect: is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Comput. Hum. Behav.* 25, 315–324. doi: 10.1016/j.chb.2008.12.020

Collard, C., and Defrancq, B. (2019). Predictors of ear-voice span, a corpus-based study with special reference to sex. *Perspectives* 27, 431–454. doi: 10.1080/0907676X.2018.1553199

Costa, B., Gutiérrez, R. L., and Rausch, T. (2020). Self-care as an ethical responsibility: A pilot study on support provision for interpreters in human crises. *Transl. Interpreting Stud.* 15, 36–56. doi: 10.1075/tis.20004.cos

Darò, V., and Fabbro, F. (1994). Verbal memory during simultaneous interpretation: effects of phonological interference. *Appl. Linguis.* 15, 365–381. doi: 10.1093/applin/15.4.365

Davitti, E., and Braun, S. (2020). Analysing interactional phenomena in video remote interpreting in collaborative settings: implications for interpreter education. *Interpreter Translator Trainer* 14, 279–302. doi: 10.1080/1750399X.2020.1800364

DeLeeuw, K. E., and Mayer, R. E. (2008). A comparison of three measures of cognitive load: evidence for separable measures of intrinsic, extraneous, and germane load. *J. Educ. Psychol.* 100, 223–234. doi: 10.1037/0022-0663.100.1.223

Desmet, B., Vandierendonck, M., and Defrancq, B. (2018). "Simultaneous interpretation of numbers and the impact of technological support," in *Interpreting and Technology*. ed. C. Fantinuoli (Language Science Press), 13–27.

Díaz-Galaz, S. (2011). The effect of previous preparation in simultaneous interpreting: preliminary results. *Across Lang. Cult.* 12, 173–191. doi: 10.1556/Acr.12.2011.2.3

Díaz-Galaz, S., Padilla, P., and Bajo, M. T. (2015). The role of advance preparation in simultaneous interpreting: a comparison of professional interpreters and interpreting students. *Interpreting* 17, 1–25. doi: 10.1075/intp.17.1.01dia

Dong, Y., Liu, Y., and Cai, R. (2018). How does consecutive interpreting training influence working memory: a longitudinal study of potential links between the two. *Front. Psychol.* 9:875. doi: 10.3389/fpsyg.2018.00875

Dragsted, B., and Hansen, I. (2009). Exploring translation and interpreting hybrids. The case of sight translation. *Meta* 54, 588–604. doi: 10.7202/038317ar

Ehrensberger-Dow, M., Albl-Mikasa, M., Andermatt, K., Heeb, A. H., and Lehr, C. (2020). Cognitive load in processing ELF: translators, interpreters, and other multilinguals. *J. English Lingua Franca* 9, 217–238. doi: 10.1515/jelf-2020-2039

Elmer, S., Klein, C., Kühnis, J., Liem, F., Meyer, M., and Jäncke, L. (2014). Music and language expertise influence the categorization of speech and musical sounds: behavioral and electrophysiological measurements. *J. Cogn. Neurosci.* 26, 2356–2369. doi: 10.1162/jocn_a_00632

Ericsson, K. A., and Simon, H. A. (1998). How to study thinking in everyday life: contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind Cult. Act.* 5, 178–186. doi: 10.1207/s15327884mca0503_3

Fantinuoli, C. (2018). "Computer-assisted interpreting: challenges and future perspectives," in *Trends in E-tools and Resources for Translators and Interpreters*. eds. G. Corpas Pastor and I. Durán-Muñoz (Leiden: Brill), 153–174.

Firth, A. (1996). The discursive accomplishment of normality: on 'lingua franca' English and conversation analysis. *J. Pragmat.* 26, 237–259. doi: 10.1016/0378-2166(96)00014-8

Flavell, J. H. (1976). "Metacognitive aspects of problem solving," in *The Nature of Intelligence*. ed. L. Resnick (Mahwah, NJ: Lawrence Erlbaum Associates), 231–236.

Fujita, R. (2021). The role of speech-in-noise in Japanese EFL learners' listening comprehension process and their use of contextual information. *Int. J. Listening* 36, 118–137. doi: 10.1080/10904018.2021.1963252

Gany, F., Kapelusznik, L., Prakash, K., Gonzalez, J., Orta, L. Y., Tseng, C. H., et al. (2007). The impact of medical interpretation method on time and errors. *J. Gen. Intern. Med.* 22, 319–323. doi: 10.1007/s11606-007-0361-7

Gerver, D. (1975). A psychological approach to simultaneous interpretation. *Meta* 20, 119–128. doi: 10.7202/002885ar

Gerver, D. (1976). "Empirical studies of simultaneous interpretation: a review and a model," in *Translation: Applications and Research*. ed. R. W. Brislin (New York, NY: Gardiner), 165–207.

Gilbert, J. K. (2004). Models and modelling: routes to more authentic science education. *Int. J. Sci. Math. Educ.* 2, 115–130. doi: 10.1007/s10763-004-3186-4

Gile, D. (1995). *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam: John Benjamins Publishing.

Gile, D. (1999). Testing the effort models' tightrope hypothesis in simultaneous interpreting – a contribution. *Hermes* 23, 153–172.

Gile, D. (2008). Local cognitive load in simultaneous interpreting and its implications for empirical research. *Forum* 6, 59–77. doi: 10.1075/forum.6.2.04gil

Gile, D. (2009). *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam: John Benjamins Publishing.

Gluszek, A., and Dovidio, J. F. (2010). The way they speak: a social psychological perspective on the stigma of nonnative accents in communication. *Personal. Soc. Psychol. Rev.* 14, 214–237. doi: 10.1177/1088868309359288

Gracia-García, R. A. (2002). "Telephone interpreting: a review of pros and cons," in *Proceedings of the 43rd Annual Conference*. ed. B. Scott (Alexandria: American Translators Association), 195–216.

Hart, S. G., and Staveland, L. E. (1988). "Development of NASA-TLX (task load index): results of empirical and theoretical research," in *Human mental Workload*. eds. P. A. Hancock and N. Meshkati (Amsterdam: North-Holland), 139–183.

Havelka, I. (2020). Video-mediated remote interpreting in healthcare: analysis of an Austrian pilot project. *Babel* 66, 326–345. doi: 10.1075/babel.00156.hav

Hervais-Adelman, A., Moser-Mercer, B., and Golestani, N. (2018). Commentary: Broca pars triangularis constitutes a "hub" of the language-control network during simultaneous language translation. *Front. Hum. Neurosci.* 12:22. doi: 10.3389/fnhum.2018.00022

Huddle, A. P., White, M. D., and Rogers, F. (2000). Using a teaching model to correct known misconceptions in electrochemistry. *J. Chem. Educ.* 77, 104–110. doi: 10.1021/ed077p104

Hymes, D. (1972). "On communicative competence," in *Sociolinguistics*. eds. J. Pride and J. Holmes (Harmondsworth: Penguin Books), 269–293.

Hymes, D. (1974). *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia, PA: University of Pennsylvania Press.

Hyönä, J., Tommola, J., and Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Q. J. Exp. Psychol.* 48, 598–612. doi: 10.1080/14640749508401407

Ivanova, A. (2000). "The use of retrospection in research on simultaneous interpreting," in *Tapping and Mapping the Processes of Translation and Interpreting: Outlooks*. eds. S. Tirkkonen-Condit and R. Jääskeläinen (Benjamins Translation Library), 27–52.

Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford: Oxford University Press.

Jiménez-Ivars, A. (2021). Telephone interpreting for asylum seekers in the US: a corpus-based study. *J. Specialised Transl.* 36a, 125–146.

Jones, D., Gill, P., Harrison, R., Meakin, R., and Wallace, P. (2003). An exploratory study of language interpretation services provided by videoconferencing. *J. Telemed. Telecare* 9, 51–56. doi: 10.1258/135763303321159701

Kahneman, D. (1973). *Attention and Effort* Prentice-Hall.

Kalina, S. (2000). Interpreting competences as a basis and a goal for teaching. *Interpreters' Newsletter* 14, 3–32.

Kalina, S. (2002). "Quality in interpreting and its prerequisites a framework for a comprehensive view," in *Interpreting in the 21st Century: Challenges and Opportunities*. eds. G. Garzone and M. Viezzi (Amsterdam: John Benjamins), 121–130.

Kalina, S. (2005). Quality assurance for interpreting processes. *Meta* 50, 768–784. doi: 10.7202/011017ar

Kalyuga, S., and Plass, J. L. (2018). "Cognitive load as a local characteristic of cognitive processes: implications for measurement approaches" in *Cognitive load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice*. ed. R. Z. Zheng (New York, NY: Routledge), 59–74.

Ketola, A. (2015). Translation diaries of an illustrated technical text: translation students' conceptions of word-image interaction. *Connexions: Int. Prof. Commun. J.* 3, 13–40.

Kirchhoff, H. (1976). "Simultaneous interpreting: interdependence of variables in the interpreting process, interpreting models and interpreting strategies," in *The Interpreting Studies reader*. eds. F. Pöchhacker and M. Schlesinger (London: Routledge), 111–119.

Ko, L. (2006). The need for long-term empirical studies in remote interpreting research: a case study of telephone interpreting. *Linguistica Antverpiensia, New Series–Themes in Translation Studies* 5, 325–338. doi: 10.52034/lanstts.v5i.167

Korpal, P. (2016). Interpreting as a stressful activity: physiological measures of stress in simultaneous interpreting. *Poznan Stud. Contemp. Ling.* 52, 297–316. doi: 10.1515/psicl-2016-0011

Korpal, P., and Jasielska, A. (2019). Investigating interpreters' empathy: are emotions in simultaneous interpreting contagious? *Targets* 31, 2–24. doi: 10.1075/target.17123.kor

Korpal, P., and Stachowiak-Szymczak, K. (2018). The whole picture: Processing of numbers and their context in simultaneous interpreting. *Pozn. Stud. Contemp. Linguist.* 54, 335–354. doi: 10.1515/psicl-2018-0013

Korpal, P., and Stachowiak-Szymczak, K. (2019). Combined problem triggers in simultaneous interpreting: Exploring the effect of delivery rate on processing and rendering numbers. *Perspect.* 28, 126–143. doi: 10.1080/0907676X.2019.1628285

Korpal, P., and Stachowiak-Szymczak, K. (2020). Combined problem triggers in simultaneous interpreting: exploring the effect of delivery rate on processing and rendering numbers. *Perspect.* 28, 126–143. doi: 10.1080/0907676X.2019.1628285

Koskinen, K., and Ruokonen, M. (2017). "Love letters or hate mail? Translators' technology acceptance in the light of their emotional narratives" in *Human Issues in Translation Technology*. ed. D. Kenny (London: Routledge), 26–42.

Krell, M., Xu, K. M., Rey, G. D., and Paas, F. (2022). "Editorial: recent approaches for assessing cognitive load from a validity perspective" in *Recent Approaches for Assessing Cognitive load from a Validity Perspective*. eds. M. Krell, K. M. Xu, G. D. Rey and F. Paas (Lausanne, Frontiers Media SA), 4–10.

Kuang, H., and Zheng, B. (2022). Note-taking effort in video remote interpreting: effects of source speech difficulty and interpreter work experience. *Perspectives*, 1–21. doi: 10.1080/0907676X.2022.2053730

Kurz, I. (1994). "A look into the "black box"- EEC probability mapping during mental simultaneous interpreting" in *Translation Studies – An Interdiscipline*. eds. M. Snell-Hornby, M. Pochhacker and K. Kaindl (Amsterdam: John Benjamins), 199–213.

Kurz, I. (1995). Watching the brain at work-an exploratory study of EEG changes during simultaneous interpreting (SI). *Interpreters' Newsletter* 6, 3–16.

Lamberger-Felber, H. (2001). Text-oriented research into interpreting-examples from a case-study. *HERMES-J. Lang. Commun. Bus.* 14, 39–64. doi: 10.7146/hjlcb.v14i26.25638

Lee, J., Aryadoust, V., Ng, L. Y., and Foo, S. (2020). A neurocognitive comparison of listening to academic lectures and natural sounds: implications for test validity. *Int. J. Listening*, 1–15. doi: 10.1080/10904018.2020.1818565

Leppink, J., Paas, F., van der Vleuten, C. P. M., van Gog, T., and van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behav. Res. Methods* 45, 1058–1072. doi: 10.3758/s13428-013-0334-1

Leung, C. (2005). Language and content in bilingual education. *Linguist. Educ.* 16, 238–252. doi: 10.1016/j.linged.2006.01.004

Li, X. (2018). Teaching beliefs and learning beliefs in translator and interpreter education: an exploratory case study. *Interpreter Translator Trainer* 12, 132–151. doi: 10.1080/1750399X.2017.1359764

Licoppe, C., and Veyrier, C. A. (2017). How to show the interpreter on screen? The normative organization of visual ecologies in multilingual courtrooms with video links. *J. Pragmat.* 107, 147–164. doi: 10.1016/j.pragma.2016.09.012

Lin, Y., Lv, Q., and Liang, J. (2018). Predicting fluency with language proficiency, working memory, and directionality in simultaneous interpreting. *Front. Psychol.* 9:1543. doi: 10.3389/fpsyg.2018.01543

Linton, P. M., Plamondon, B. D., Dick, A. O., Bittner, A. C., and Christ, R. E. (1989). "Operator workload for military system acquisition," in *Applications of Human Performance Models to System design*. eds. G. R. McMillan, D. Beevis and Breda L. Van, Cham: Springer.

Locatis, C., Williamson, D., Gould-Kabler, C., Zone-Smith, L., Detzler, I., Roberson, J., et al. (2010). Comparing in-person, video, and telephonic medical interpretation. *J. Gen. Intern. Med.* 25, 345–350. doi: 10.1007/s11606-009-1236-x

Low, A. R. L., and Aryadoust, V. (2021). Investigating test-taking strategies in listening assessment: a comparative study of eye-tracking and self-report questionnaires. *Int. J. Listening*, 1–20. doi: 10.1080/10904018.2021.1883433

Luximon, A., and Goonetilleke, R. S. (2001). Simplified subjective workload assessment technique. *Ergonomics* 44, 229–243. doi: 10.1080/00140130010000901

Ma, X., and Li, D. (2021). A cognitive investigation of 'chunking'and 'reordering'for coping with word-order asymmetry in English-to-Chinese sight translation: evidence from an eye-tracking study. *Interpreting* 23, 192–221. doi: 10.1075/intp.00057.ma

Ma, X., Li, D., and Hsu, Y. Y. (2021). Exploring the impact of word order asymmetry on cognitive load during Chinese–English sight translation: evidence from eye-movement data. *Targets* 33, 103–131. doi: 10.1075/target.19052.ma

Martin, S. (2018). "A critical analysis of the theoretical construction and empirical measurement of cognitive load," in *Cognitive load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice*. ed. R. Z. Zheng (London: Routledge), 29–44.

Mazza, C. (2001). Numbers in simultaneous interpretation. *Interpreters' Newsletter* 14, 87–104.

McNamara, D. S., de Vega, M., and O'Reilly, T. (2007). "Comprehension skill, inference making, and the role of knowledge," in *Higher level Language Processes in the brain: Inference and Comprehension Processes*. eds. F. Schmalhofer and C. A. Perfetti (Mahwah, NJ: Erlbaum), 233–253.

Mellinger, C. D., and Hanson, T. A. (2018). Interpreter traits and the relationship with technology and visibility. *Transl. Interpret. Stud.* 13, 366–392. doi: 10.1075/tis.00021.mel

Mellinger, C. D., and Hanson, T. A. (2019). Meta-analyses of simultaneous interpreting and working memory. *Interpreting* 21, 165–195. doi: 10.1075/intp.00026.mel

Meshkati, N. (1988). "Toward development of a cohesive model of workload," in *Human Mental Workload*. eds. P. A. Hancock and N. Meshkati (Amsterdam: North–Holland), 305–314.

Meuleman, C., and Van Besien, F. (2009). Coping with extreme speech conditions in simultaneous interpreting. *Interpreting* 11, 20–34. doi: 10.1075/intp.11.1.03meu

Mizuno, A. (2005). Process model for simultaneous interpreting and working memory. *Meta* 50, 739–752. doi: 10.7202/011015ar

Moray, N. (1967). Where is capacity limited: A survey and a model. *Acta Psychol.* 27, 84–92.

Moser, B. (1978). "Simultaneous Interpretation: a hypothetical model and its practical application," in *Language Interpretation and Communication. Proceedings of the NATO Symposium*. eds. D. Gerver and H. W. Sinaiko (New York, NY: Plenum Press), 353–368.

Moser-Mercer, B. (1997). "Skill components in simultaneous interpreting," in *Conference Interpreting: Current Trends in Research*. eds. Y. Gambier, D. Gile and C. Taylor (Amsterdam: John Benjamins), 133–148.

Moser-Mercer, B. (2000). Simultaneous interpreting: cognitive potential and limitations. *Interpreting* 5, 83–94. doi: 10.1075/intp.5.2.03mos

Moser-Mercer, B. (2003). Remote interpreting: Assessment of human factors and performance parameters. Available at: https://aiic.net/page/1125/remote-interpreting-assessment-of-human-factors-and-pe/lang/1 (Accessed July 6, 2021).

Moser-Mercer, B. (2005a). Remote interpreting: the crucial role of presence. *Bull. VALS-ASLA* 81, 73–97.

Moser-Mercer, B. (2005b). Remote interpreting: issues of multi-sensory integration in a multilingual task. *Meta* 50, 727–738. doi: 10.7202/011014ar

Moser-Mercer, B., Frauenfelder, U. H., Casado, B., and Kunzli, A. (2000). "Searching to define expertise in interpreting," in *Language Processing and Simultaneous Interpreting*. eds. B. Englund Dimitrova and K. Hyltenstam (Amsterdam: John Benjamins), 107–131.

Mouzourakis, P. (1996). Videoconferencing: techniques and challenges. *Interpreting* 1, 21–38. doi: 10.1075/intp.1.1.03mou

Mouzourakis, P. (2003). That feeling of being there: vision and presence in remote interpreting. Available at: http://www.aiic.net/ViewPage.cfm?page_id=1173 (Accessed July 6, 2021).

Mouzourakis, P. (2006). Remote interpreting: a technical perspective on recent experiments. *Interpreting* 8, 45–66. doi: 10.1075/intp.8.1.04mou

Muñoz, E., Calvo, N., and García, A. M. (2018). Grounding translation and interpreting in the brain: what has been, can be, and must be done. *Perspectives* 27, 483–509. doi: 10.1080/0907676X.2018.1549575

Napier, J. (2002). University interpreting: Linguistic issues for consideration, The Journal of Deaf Studies and Deaf Education. *The Journal of Deaf Studies and Deaf Education* 7, 281–301. doi: 10.1093/deafed/7.4.281

Napier, J., Skinner, R., and Braun, S. (Eds.). (2018). *Here or There: Research on Interpreting Via Video Link*. Washington, DC: Gallaudet University Press.

Oléron, P., and Nanpon, H. (1965). Recherches sur la traduction simultanée [studies of simultaneous translation]. *J. Psychol. Norm. Pathol.* 62, 73–94.

Ouwehand, K., van der Kroef, A., Wong, J., and Paas, F. (2022). Measuring cognitive load: are there more valid alternatives to Likert rating scales? *Front. Psychol.* 6, 146–158. doi: 10.3389/feduc.2021.702616

Oviatt, S. L., and Cohen, P. R. (1992). Spoken language in interpreted telephone dialogues. *Comput. Speech Lang.* 6, 277–302. doi: 10.1016/0885-2308(92)90021-U

Ozolins, U. (2011). Telephone interpreting: understanding practice and identifying research needs. *Transl. Interpreting* 3, 33–47.

Paas, F., Tuovinen, J. E., Tabbers, H., and Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educ. Psychol.* 38, 63–71. doi: 10.1207/S15326985EP3801_8

Paas, F. G. W. C., and Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educ. Psychol. Rev.* 6, 351–371. doi: 10.1007/BF02213420

Paas, F., and van Merriënboer, J. J. (2020). Cognitive-load theory: methods to manage working memory load in the learning of complex tasks. *Curr. Dir. Psychol. Sci.* 29, 394–398. doi: 10.1177/0963721420922183

Parsons, K., McCormac, A., Butavicius, M., Pattinson, M., and Jerram, C. (2014). Determining employee awareness using the human aspects of information security questionnaire (HAIS-Q). *Comput. Secur.* 42, 165–176. doi: 10.1016/j.cose.2013.12.003

Pio, S. (2003). The relation between ST delivery rate and quality in simultaneous interpretation. *Interpreters' Newsl.* 14, 69–100.

Plevoets, K., and Defrancq, B. (2018). The cognitive load of interpreters in the European Parliament: a corpus-based study of predictors for the disfluency uh (m). *Interpreting* 20, 1–32. doi: 10.1075/intp.00001.ple

Pöchhacker, F. (2016). *Introducing Interpreting Studies. 2nd Edn*. London: Routledge.

Powell, M. B., Manger, B., Dion, J., and Sharman, S. J. (2017). Professionals' perspectives about the challenges of using interpreters in child sexual abuse interviews. *Psychiatry Psychol. Law* 24, 90–101. doi: 10.1080/13218719.2016.1197815

Prandi, B. (2018). "An exploratory study on CAI tools in simultaneous interpreting: theoretical framework and stimulus validation," in *Interpreting and Technology*. ed. C. Fantinuoli (Berlin: Language Science Press), 29–59.

Pretorius, A., and Cilliers, P. J. (2007). Development of a mental workload index: a systems approach. *Ergonomics* 50, 1503–1515. doi: 10.1080/00140130701379055

Price, C. J., Green, D. W., and von Studnitz, R. (1999). A functional imaging study of translation and language switching. *Brain* 122, 2221–2235. doi: 10.1093/brain/122.12.2221

Reid, G. B., and Nygren, T. E. (1988). "The subjective workload assessment technique: a scaling procedure for measuring mental workload," in *Human Mental Workload*. eds. P. A. Hancock and N. Meshkati (Amsterdam: North-Holland), 185–218.

Riccardi, A. (1998). "Interpreting strategies and creativity," in *Translators' Strategies and Creativity: Selected Papers from the 9th International Conference on Translation and Interpreting*. eds. A. Beylard-Ozeroff, J. Králová and B. Moser-Mercer (Amsterdam: Benjamins Translation Library), 171–180.

Riccardi, A. (2005). On the evolution of interpreting strategies in simultaneous interpreting. *Meta* 50, 753–767. doi: 10.7202/011016ar

Rinne, J. O., Tommola, J., Laine, M., Krause, B. J., Schmidt, D., Kaasineni, V., et al. (2000). The translating brain: cerebral activation patterns during simultaneous interpreting. *Neurosci. Lett.* 294, 85–88. doi: 10.1016/S0304-3940(00)01540-8

Rojo López, A. M., Foulquié-Rubio, A. I., Espín López, L., and Martínez Sánchez, F. (2021). Analysis of speech rhythm and heart rate as indicators of stress on student interpreters. *Perspectives* 29, 591–607. doi: 10.1080/0907676X.2021.1900305

Rojo López, A. M., and Martín, R. M. (2022). "Translation process research," in *The Routledge Handbook of Translation and Methodology*. eds. F. Zanettin and C. Rundle (London: Routledge), 356–372.

Rosiers, A., Eyckmans, J., and Bauwens, D. (2011). A story of attitudes and aptitudes? Investigating individual difference variables within the context of interpreting. *Interpreting* 13, 53–69. doi: 10.1075/intp.13.1.04ros

Roziner, I., and Shlesinger, M. (2010). Much ado about something remote: stress and performance in remote interpreting. *Interpreting* 12, 214–247. doi: 10.1075/intp.12.2.05roz

Runcieman, A. J. (2020). Community Interpreting and the Covid-19 crisis: Present relevancy and future directions. Tilburg Papers in Culture Studies, 242, 1–22. https://pure.uvt.nl/ws/portalfiles/portal/48995997/TPCS_242_Runcieman.pdf

Schnaubert, L., and Schneider, S. (2022). Analysing the relationship between mental load or mental effort and metacomprehension under different conditions of multimedia design. *Front. Psychol.* 6, 159–177. doi: 10.3389/feduc.2021.648319

Schultheis, H., and Jameson, A. (2004). "Assessing cognitive load in adaptive hypermedia systems: physiological and behavioral methods," in *Adaptive Hypermedia and Adaptive web-based Systems*. eds. Bra P. D. E. De and W. Nejdl (Berlin: Springer) 225–234.

Seeber, K. (2017). "Multimodal processing in simultaneous interpreting," in *The handbook of translation and cognition*. eds J. W. Schwieter and A. Ferreira (Wiley Blackwell), 84–92.

Seeber, K. G. (2011). Cognitive load in simultaneous interpreting: existing theories – new models. *Interpreting.* 13, 176–204. doi: 10.1075/intp.13.2.02see

Seeber, K. G. (2013). Cognitive load in simultaneous interpreting: measures and methods. *Targets* 25, 18–32. doi: 10.1075/target.25.1.03see

Seeber, K. G., Keller, L., Amos, R., and Hengl, S. (2019). Expectations vs. experience: attitudes towards video remote conference interpreting. *Interpreting* 21, 270–304. doi: 10.1075/intp.00030.see

Seeber, K. G., Keller, L., and Hervais-Adelman, A. (2020). When the ear leads the eye–the use of text during simultaneous interpretation. *Lang. Cognit. Neurosci.* 35, 1480–1494. doi: 10.1080/23273798.2020.1799045

Seeber, K. G., and Kerzel, D. (2011). Cognitive load in simultaneous interpreting: model meets data. *Int. J. Biling.* 16, 228–242. doi: 10.1177/1367006911402982

Seleskovitch, D., and Lederer, M. (1989). Pdagogie raisonne de l'interprtation [Reasoned Pedagogy of Interpretation]. Paris: Didier rudition.

Setton, R. (1999). *Simultaneous Interpretation: A Cognitive-pragmatic Analysis*. Amsterdam: Benjamins Publishing.

Setton, R. (2001). Deconstructing SI: a contribution to the debate on component processes. *Interpreters' Newsl.* 11, 1–26.

Setton, R. (2002). Pragmatic analysis as a methodology: a reply to Gile's review of Setton (1999). *Targets* 14, 353–360. doi: 10.1075/target.14.2.08set

Shlesinger, M. (2000). "Interpreting as a cognitive process: how can we know what really happens?" in *Tapping and Mapping the Processes of Translation and Interpreting: Outlooks*. eds. S. Tirkkonen-Condit and R. Jääskeläinen (Amsterdam: Benjamins Translation library), 3–16.

Skinner, R., Napier, J., and Braun, S. (2018). "Interpreting via video link: mapping of the field," in *Here or there: Research on Interpreting via Video Link*. eds. J. Napier, R. Skinner and S. Braun (Washington DC: Gallaudet University Press), 11–35.

Su, W., and Li, D. (2019). Identifying translation problems in English-Chinese sight translation: an eye-tracking experiment. Translation and interpreting studies. *J. Am. Transl. Interpreting Stud. Assoc.* 14, 110–134. doi: 10.1075/tis.00033.su

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cogn. Sci.* 12, 257–285. doi: 10.1016/0364-0213(88)90023-7

Sweller, J. (2018). "The role of independent measures of load in cognitive load theory," in *Cognitive Load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice*. ed. R. Z. Zheng (London: Routledge), 3–8.

Sweller, J., Ayres, P., and Kalyuga, S. (2011). *Cognitive Load Theory*. Berlin: Springer.

Sweller, J., van Merrienboer, J., and Paas, F. (1998). Cognitive architecture and instructional design. *Educ. Psychol. Rev.* 10, 251–296. doi: 10.1023/A:1022193728205

Sweller, J., van Merriënboer, J. J., and Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* 31, 261–292. doi: 10.1007/s10648-019-09465-5

Timarová, S. (2008). "Working memory and simultaneous interpreting," in *Translation and its Others: Selected Papers of the CETRA Research Seminar in Translation Studies*. ed. P. Boulogne (Belgium: KU Leuven Centre for Translation Studies), 1–28.

Timarová, S., and Salaets, H. (2011). Learning styles, motivation and cognitive flexibility in interpreter training: self-selection and aptitude. *Interpreting* 13, 31–52. doi: 10.1075/intp.13.1.03tim

Tiselius, E., and Sneed, K. (2020). Gaze and eye movement in dialogue interpreting: an eye-tracking study. *Biling. Lang. Congn.* 23, 780–787. doi: 10.1017/S1366728920000309

Tymoczko, M. (2012). The neuroscience of translation. *Target* 24, 83–102. doi: 10.1075/target.24.1.06tym

Van de Putte, E., De Baene, W., García-Pentón, L., Woumans, E., Dijkgraaf, A., and Duyck, W. (2018). Anatomical and functional changes in the brain after simultaneous interpreting training: a longitudinal study. *Cortex* 99, 243–257. doi: 10.1016/j.cortex.2017.11.024

Van Merriënboer, J. J., Kester, L., and Paas, F. (2006). Teaching complex rather than simple tasks: balancing intrinsic and germane load to enhance transfer of learning. *Appl. Cogn. Psychol.* 20, 343–352. doi: 10.1002/acp.1250

Vianna, B. (2005). Simultaneous interpreting: A relevance-theoretic approach. *Intercult. Pragmat.* 2, 169–190. doi: 10.1515/iprg.2005.2.2.169

Vranjes, J., Brône, G., and Feyaerts, K. (2018a). Dual feedback in interpreter-mediated interactions: on the role of gaze in the production of listener responses. *J. Pragmat.* 134, 15–30. doi: 10.1016/j.pragma.2018.06.002

Vranjes, J., Brône, G., and Feyaerts, K. (2018b). On the role of gaze in the organization of turn-taking and sequence organization in interpreter-mediated dialogue. *Lang. Dialogue* 8, 439–467. doi: 10.1075/ld.00025.vra

Vranjes, J., and Oben, B. (2022). Anticipation and timing of turn-taking in dialogue interpreting: a quantitative study using mobile eye-tracking data. *Targets*. doi: 10.1075/target.20121.vra

Wadensjö, C. (1999). Telephone interpreting & the synchronization of talk in social interaction. *Translator* 5, 247–264. doi: 10.1080/13556509.1999.10799043

Wang, J. (2018). "It keeps me on my toes": Interpreters' perceptions of challenges in telephone interpreting and their coping strategies. *Target* 30, 430–462. doi: 10.1075/target.17012.wan

Wang, J., Antonenko, P., Keil, A., and Dawson, K. (2020). Converging subjective and psychophysiological measures of cognitive load to study the effects of instructor-present video. *Mind Brain Educ.* 14, 279–291. doi: 10.1111/mbe.12239

Wen, H., and Dong, Y. (2019). How does interpreting experience enhance working memory and short-term memory: a meta-analysis. *J. Cogn. Psychol.* 31, 769–784. doi: 10.1080/20445911.2019.1674857

Weng, Y., Zheng, B., and Dong, Y. (2022). Time pressure in translation: Psychological and physiological measures. *Target*. doi: 10.1075/target.20148.wen

Wessling, D. M., and Shaw, S. (2014). Persistent emotional extremes and video relay service interpreters. *J. Interpretation* 23, 6.

Wickens, C. D. (1984). "Processing resources in attention" in *Varieties of Attention*. eds. R. Parasuraman and D. R. Davies (New York: Academic Press), 63–102.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theor. Issues Ergon. Sci.* 3, 159–177. doi: 10.1080/14639220210123806

Xie, B., and Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single-and multi-task environments. *Work Stress* 14, 74–99. doi: 10.1080/026783700417249

Yin, B., Chen, F., Ruiz, N., and Ambikairajah, E. (2008). "Speech-based cognitive load monitoring system," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing* (Piscataway, NJ: IEEE), 2041–2044.

Young, M. S., Brookhuis, K. A., Wickens, C. D., and Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics* 58, 1–17. doi: 10.1080/00140139.2014.956151

Young, M. S., and Stanton, N. A. (2005). "Mental workload," in *Handbook of human Factors and Ergonomics Methods*. eds. N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas and H. W. Hendrick (London: Taylor & Francis), 390–401.

Zheng, R. Z. (2018). *Cognitive load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice* London, Routledge.

Zheng, B., and Zhou, H. (2018). 隐喻表达过程：视译中听译时间差的眼动研究 [Revisiting processing time for metaphorical expressions: An eye-tracking study on eye-voice span during sight translation]. *Wai Yu Jiao Xue Yu Yan Jiu [Foreign Language Teaching and Research]* 50, 744–759. doi: 10.1016/j.bandc.2020.105584

Zheng, B., Báez, S., Su, L., Xiang, X., Weis, S., Ibáñez, A., et al. (2020). Semantic and attentional networks in bilingual processing: fMRI connectivity signatures of translation directionality. *Brain Cogn.* 143:105584. doi: 10.1016/j.bandc.2020.105584

Zhu, X., and Aryadoust, V. (In press) (2022). A Scientometric review of translation and interpreting research in the early 21st century. *Targets.*

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 23 August 2022
DOI 10.3389/fpsyg.2022.754732

OPEN ACCESS

EDITED BY
Hong Xu,
Nanyang Technological University,
Singapore

REVIEWED BY
Alexandru Nicolae Ungureanu,
Università di Torino, Italy
Jovan Gardasevic,
University of Montenegro, Montenegro

*CORRESPONDENCE
Christian Krupitzer
christian.krupitzer@uni-hohenheim.de

SPECIALTY SECTION
This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

RECEIVED 06 August 2022
ACCEPTED 01 August 2022
PUBLISHED 23 August 2022

CITATION
Krupitzer C, Naber J, Stauffert J-P, Mayer J, Spielmann J, Ehmann P, Boci N, Bürkle M, Ho A, Komorek C, Heinickel F, Kounev S, Becker C and Latoschik ME (2022) CortexVR: Immersive analysis and training of cognitive executive functions of soccer players using virtual reality and machine learning.
Front. Psychol. 13:754732.
doi: 10.3389/fpsyg.2022.754732

CortexVR: Immersive analysis and training of cognitive executive functions of soccer players using virtual reality and machine learning

Christian Krupitzer[1]*, Jens Naber[2], Jan-Philipp Stauffert[3], Jan Mayer[4], Jan Spielmann[4], Paul Ehmann[4], Noel Boci[2], Maurice Bürkle[2], André Ho[2], Clemens Komorek[2], Felix Heinickel[5], Samuel Kounev[5], Christian Becker[2] and Marc Erich Latoschik[3]

[1]Department Food Informatics and Computational Science Lab, Universität Hohenheim, Stuttgart, Germany, [2]Chair of Information Systems II, Universtiät Mannheim, Mannheim, Germany, [3]HCI Group, Universität Würzburg, Würzburg, Germany, [4]TSG Research Lab, Zuzenhaussen, Germany, [5]Software Enngineering Group, Universität Würzburg, Würzburg, Germany

Goal: This paper presents an immersive Virtual Reality (VR) system to analyze and train Executive Functions (EFs) of soccer players. EFs are important cognitive functions for athletes. They are a relevant quality that distinguishes amateurs from professionals.

Method: The system is based on immersive technology, hence, the user interacts naturally and experiences a training session in a virtual world. The proposed system has a modular design supporting the extension of various so-called game modes. Game modes combine selected game mechanics with specific simulation content to target particular training aspects. The system architecture decouples selection/parameterization and analysis of training sessions *via* a coaching app from an Unity3D-based VR simulation core. Monitoring of user performance and progress is recorded by a database that sends the necessary feedback to the coaching app for analysis.

Results: The system is tested for VR-critical performance criteria to reveal the usefulness of a new interaction paradigm in the cognitive training and analysis of EFs. Subjective ratings for overall usability show that the design as VR application enhances the user experience compared to a traditional desktop app; whereas the new, unfamiliar interaction paradigm does not negatively impact the effort for using the application.

Conclusion: The system can provide immersive training of EF in a fully virtual environment, eliminating potential distraction. It further provides an easy-to-use analyzes tool to compare user but also an automatic, adaptive training mode.

KEYWORDS

sports analytics, virtual worlds training simulations, training of cognitive functions, executive functions, healthcare, machine learning, training

1. Introduction

Principle performance characteristics of many professional sport activities are continuously evolving, for example, professional soccer is characterized by shorter ball possession times, increased need of accurate passing rates, or longer running distances. The average duration of ball possession per contact for soccer players decreased from 2.8 s in 2006 to 1.1 s in 2010 (Carling, 2010). Consequently, players need to train to react faster. However, today's high training load with repeated physical training of athletic skills in combination with an ever increased number of matches in various competition contexts already puts significant physical strain on players. Hence, professional clubs are seeking for alternatives. A promising approach focuses on non-physical training of so-called Executive Functions (EFs) to increase the players' *Soccer IQ* (Ingle, 2016). EFs encompass the cognitive abilities to evaluate and make situational decisions (Strauss et al., 2006). Hence, they are mental determinants of players' behavior and can be an important discriminant for comparing players as well as determining suitable positions for youth players as different positions have different requirements regarding the EFs (Verburgh et al., 2014).

EFs can effectively be increased by certain computer games (Green et al., 2012; Vestberg et al., 2012; Verburgh et al., 2014) which already led to an adoption of similar approaches by professional sports. One of the most elaborated systems, the Helix, is based on Mixed Reality (MR). Here, soccer players stand in the focal center of an 180° large screen display to partly immerse them into a virtual soccer game. During training, players have to execute a multi-object tracking task in form of tracking non-player characters (NPCs) running around on the virtual soccer field. Due to the large size of the room-filling screen, this tasks demands high capabilities in working memory, peripheral vision, and cognitive flexibility. Recently, the Helix has been extend toward a 360° screen-based systems; however, the mentioned shortcomings are still present.

However, the current MR-based Helix has several shortcomings. First, it's NPC logic is very rudimentary and based on random paths. As a consequence, users with a good understanding of the game do not benefit from that. Second, the player has to tell a coach which NPCs should be selected. Thus, the reaction time of a player is not measurable, hence, a detailed analysis of the player's performance taking the trade-off between accuracy and reaction time into account is not feasible. Third, while the Helix already allows a semi-immersion for players based on the large-screen installation, it restricts analysis and training to one user at a time, it is a complex and expensive system, and does not scale well with large teams. The semi-immersiveness of the current approach is potentially disadvantageous. While a high degree of immersion does not necessarily have to be a goal for any VR system (Bowman and McMahan, 2007), it is usually beneficial when it comes to the elicitation of a strong effect of a place illusion and spatial presence. The semi-immersiveness also risks to have players notice the physical space around them, e.g., when the visual attention is directed to the vicinity of screen borders.

This paper presents *CortexVR*, an immersive VR system for the analysis and training of EFs of soccer players[1]. We present an immersive system that aims at eliminating the need of a large screen system that suffers the issue of distraction while providing a whole new type of user interaction and training experience. Bird (2020) showed the applicability of using virtual reality head-mounted displays within applied sport psychology. Accordingly, our system is based on a head-mounted display (HMD) VR kit (the *Oculus Rift*) that effectively shuts out visual and auditory distractors from the physical space around users. This avoids breaks in presence and fosters undisturbed training of EFs. The system platform increases mobility, lowers hardware costs, and allows an increased number of users concurrently training. *CortexVR* incorporates automated assessment of reaction times. The system's modular design fosters extensibility and includes necessary control functions for training personal *via* a dedicated app. The current version includes realistic NPC paths captured from a real professional soccer match in the German Bundesliga. Additionally, we offer a game mode which dynamically adjusts the game level depending on the performance of the user. The system is tested for VR-critical performance criteria and evaluated for the ease of use and user experience. Subjective ratings for overall usability show that the user experience created by using VR increases the user experience while analysis/training of EFs. Additionally, we evaluate the reinforcement learning approach of the adaptive game mode.

The remainder of this paper is structured as follows. Section 2 discusses related work: the relevant theoretical fundamentals of EFs, works in the field of training EFs, and the application of VR in sports. Subsequently, Section 3 covers the analysis of the requirements for the platform based on an analysis of the Helix and stakeholder interviews. Within Section 4, we describe the technology stack of the *CortexVR* platform as well as the adaptive game mode based on reinforcement learning. Section 5 describes the qualitative evaluation with a group of users and the quantitative of the adaptive game mode, followed by the discussion of the results in Section 6. Section 7 concludes this paper.

2. Related work

This section introduces the theoretical foundations of EFs. Following, it presents different approaches for training EFs as well as VR-based sports training in general. Lastly, we discuss the related work and motivate the research gap for this work.

---

[1] A video of the prototype can be found at: https://www.youtube.com/watch?v=gQM2mP5Mjs0.

2.1. Executive functions

Executive Functions (EFs) denote cognitive processes controlling human actions in different environments (Strauss et al., 2006). EFs usually are separated into two distinct types: The work here focuses on basic cognitive processes including cognitive inhibition, working memory, and cognitive flexibility. The higher order EFs (e.g., reasoning and/or problem solving) are not considered.

Cognitive inhibition refers to the blocking out or tuning out of information that is irrelevant to the task at hand (Harnishfeger, 1995). This mental process can be intentional or unintentional and can manifest itself in a variety of ways (Harnishfeger, 1995). The working memory is a conceptualization of human memory of limited capacity responsible for storage and manipulation of information over brief temporal intervals (Baddeley, 2010). Cognitive flexibility refers to the ability to adapt to the transition from thinking and reasoning about one concept to a different one (Scott, 1962).

EFs develop gradually across one's lifespan. They can be improved at any point in life. Neuropsychological tests, as Stroop test and rating scales, are used to measure EFs. Notably, there were several approaches to use VR for the assessment of EF for various use cases (see, e.g., Pugnetti et al., 1998; Lalonde et al., 2013; Climent-Martinez et al., 2014). EFs are considered to have a high relevance for sports. In several experiments soccer players performed better than non-athletes concerning their EFs, especially in decision making tasks (Vestberg et al., 2012; Verburgh et al., 2014). Also for other sports, the training of EFs seems to be promising (Holfelder et al., 2020; Koch and Krenn, 2021). De Waelle et al. (2021) identified that the strength of EFs is higher for team sports player than for self-paced sports and that those effects are already present for young athletes.

2.2. Training of executive functions

2.2.1. VR-based training

Since EFs are highly relevant for the success of a professional soccer player, it is beneficial to train these functions to increase player performance. However, this is not restricted to high performance sports. Kubesch and Walk (2009) discuss the effects of training EFs for children. The training of these functions is often used in treatment of several, especially cognitive, diseases, such as Parkinson's disease (Sammer et al., 2006) or children with Attention Deficit Hyperactivity Disorder (ADHD) (Tamm et al., 2014). It has extensively been explored for neurological deficits in the area of neurorehabilitation (McGeorge et al., 2001; Lo Priore et al., 2003; Weiss et al., 2006). The analysis of the EFs is neglected in all those approaches.

2.2.2. Gamification-based training

Studies showed that action video games can be used to train specific EFs, e.g., gain advantages in task-switching (Green et al., 2012), increase the processing speed (Dye et al., 2009), or enhance the development of perceptual templates (Bejjanki et al., 2014). While these studies show that video games have a positive training effect, they are rarely used systematically to train the performance of professional athletes.

2.3. VR-based sports training

Akbaş et al. (2019) provide an overview of the application of VR for competitive sports. They identified three categories of applications: performance analysis, simulation improvement, and virtual training.

First approaches of VR-based sports training included simple videos as VR lessons (Success Series, 2019) for amateur sports or advanced systems with elaborated special-purpose training equipment aimed at gyms (Icaros, 2019). But VR is not only relevant for the private sports sector; it is also discovered in professional sports for training and analysis. Bideau et al. (2004, 2010) analyzed and trained the movement of handball goal keepers with synthetic VR opponents. Additionally, there are first commercial products for VR sports training emerging. *NeuroTrainer* aims at training neurological and cognitive functions of athletes (NeuroTrainer, 2019). A more physical approach is the *Virtual Goalie* by Reaction VR Sports which makes use of the motion tracking controllers of the HTC Vive and Oculus Rift to train the reaction of lacrosse goalies (Reaction VR Sports, 2019). *Mi Hiepa Sports* introduced a mixed reality approach for soccer training: motion tracking sensors are attached to different parts of the player's body and his movement is fully captured in the virtual environment (Mi Hiepa Sports, 2019). This can be used to train technical skills by studying a 360° perspective. *StriVR* provides a VR tool to train different plays in a high repetition without the physical demand (StriVR, 2019). This tool is dominantly used in American Football to train play calls. A further example is EonSports VR (EonSports VR, 2019), which is mainly used in baseball for technical and tactical training. Besides professional sports training, VR helped to enable exercises of groups in a remotely fashion during the COVID-19 pandemic. For example, Gao et al. (2020) present a study on exercises for health and wellness of older adults during the COVID-19 Pandemic.

2.4. Distinction from related work

The main advantage of VR used for sports analysis and training is the possibility to achieve a high number of repetitions and a good memorization effect without the physical stress of conventional athletics training, as could be illustrated by various applications. Further, as the technology disappears, a new, pervasive user experience is generated which supports the training effects as it avoids distraction. Therefore, it is especially useful during physiotherapy and regeneration, but also as a low-impact add-on to conventional physical training. VR-based training of EFs has shown to be effective in neurorehabilitation (McGeorge et al., 2001; Lo Priore et al., 2003; Weiss et al., 2006) but the current state-of-the-art in professional soccer still relies on gamified approaches. To the best of our knowledge, there currently is no research available which uses VR to analyze/train EFs and applies this to professional soccer as is illustrated by the approach in this paper. Hence, this paper provides the first approach that transfers the analysis/training of EFs in an immersive VR system—which has been proven to be beneficial in neurorehabilitation—into the domain of sports, namely into professional soccer.

3. Requirements analysis

The soccer club TSG Hoffenheim uses the Helix to analyze and train the EFs of professional first league soccer players—both male and female—as well as the youth players. The players have to track NPCs in a simulated game environment. The Helix currently comprises an 180° screen by integrating the stream of six different projectors in a wide-angle screen. A recent extension of the system provides a 360° screen. The Helix app integrates two roles: the coach and the user, i.e., the soccer players that train their EFs. The coach uses a tablet to configure the game settings before starting the game. The user is then confronted with a randomly created sequence of soccer NPCs running on the field. Their task is to tell the coach which NPCs were highlighted in the beginning of the sequence. The coach then selects these NPCs in the application. Figure 1 shows on the left the original Helix system with both roles, coach and user, simultaneously active during the selection of the NPCs[2]. We did an analysis of the limitations of the original Helix system and conducted interviews with stakeholders. Based on this material and the analysis of related approaches, we derived several requirements for a computing system to support immersive cognitive training of EFs for professional soccer players. In the following, we describe these functional requirements.

First, an important requirement is cross-platform support. Using VR, still requires setup and maintenance effort. As the necessary IT skills cannot be presupposed in a soccer club, we decided to design the application in a way that it can be used as immersive VR application but also as normal desktop application with mouse/key support (requirement $R_{App}1$). This way, soccer clubs can study the benefits of immersion for training of EFs but have a backup option in case that the organizational structures cannot guarantee the use of VR devices. However, we suppose that immersion has benefits for training EFs compared to the normal desktop application.

The coach should be able to configure games (requirement $R_{Config}1$), setup the training session (requirement $R_{Config}2$), and analyze the player's performance (requirement $R_{Config}3$). As this functionality neither affects the training itself nor should be running as a VR app, the configuration application should be a stand-alone tool (requirement $R_{Config}4$).

One limitation in the original Helix system is that the coach enters the data collected verbally from the user, resulting in a delay while data input. Hence, the reaction time of a soccer player is not trackable. To overcome this, the user should be able to directly interact with the application (requirement $R_{App}2$) through a controller or gesture-based input.

Soccer players with a good anticipation of the game should benefit from that while tracking. Consequently, we decided to include the option to choose between player movements extracted from real spatiotemporal soccer tracking data or random paths (requirement $R_{App}3.1$). Furthermore, it should be possible to use specific game situations like corner kicks or kick-offs extracted from real soccer game data (requirement $R_{App}3.2$). The reproducibility of those sequences—in contrast to the random paths of the original Helix system—helps to systematically analyze and compare the EFs of players.

In addition to the configuration of the game itself, the *CortexVR* application should also offer new game modes to enhance the gameplay and make it more diversified (requirement $R_{App}4$). The user therefore is not bound to the repetition of the same game but can find challenges in new variants which triggers the gamification aspects. Thus, it is important to strive for modularity (requirement $R_{App}5$)—such as offering modules for (i) generating the game situations, (ii) rendering the game situations, or (iii) evaluating user data—as several functionalities are relevant across game modes. Further, the system needs to follow a design that allows to extend the system with additional game modes (requirement $R_{App}6$).

4. Design and implementation

As the requirements $R_{Config}1$–$R_{Config}3$ demand that game configuration and evaluation are running in a dedicated app (see requirement $R_{Config}4$), it was decided to implement those in the so called *Coach App*. Figure 2 visualizes the interaction between *Coach App* and *CortexVR*. By launching the *Coach App*, the instructor—i.e., a coach or staff member—starts the system. Then, the session configuration takes place through which the instructor decides on the game parameters and stores them to a respective JSON file. In the next step, after starting the Unity3D App and loading the JSON file, the scene

---

[2] A video of the original Helix system is available at: https://www.youtube.com/watch?v=M7Tmg1mmgHw&t=11s.

FIGURE 1
The left part of the figure shows the original *Helix* for training of EFs on a 180° screen. It is a room-filling 180° screen (source: Michael Horeni, 2017). The right part of the figure shows the new *CortexVR* for VR-supported training of EFs with an Oculus Rift used by one of the authors.

is initialized. The original Helix application generates random paths for the movement of NPCs. For generating realistic player movement paths (see requirement $R_{App}3.1$) and extracting clearly defined game situations (see requirement $R_{App}3.2$), we use the spatiotemporal data of a German Bundesliga soccer game. It was recorded by the widely-used tracking system *TRACAB Image Tracking* and provides two-dimensional positions of all players and three-dimensional position of the ball. After playing the configured numbers of sessions, the Unity3D App writes the game log data into a shared SQLite database before it is closed. The instructor can then use the *Coach App* for analyzing the player's performance. In the following, we explain the design and implementation of the *CortexVR* and the *Coach App*. Afterwards, we describe the design of the three game modes.

4.1. Coach App

The *Coach App* enables the management of user and team data, the initialization of training sessions as well as the analysis of user performance. It is a standalone .NET based application implemented in C# using the Windows Presentation Foundation (WPF). Using the UI, the coach can define the training session for a soccer player (see requirement $R_{Config}1$). This includes the definition of the game mode and the level. This way, the coach can define sessions for analysis and comparison of players before starting the app (see requirement $R_{Config}2$).

The *Coach App* integrates an overview for the analysis of a player's performance (see requirement $R_{Config}3$). The game produces session data that captures the user's performance in terms of accuracy and reaction time. Also, the comparison between the performance of the user under consideration and a certain peer group is offered by the *Coach App*. The peer group as well as all parameters of the training sessions can be filtered to ensure a customized analysis.

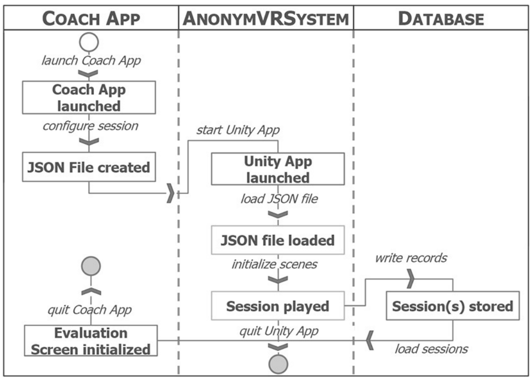


FIGURE 2
Interaction between the *Coach App* and the *CortexVR* application.

4.2. Cortex VR

The *CortexVR* app displays the content of the training session and interacts with the user. It loads the session configuration, starts the game mode, and stores the data for the evaluation of the user's performance. We implemented the *CortexVR* app using the Unity3D Game Engine. Unity3D enables cross-compilation for various platforms including VR devices (see requirement $R_{App}1$).

All game modes use the same control procedure and the same UI. Accordingly, these elements are modularized which fulfills requirement $R_{App}5$. We designed the games modes after discussions with staff and coaches of the soccer club TSG Hoffenheim. For fulfilling the requirement of direct interaction between the app and the user (see requirement $R_{App}2$), we decided to use the *Oculus Rift* VR device as it includes—besides a headset—two gesture-based controllers and a set of sensors called Constellation system which is responsible for the

FIGURE 3
Setup of the evaluation for the mouse/key control-based application used by one of the authors.

tracking of the position of the user's head as well as other VR devices. However, it would be possible to migrate our *CortexVR* application to the Occulus Go and avoid the necessity of having a dedicated base station machine.

We implemented two variants of the application: a low immersive desktop-based version and a high immersive VR version using the Oculus Rift. Both offer the same set of functionality. However, they differ in immersiveness and user interaction. The desktop version relies on mouse and keyboard input (see Figure 3), the VR version uses the input of the controllers and sensors to display a laser beam like pointer for enabling a natural interaction in a virtual environment.

For triggering the gamification effect, we created a realistic environment for the user. As stadium we used an asset from the Unity3D Asset Store. A sound game object that holds a mp3 with stadium sounds in the audio source component simulates a real stadium atmosphere. We created two versions of a NPC using *Adobe Fuse CC*: one with a blue T-shirt and black shorts; the other with a red T-shirt and red shorts. All other visual parameters are the same, including their body specifications and facial features, to increase the level of complexity.

4.3. Game modes

The original Helix system targeted the training of the working memory as one specific sub-function of EFs. However, it only provides a single game mode for this specific purpose. We converted the tracking player mode to our new *CortexVR* system. Supported by our modularized approach (see requirement $R_{App}5$) and as new game modes extend a common abstract class (see requirement $R_{App}6$), it is really convenient to add new game modes to the system by re-using several existing modules. As divergent learning experiences in sports can enhance training of the cognitive executive functions (Buning et al., 2021), we add two additional game modes (see requirement $R_{App}4$). In the following, we present the design of the three implemented game modes.

4.3.1. Tracking players

The *Tracking Players* game mode consists of a fair number of NPCs being targeted at the beginning. Then, the user tries to follow them and pointing them out at the end of the game session. The key difference to the original Helix is that the user can now move within the environment and manipulate the game camera which as a result means that he is not restricted to a frontal view only anymore. Accordingly, we shift the game environment from a 180° screen to a more realistic 360° screen. This widens the field of view, but also offers a new challenge due to the possibility of NPCs being off screen. Hence, the use of the working memory is stressed additionally, as the user has to remember the different positions of the NPCs as well as which of them had been marked in the beginning.

4.3.2. Count players

The *Count Players* game mode aims to improve the EFs of the users by letting them keep track of the count of current NPCs on the screen. As NPCs can appear from the left and right, but also from behind, the eyes are trained to keep the overview on fixed positions of the display instead of following specific objects as in the Tracking Players game mode. This game mode addresses the training of the working memory but also of the inhibition, as the user spontaneously has to decide if a player has been counted already. The camera is fixed which means that moving and looking around is disabled. To vary the difficulty, the position can be adjusted, e.g., inside or near the NPCs but also bird-like views that simplify the task. This game mode does not require soccer knowledge because the user does not need to interpret the movement of NPCs.

4.3.3. Find ball

In the *Find Ball* mode, the user sees a game situation without the ball and needs to estimate the position of the ball. Therefore, the soccer pitch is split into equally sized areas. The user needs to analyze the overall formation and position of the NPCs. The game situations are extracted from real game data, accordingly, soccer-related knowledge helps to analyze the game situation. However, animations like shooting and passing are not available which raises the difficulty. This game mode supports the training of cognitive flexibility, as it targets the ability of players to transfer their game understanding to the unknown situations that miss important information, such as the ball and animations of players' actions.

4.3.4. Reinforcement learning-supported adaptive game mode

The additional *track adaptive* game mode extends the *tracking players* mode and offers the functionality of adapting the speed of the application depending on the performance of the user. Accordingly, this increases the difficulty quickly and supports an improved analysis as players are higher demanded. We decided to rely on reinforcement learning (RL) as this machine learning technique is often applied in artificial game intelligence [e.g., in Backgammon (Tesauro, 1995), Hearts (Sturtevant and White, 2007), or first-person shooters (McPartland and Gallagher, 2011)]. We applied Q-learning for predicting the attribute `Speed` of NPCs using the current state and a list of possible actions. The objective of Q-learning is to learn the quality of actions (policy) for telling an agent which action to perform based on the expected rewards. As state, we consider the speed, the amount of correctly played rounds, the accuracy, and the time. The speed is split into 10 areas, reflecting velocities between 13 and 40 $\frac{\text{km}}{\text{h}}$ (in steps of 3 $\frac{\text{km}}{\text{h}}$). Successful rounds are those in which the user has tracked all NPCs correctly. As we always track two players, the accuracy is 0, 0.5, or 1 for each round. For the time, we assume that users requires at least 1.5 s and ignore all values above 30 s.

$$T(n) = \begin{cases} \lfloor \text{SubmitTime}/3 \rceil - 1 & \text{if } n \leq 30 \\ 9 & \text{if } n > 30 \end{cases} \quad (1)$$

As actions, we focus on adjusting the velocity of NPCs by increasing or reducing the velocity in multiples of 3 $\frac{\text{km}}{\text{h}}$. To keep the speed always between 13 and 40 $\frac{\text{km}}{\text{h}}$, only specific actions are possible. For example, when starting with 13 $\frac{\text{km}}{\text{h}}$, increases can be arbitrarily (9 actions) or the speed stays the same (1 action). For determining the quality of an action, positive or negative rewards are assigned. Our reward function takes into account if the player already played more than one correct round and the correctly played rounds on the current level. Further, it distinguishes if the speed has recently decreased or increased.

5. Evaluation

This section presents the evaluation of our approach for an immersive system to support the analysis and training of EFs. The evaluation is composed of an analysis of important technical properties, a user study as well as a quantitative analysis of the adaptive game mode's RL method. In the following, we describe in consecutive order the evaluation.

5.1. Technical evaluation of *CortexVR*

To guarantee the performance of the VR mode, we conducted motion to photon latency measurements with frame counting.

5.1.1. Methodology

A camera recorded both the physical controller and the screen at 240 frames per second. The latency is the time between the start of a movement of the physical controller until the start of the movement of its virtual counterpart. The ideal case of recording both the Oculus Rift screen and the physical controller is not possible as the lenses render the image unusable. The measurement is conducted in two steps: The motion to photon latency is determined by observing the real controller in front of and the virtual controller on the computer's monitor screen. The reaction time difference between the monitor screen and the Oculus Rift screen is determined by observing how fast they react to color changes spanning the entire image. Frames per second are derived with Unity's unscaledDeltaTime. It indicates for each frame how much time has passed since the previous frame. The technical measures are taken on a desktop PC with Intel Core i7 7,700 k 4 × 4.2 GHz, 16 GB RAM, and Nvidia Geforce GTX 1080.

5.1.2. Results

The measured latency between controller movement and its virtual counterpart on the monitor is 66.6 ms (SD = 21.6 ms). The difference between the monitor screen and the Oculus Rift screen is 32.2 ms (SD = 5.8 ms). This leads to a motion to photon latency between the physical controller and the Oculus Rift screen of 34.4 ms (SD = 27.4ms). The application performed at a mean of 78.39 frames per second, i.e., a mean frame time of 12.75 ms (SD = 3.96 ms). Accordingly, the performance of the VR presentation is fast enough so that the technical implementation of the VR mode does not negatively influence the user experience (UX).

5.2. User study

For the evaluation of our approach, we focus the analysis of the UX in a virtual world using the VR-based solution in contrast to the traditional desktop app. As we expect that most users did not have experience in VR applications before, we assume that those might expect a higher effort in using the VR app due to initial familiarization. Accordingly, we want to discuss the effort expected by users for a new interaction type. Especially, we want to guarantee that the analysis of EFs does not suffer from less experience in using the VR technology. However, we still expect that the integration of VR improves the gamification and, hence, the UX. Accordingly, we investigate the following two hypotheses:

- *H1: VR has increased expected effort for users and, hence, affect negatively the analysis of EFs.*
- *H2: VR enhances the Hedonic Motivation and, hence, affect positively the analysis of EFs.*

5.2.1. Procedure

Each participant of the user study had to play the three game modes in the following specified order: (i) *Track Players* game, (ii) *Count Players* game, and (iii) *Find Ball* game. For each game mode, the participants were asked to play it four times to eliminate learning effects using both types of control, VR and non-VR control. The order of the two control types was chosen randomly with a toss coin for each participant at the beginning, but was equal for all three game modes for the user. Throughout the evaluation, the participants were supported by our staff for questions. Further, we provided a short introduction. Afterwards, participants answered a questionnaire.

5.2.2. Setup

We used the Oculus Rift VR device for the graphics output. The participants could interact with the application using the included Touch controllers and sensors. The configuration of the VR device and the controllers were done beforehand by our staff. As base station, we used a desktop PC with Intel Core i7 8,700 k 6 × 3.7 GHz, 32 GB RAM, and Nvidia Geforce GTX 1080Ti running with Windows 10 and Unity3D 2018.2.13. The right part of Figure 1 illustrates the environment with the Oculus Rift. Except of the control and the fact, that the content is shown on a regular display, the non-VR variant (see Figure 3) is identical to the VR one.

5.2.3. Methodology

We focus on the two dimensions of Effort Expectancy (cf. hypothesis *H1*) and Hedonic Motivation (cf. hypothesis *H2*). Effort Expectancy describes "the degree of ease for using the system the user expects" (Venkatesh et al., 2003). We expect that users might be unfamiliar with VR applications, hence, have a higher effort for using it. We operationalized the construct Effort Expectancy for each game mode using the question items provided by Venkatesh et al. (2003) regarding the ease of use and learnability. However, we expect that VR apps will provide a better user experience. Hedonic motivation describes "the fun or pleasure derived from using a technology" (Venkatesh et al., 2012). It showed to play an important role in determining technology acceptance and use, especially in gamification settings as the *CortexVR* application. We conceptualized Hedonic Motivation using question items from Venkatesh et al. (2012) and one additional item regarding the game atmosphere. Accordingly, we evaluate the trade-off of user effort for using (unfamiliar) VR and the benefits of a better analysis of EFs through a more realistic UX. Table 1 shows the question items for the user study. The question items are operationalized with 5-points Likert scales from 1 (strongly disagree) to 5 (strongly agree) and the option to skip question items. Except of the question item HC_4, we asked the same questions for both versions of the *CortexVR*–the one with VR control and the non-VR version with mouse/key control—and for each game mode.

We decided to follow this confounding-based approach as the *CortexVR* introduces new game modes that differ from the original Helix application which would result in an unequal comparison. Further, the integration of original player paths by the *CortexVR* application creates a different training setting. Additionally, it is also relevant for soccer clubs to know whether VR offers benefits for the training compared to a desktop application before buying VR hardware. Hence, we do not perform a comparison with the original Helix application.

5.2.4. Participants

Thirty-seven participants took part in the study. The age ranged between 21 and 35 (average = 24.89 years; standard deviation = 2.99 years). We had 25 males and 12 female users of which 13 used VR applications before, whereas 24 were first-time users. Sixteen users play regularly video games, 21 not.

5.2.5. Results

The box-plot in Figure 4 visualizes the results of the question items of the user study. The results are grouped by question items. As shown in the box-plot, the results are close to each

TABLE 1 Constructs of the user study (GameModeX stands for the (i) Track, (ii) Count, or (iii) Find Ball game modes).

ID	Question items for the user study
	Effort expectancy (Venkatesh et al., 2003)
EE_1	Learning how to use GameModeX with VR/mouse is easy for me.
EE_2	My interaction with the GameModeX with VR/mouse is clear and understandable.
EE_3	I find the GameModeX with VR/mouse easy to use.
EE_4	It is easy for me to become skillful at using the GameModeX with VR/mouse.
	Hedonic motivation (Venkatesh et al., 2012)
HC_1	Playing the GameModeX with VR/mouse is fun.
HC_2	Playing the GameModeX with VR/mouse is enjoyable.
HC_3	Playing the GameModeX with VR/mouse is very entertaining.
HC_4	The application creates a better stadium atmosphere using the Oculus Rift VR device.

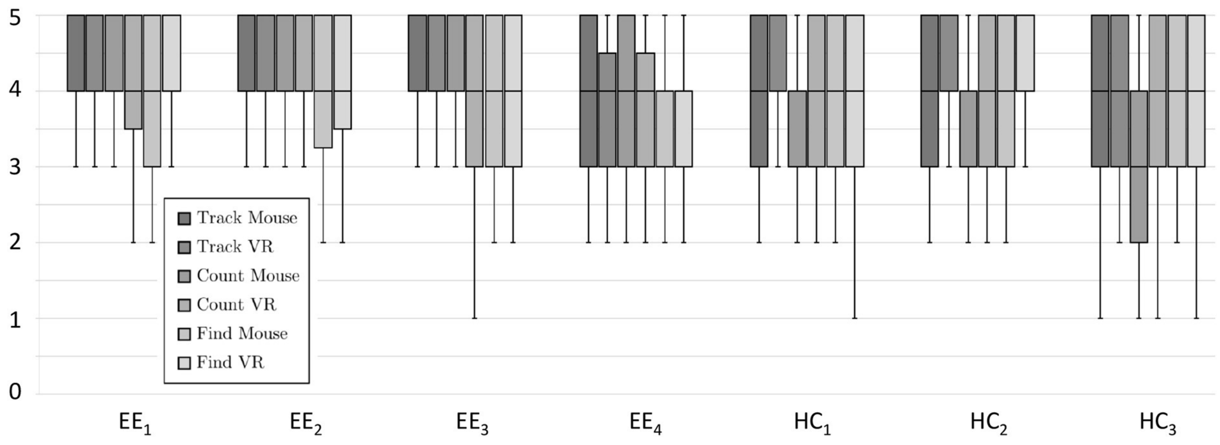


FIGURE 4
Box-plots of the effort expectancy and the hedonic motivation results (5-points Likert scales). Additionally, the "x" shows the mean value.

other for non-VR and VR, especially for Effort Expectancy. Additionally, we did a *t*-test for Effort Expectancy and Hedonic Motivation for analyzing significant effects between the use of the VR and non-VR applications and calculated the effect size using Cohen's d. The results indicate that differences in Effort Efficiency between the non-VR (M = 4.12, SD = 0.24) and the VR version [M = 4.00, SD = 0.19, $t_{(37)} = 1.31$, $p = 0.2$, $d = 0.5356$] are not significant. For Hedonic Motivation, we found that compared to the non-VR version (M = 3.69, SD = 0.20) the user experience is significantly improved with the introduction of VR [M = 3.94, SD = 0.21, $t_{(37)} = -2.44$, $p = 0.03$, $d = -1.1508$].

5.3. Adaptive game mode analysis

We applied a quantitative analysis for evaluating the performance of the reinforcement learning algorithm. Using the original implementation of the adaptive tracking game mode bears the risks of having a too steep increase of the difficulty (in case the learner does perform non-optimally) and corrupting the performance of users. Hence, we decided to use data of another study for simulating the use of the tracking game mode. This is possible, as the speed is adjusted only after played rounds, hence, we can use the data to mocking playing a round.

5.3.1. Methodology

We have collected data from 110 different participants, each completed one training session. A training session's data set contains data for each possible level between 13 and $40\frac{km}{h}$. The speed increased by $3\frac{km}{h}$ after three rounds. This results in 30 played rounds per participant. Using this data, we simulated training sessions with the adaptive game mode. We executed one hundred cycles for evaluating the learner. A cycle consisted of one training session with the data from each participant. We feed the learner with the results of a participant for the calculated speed. Then, the learner calculated for each participant the speed of the next level and, again, we used the data for the speed of the participant. We performed thirty rounds (10 levels with three rounds each). Since the reinforcement learning algorithm is updated after each level, we perform around $100,000$ updates.

5.3.2. Results

We analyzed the cumulative reward and the regret of a selected action. The results are described below.

5.3.2.1. Reward

First, we consider the reward that the reinforcement learning algorithm is supposed to apply a beneficial policy. It can be negative for actions not useful for achieving the algorithm's goal and positive for expedient ones. The reward is calculated at each update for the executed action. Figure 5 shows the cumulative reward for all updates of the simulation. A mostly linear increase can be recognized. This linear-gradient indicates that the algorithm found a policy not changing significantly. Further, it indicates the policy granting a positive reward on average. Still, the decrease at the beginning of the process needs around 3,300 updates to achieve a break-even point. From this point on, the previously mentioned linear-gradient commences.

The cross-validation of the reward into ten user groups show that there are clear differences depending on the composition of a group as visible by the gradient of the different functions (see Figure 6). Accordingly, an approach for learning in this use case should cluster the data beforehand into user groups.

5.3.2.2. Regret

The regret can be defined as the difference in the reward achieved for a chosen action in comparison to the reward

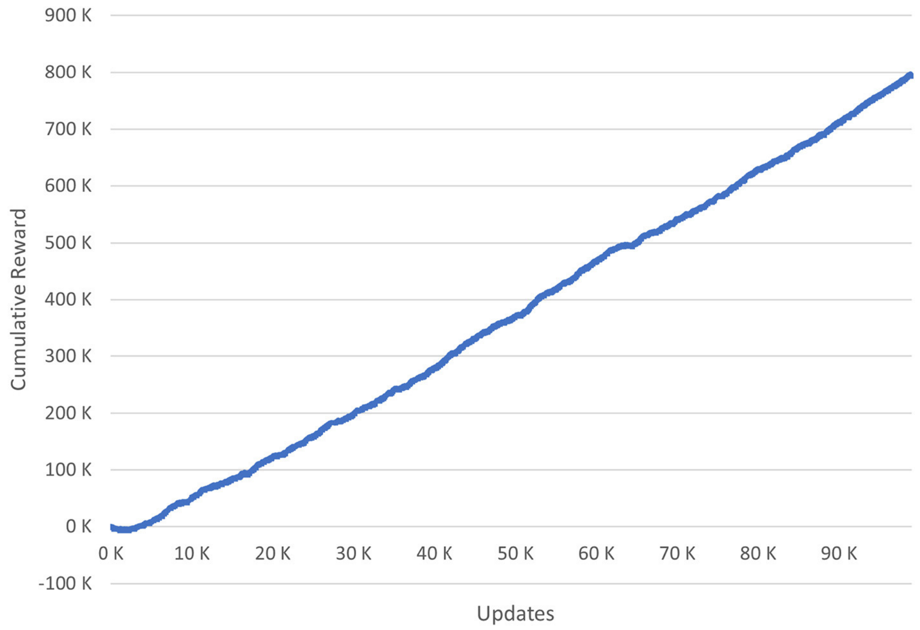


FIGURE 5
The cumulative reward of the simulation.

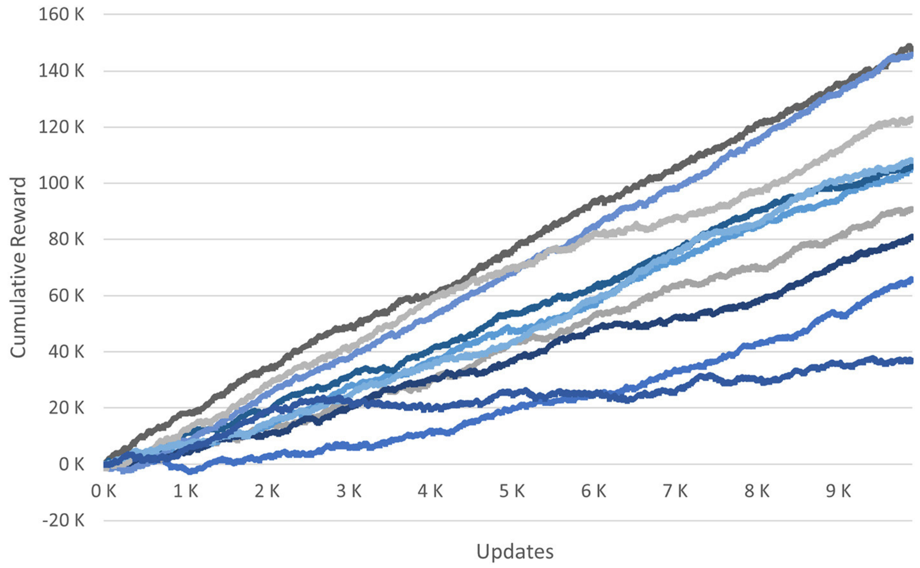


FIGURE 6
Ten-fold cross-validation for the cumulative reward.

of the best possible action. We define a normalized regret as $normalizedregret = regret(executedaction)/regret(worstaction)$ whereas the worst action the one with the highest regret is. As a result, the normalized regret stays between 0 (best possible action) and 1 (worst possible action). Figure 7 shows the normalized regret for each update (average for 1,000 steps each) and the development of the score. As one can see, the regret on average decreases with increasing rounds, i.e., with more training. Fluctuations in the regret shows the regularly exploration of new actions.

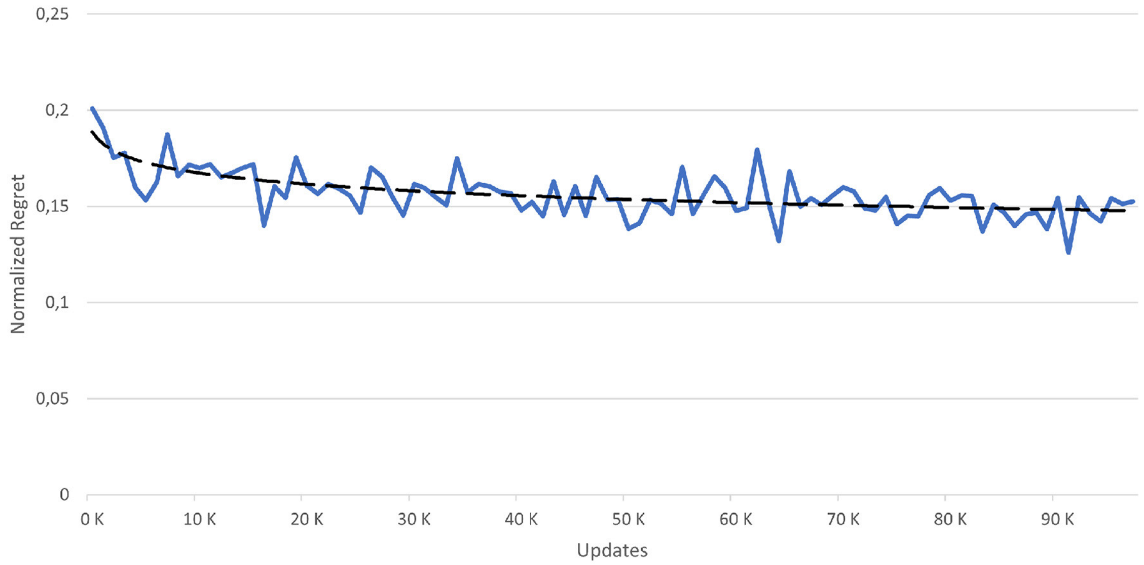


FIGURE 7
Regret for the reinforcement learning approach.

6. Discussion

In this section, we discuss the suitability of the *CortexVR* and *Coach App* applications based on the evaluation. Therefore, this section is divided according to the hypotheses.

In contrast to the former Helix, users of *CortexVR* can directly interact with the application (see requirement $R_{App}2$), either using mouse/key control or the Oculus Rift VR device (see requirement $R_{App}1$). As this avoids the need of having a room-filling projection system, the *CortexVR* application is a stand-alone tool (see requirement $R_{Config}4$) with minimal setup effort. The results show that either mouse/key control or VR is superior in terms of Effort Expectancy (rejecting hypothesis *H1*) depending on the game mode. However, a t-test did not show significance. We could find only weak correlations between the moderators and these results. Some minor correlations exist between the variable of having experience in using VR devices with a satisfaction of the VR control. It is feasible that users with higher experience in using VR applications will show significant better results. Further, training effects might change the results, too. This could be investigated in an extended long-term user study. The metrics shows the superiority of VR regarding Hedonic Motivation (supporting hypothesis *H2*). Especially all participants (fully) agreed that VR creates a better atmosphere. This improves the gamification effects which are a relevant aspect for the effectiveness of the training sessions.

The user study indicates that VR enhances the Hedonic Motivation (cf. hypothesis *H2*). As this might by a result of the novelty of using VR for some participants, we need to investigate the results in more detail within a longitudinal study. We could further investigate whether familiarity with the system improves the handling times with VR-based control. This is an important factor for measuring a user's performance, i.e., supporting the analysis and comparison of players' EFs.

In contrast to the original Helix system, our implementation of the *CortexVR* system supports the analysis of the reaction times of a player in addition to the accuracy. However, the current evaluation focused on comparing the VR-based and the non-VR versions of the application and purposely did not measure efficiency of users in terms of reaction times. These times are highly dependent on the interaction style and input interface, here mouse and 3D controllers, and, especially, those depend on the experience of users with the user interface. However, this would be an analysis of the user interface rather than the efficiency of the system for supporting the users in their task. Still, the new functionality of measuring reaction times enable to use those measures in studies to efficiently compare user as well as observe the training progress of players.

Our system not only supports the analysis of EFs, but can be also used for training of EFs. The most effective method to prove the efficiency of our system in training EFs would be a long-term study of a user group. However, this is not the scope of this paper. Still, a long-term user study for analyzing the effects of training the EFs using our *CortexVR* system is part of our future work, however, this is barely feasible due to the fluctuating nature (due to player transfers or injuries) in the composition of soccer teams.

However, there is a lack of theoretical and empirical justification of the included training tasks. We designed the games modes after discussions with staff and coaches of the soccer clubTSG HoffenheimẆhereas the staff have a lot of experience in this area and also scientific personnel was involved, we did not follow a systematic approach. Open questions include: How can be assured that the games selected improve the proposed EFs? How to set the parameters (number of players to be tracked or counted, number of distractors and degree of similarity with the targets, etc.) of each game in relation to the player's EF level? How to systematically define the complexity of the situations in which the players have to estimate the ball position depending on their EFs?

7. Conclusion

This paper presented our approach for training and analyzing EFs, a set of cognitive functions. Different studies have shown that high class soccer players have a dominant level of EFs compared to low class players or non players (Green et al., 2012; Vestberg et al., 2012; Verburgh et al., 2014). In cooperation with the German Bundesligasoccer club TSG Hoffenheim, we designed and implemented the *CortexVR* for VR-based training of the EFs. The *CortexVR* application is complemented by the *Coach App* for configuration of training sessions and analysis of the user performance data.

We evaluated both systems. For the *CortexVR* we designed a user study based on the constructs Effort Expectancy (Venkatesh et al., 2003) and Hedonic Motivation (Venkatesh et al., 2012). Thirty-seven users played the three game modes of the *CortexVR* app with a traditional mouse/key control and using the Oculus Rift VR device with the Oculus controllers. The results support our hypotheses. Regarding the Hedonic Motivation (hypothesis *H2*), the VR-based control is significantly dominating and supports the usage experience. Additionally, the user study showed that depending on the game mode, either the VR-based or the mouse/key-based control is superior in terms of Effort Expectancy (hypothesis *H1*). However, detailed tests in a longitudinal study on the effects for training by using VR are subject to future work.

In this paper, we focused the domain of professional soccer players. However, our approach can be applied for training of EFs in other domains, e.g., for the treatment of Parkinson's disease (Sammer et al., 2006) or children with ADHD (Tamm et al., 2014). Additionally, it might be used for education and training in domains where a fast reaction (in emergency and exceptional situations) is critical, e.g., police, fire brigade, or public transportation. Through the modularity of our system, it can be easily customized and extended, e.g., with new game modes or another game engine. Both measurements of the adaptive game mode, reward, and regret, indicate the correctness of the learner. Still, the benefits of adaptation of the game level have to be analyzed in detail. This was out of scope of this paper but is part of our future work.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

CKr and JN were initially writing the paper, supporting the user study, and leading the project. JS and ML were responsible for revising the paper and provided advice in the conceptualization of the study and the VR parts. JM, JS, and PE contributed to the conceptualization of the prototype and supported the development with scientific advise regarding the Executive Functions. NB, MB, AH, and CKo were responsible for the development of the CortexVR software. FH was responsible for the development of the adaptive game mode. SK and CB acted as scientific advisor supporting the conceptualization and advising the paper writing process. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Akbaş, A., Marszałek, W., Kamieniarz, A., Polechooński, Słomka, K. J., and Juras, G. (2019). Application of virtual reality in competitive athletes—a review. *J. Hum. Kinet.* 69, 5–16. doi: 10.2478/hukin-2019-0023

Baddeley, A. (2010). Working memory. *Curr. Biol.* 20, R136–R140. doi: 10.1016/j.cub.2009.12.014

Bejjanki, V. R., Zhang, R., Li, R., Pouget, A., Green, C. S., Lu, Z.-L., et al. (2014). Action video game play facilitates the development of better perceptual templates. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16961–16966. doi: 10.1073/pnas.1417056111

Bideau, B., Kulpa, R., Vignais, N., Brault, S., Multon, F., and Craig, C. (2010). Using virtual reality to analyze sports performance. *IEEE Comput. Graph. Appl.* 30, 14–21. doi: 10.1109/MCG.2009.134

Bideau, B., Multon, F., Kulpa, R., Fradet, L., and Arnaldi, B. (2004). "Virtual reality applied to sports: Do handball goalkeepers react realistically to simulated synthetic opponents?," in *Proceedings of the 2004 ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry*, 210–216. doi: 10.1145/1044588.1044632

Bird, J. M. (2020). The use of virtual reality head-mounted displays within applied sport psychology. *J. Sport Psychol. Action* 11, 115–128. doi: 10.1080/21520704.2018.1563573

Bowman, D. A., and McMahan, R. P. (2007). Virtual reality: how much immersion is enough? *Computer* 40, 36–43. doi: 10.1109/MC.2007.257

Büning, C., Jürgens, L., and Lausberg, H. (2021). Divergent learning experiences in sports enhance cognitive executive functions and creativity in students. *Phys. Educ. Sport Pedag.* 26, 402–416. doi: 10.1080/17408989.2020.1812056

Carling, C. (2010). Analysis of physical activity profiles when running with the ball in a professional soccer team. *J. Sports Sci.* 28, 319–326. doi: 10.1080/02640410903473851

Climent-Martinez, G., Luna-Lario, P., Bombin-Gonzalez, I., Cifuentes-Rodriguez, A., Tirapu-Ustarroz, J., and Diaz-Orueta, U. (2014). Neuropsychological evaluation of the executive functions by means of virtual reality. *Rev. Neurol.* 58, 465–475. doi: 10.33588/rn.5810.2013487

De Waelle, S., Laureys, F., Lenoir, M., Bennett, S. J., and Deconinck, F. J. (2021). Children involved in team sports show superior executive function compared to their peers involved in self-paced sports. *Children* 8:264. doi: 10.3390/children8040264

Dye, M. W., Green, C. S., and Bavelier, D. (2009). Increasing speed of processing with action video games. *Curr. Direct. Psychol. Sci.* 18, 321–326. doi: 10.1111/j.1467-8721.2009.01660.x

EonSports VR (2019). *EonSports VR Training*. Available online at: http://eonsportsvr.com/vr-sports-training (accessed March 25, 2019).

Gao, Z., Lee, J. E., McDonough, D. J., and Albers, C. (2020). Virtual reality exercise as a coping strategy for health and wellness promotion in older adults during the COVID-19 pandemic. *J. Clin. Med.* 9:1986. doi: 10.3390/jcm9061986

Green, C. S., Sugarman, M. A., Medford, K., Klobusicky, E., and Bavelier, D. (2012). The effect of action video game experience on task-switching. *Comput. Hum. Behav.* 28, 984–994. doi: 10.1016/j.chb.2011.12.020

Harnishfeger, K. K. (1995). "The development of cognitive inhibition: theories, definitions, and research evidence," in *Interference and Inhibition in Cognition* (Elsevier), 175–204. doi: 10.1016/B978-012208930-5/50007-6

Holfelder, B., Klotzbier, T. J., Eisele, M., and Schott, N. (2020). Hot and cool executive function in elite- and amateur- adolescent athletes from open and closed skills sports. *Front. Psychol.* 11:694. doi: 10.3389/fpsyg.2020.00694

Icaros (2019). *Active VR Icaros*. Available online at: https://www.icaros.com/ (accessed March 3, 2019).

Ingle, S. (2016). *Are We a Step Closer to Being Able to Measure Football IQ*? Available online at: https://www.theguardian.com/football/blog/2016/dec/04/barcelona-andres-iniesta-scope-embrace-brain-game-real-madrid (accessed March 25, 2019).

Koch, P., and Krenn, B. (2021). Executive functions in elite athletes - comparing open-skill and closed-skill sports and considering the role of athletes' past involvement in both sport categories. *Psychol. Sport Exerc.* 55:101925. doi: 10.1016/j.psychsport.2021.101925

Kubesch, S., and Walk, L. (2009). Körperliches und kognitives training exekutiver funktionen in kindergarten und schule. *Sportwissenschaft* 39, 309–317. doi: 10.1007/s12662-009-0079-2

Lalonde, G., Henry, M., Drouin-Germain, A., Nolin, P., and Beauchamp, M. H. (2013). Assessment of executive function in adolescence: a comparison of traditional and virtual reality tools. *J. Neurosci. Methods* 219, 76–82. doi: 10.1016/j.jneumeth.2013.07.005

Lo Priore, C., Castelnuovo, G., Liccione, D., and Liccione, D. (2003). Experience with v-store: considerations on presence in virtual environments for effective neuropsychological rehabilitation of executive functions. *Cyberpsychol. Behav.* 6, 281–287. doi: 10.1089/109493103322011579

McGeorge, P., Phillips, L. H., Crawford, J. R., Garden, S. E., Sala, S. D., Milne, A. B., et al. (2001). Using virtual environments in the assessment of executive dysfunction. *Presence* 10, 375–383. doi: 10.1162/1054746011470235

McPartland, M., and Gallagher, M. (2011). Reinforcement learning in first person shooter games. *IEEE Trans. Comput. Intell. AI Games* 3, 43–56. doi: 10.1109/TCIAIG.2010.2100395

Mi Hiepa Sports (2019). *The Global Elite Football VR Platform*. Available online at: http://mihiepa.com/ (accessed March 25, 2019).

Michael Horeni (2017). *Innovation—Da Sind Wir Spitze*. Available online at: https://www.faz.net/aktuell/sport/fussball/wohin-rollt-der-ball/1899-hoffenheim-maezen-dietmar-hopp-im-interview-ueber-innovation-15128374/helix-auf-der-15131800.html

NeuroTrainer (2019). *Connect Neuroscience to In-Game Performance*. Available online at: https://neurotrainer.com/ (accessed March 3, 2019).

Pugnetti, L., Mendozzi, L., Attree, E. A., Barbieri, E., Brooks, B. M., Cazzullo, C. L., et al. (1998). Probing memory and executive functions with virtual reality: past and present studies. *CyberPsychol. Behav.* 1, 151–161. doi: 10.1089/cpb.1998.1.151

Reaction VR Sports (2019). *Virtual Goalie—Virtual Reality Lacrosse Goalie Training*. Available online at: https://www.reactionvrsports.com/virtual-goalie-product/ (accessed March 3, 2019).

Sammer, G., Reuter, I., Hullmann, K., Kaps, M., and Vaitl, D. (2006). Training of executive functions in Parkinson's disease. *J. Neurol. Sci.* 248, 115–119. doi: 10.1016/j.jns.2006.05.028

Scott, W. A. (1962). Cognitive complexity and cognitive flexibility. *Sociometry* 25, 405–414. doi: 10.2307/2785779

Strauss, E., Sherman, E. M., and Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. American Chemical Society.

StriVR (2019). *StriVR Sports*. Available online at: https://www.strivr.com/sports/ (accessed March 3, 2019).

Sturtevant, N. R., and White, A. M. (2007). "Feature construction for reinforcement learning in hearts," in *Computers and Games, Vol. 4630 of Lecture Notes in Computer Science*, eds D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, H. J. van den Herik, P. Ciancarini, and H. H. L. M. Donkers (Berlin; Heidelberg: Springer Berlin Heidelberg), 122–134. doi: 10.1007/978-3-540-75538-8_11

Success Series (2019). *Success Series*. Available online at: http://www.successseries.com/ (accessed March 3, 2019).

Tamm, L., Nakonezny, P. A., and Hughes, C. W. (2014). An open trial of a metacognitive executive function training for young children with ADHD. *J. Attent. Disord.* 18, 551–559. doi: 10.1177/1087054712445782

Tesauro, G. (1995). Temporal difference learning and td-gammon. *Commun. ACM* 38, 58–68. doi: 10.1145/203330.203343

Venkatesh, V., Morris, M., Davis, G., and Davis, F. (2003). User acceptance of information technology: toward a unified view. *MIS Q.* 27, 425–478. doi: 10.2307/30036540

Venkatesh, V., Thong, J. Y. L., and Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS Q.* 36, 157–178. doi: 10.2307/41410412

Verburgh, L., Scherder, E. J. A., van Lange, P. A., and Oosterlaan, J. (2014). Executive functioning in highly talented soccer players. *PLoS ONE* 9:e91254. doi: 10.1371/journal.pone.0091254

Vestberg, T., Gustafson, R., Maurex, L., Ingvar, M., and Petrovic, P. (2012). Executive functions predict the success of top-soccer players. *PLoS ONE* 7:e34731. doi: 10.1371/journal.pone.0034731

Weiss, P. L., Kizony, R., Feintuch, U., and Katz, N. (2006). Virtual reality in neurorehabilitation. *Textb Neural Repair Rehabil.* 51, 182–197. doi: 10.1017/CBO9780511545078.015

frontiers | Frontiers in Neuroscience

TYPE Review
PUBLISHED 08 September 2022
DOI 10.3389/fnins.2022.811736

OPEN ACCESS

EDITED BY
Hong Xu,
Nanyang Technological University,
Singapore

REVIEWED BY
Paul Ferrari,
Helen DeVos Children's Hospital,
United States
Mohammad Reza Daliri,
Iran University of Science
and Technology, Iran

*CORRESPONDENCE
Célia Loriette
celia.loriette@isc.cnrs.fr
Suliann Ben Hamed
benhamed@isc.cnrs.fr

SPECIALTY SECTION
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

RECEIVED 09 November 2021
ACCEPTED 23 August 2022
PUBLISHED 08 September 2022

CITATION
Loriette C, Amengual JL and Ben Hamed S (2022) Beyond the brain-computer interface: Decoding brain activity as a tool to understand neuronal mechanisms subtending cognition and behavior.
Front. Neurosci. 16:811736.
doi: 10.3389/fnins.2022.811736

Beyond the brain-computer interface: Decoding brain activity as a tool to understand neuronal mechanisms subtending cognition and behavior

Célia Loriette*, Julian L. Amengual and Suliann Ben Hamed*

Institut des Sciences Cognitives Marc Jeannerod, CNRS UMR 5229, Université Claude Bernard Lyon 1, Bron, France

One of the major challenges in system neurosciences consists in developing techniques for estimating the cognitive information content in brain activity. This has an enormous potential in different domains spanning from clinical applications, cognitive enhancement to a better understanding of the neural bases of cognition. In this context, the inclusion of machine learning techniques to decode different aspects of human cognition and behavior and its use to develop brain–computer interfaces for applications in neuroprosthetics has supported a genuine revolution in the field. However, while these approaches have been shown quite successful for the study of the motor and sensory functions, success is still far from being reached when it comes to covert cognitive functions such as attention, motivation and decision making. While improvement in this field of BCIs is growing fast, a new research focus has emerged from the development of strategies for decoding neural activity. In this review, we aim at exploring how the advanced in decoding of brain activity is becoming a major neuroscience tool moving forward our understanding of brain functions, providing a robust theoretical framework to test predictions on the relationship between brain activity and cognition and behavior.

KEYWORDS

brain decoding, brain–computer interfaces, machine learning, electrophysiology, fMRI, neurofeedback, cognition, attention

Introduction

One of the major challenges of system neurosciences is to understand how brain functions subtends cognition and behavior. This knowledge is essential not only for a better description of how the brain works, but also to develop strategies to boost cognition and to ameliorate, or even restore, cognitive functions affected by neurological diseases. During the last decades, a humongous technical, theoretical, and clinical effort has been invested in exploring and analyzing the activity of the brain. One important advance in this context that is gaining a huge momentum in the field is the inclusion of machine learning to the analysis of brain activity (Abraham et al., 2014; Glaser et al., 2020; Iturrate et al., 2020). Machine learning is a branch of artificial intelligence which consists in the development of algorithms that imitate the way humans learn from data. The main motivation of the application of machine learning in neuroscience is the development of brain–computer interface technologies (BCI), which comes with the idea that computers might be able to mimic some of the brain's most basic cognitive capacities (Johnson, 2000). In other words, computers might reproduce the type of computations (or cognitive operations) performed by the brain. This does not imply biomimetism. In particular, it does not imply that machine learning imitates how humans learn. For instance, these techniques might be able to learn predicting a specific output (specific movement or a physical attribute of an object) from given states of a system (the activity of a region of the brain) (Glaser et al., 2020) or identifying which activity patterns are more alike to a certain behavioral counterpart (Richards et al., 2019). To achieve these goals, machine learning refers to a large list of methods spanning from supervised linear regression algorithms to other more sophisticated and complex learning tools such as deep learning neural networks.

A seminal application of machine learning in the field of neuroscience consists in the classification of brain activity patterns to predict observable outputs such as motor behavior (e.g., arm or leg movements) and visual inputs (physical attributes of stimulus) (Zafar et al., 2015; Branco et al., 2017; Contini et al., 2017; Tam et al., 2019; Kashefi and Daliri, 2021; Nazari et al., 2021) achieved in real-time, which has allowed the development of brain–computer interface technologies (BCI). Nowadays, such technologies have allowed, for instance, the real-time reconstruction of an image seen by a subject from the analysis of the concurrent visual responses recorded in the subject's occipital cortex (Shen et al., 2019; Huang et al., 2020) or to control a robotized arm by using exclusively the neural activity recorded over the motor cortex of the subject. This has led to the development of the research field of neuroprosthetics, with a myriad of potential applications to restore lost motor and sensory brain functions (Hochberg et al., 2006, 2012; Adewole et al., 2016; Bouton et al., 2016; Lebedev and Nicolelis, 2017). However, and while most of the research in neuroprosthetics fall into the use of signals associated to observable outputs such as those commented above, a whole range of cognitive functions remain unexploited in this context, such as attention, memory, visual imagery and motivation, and predicting these covert cognitive functions from brain activity remains challenging (Astrand et al., 2014b).

In addition to the development of BCI interfaces to provide tools for neuroprosthetics, a new perspective on machine learning in neuroscience is emerging, which consist on its use to genuinely model different aspects of brain processes and, thus, increase our understanding of brain functions (Hebart and Baker, 2018). In this context, relevant advances have been achieved in understanding covert spatial attention mechanisms, which corresponds to the ability of a subject to select relevant sensory information while ignoring other inputs or stimulations, independently of eye position. Visual covert attention is known to rely on a well described brain network involving the prefrontal cortex and specifically the frontal eye fields (FEF), the intra-parietal sulcus (IPS) and striate and extra-striate visual areas (Shulman et al., 1999; Corbetta and Shulman, 2002; Bisley, 2011; Bogadhi et al., 2018). Behavioral evidence has suggested that attention rhythmically samples space (Buschman and Miller, 2007; VanRullen et al., 2007; Fiebelkorn et al., 2013, 2018; Dugué et al., 2015, 2016; VanRullen, 2016, 2018; Gaillard and Ben Hamed, 2020; Gaillard et al., 2020). The use of machine learning tools to predict the position of the attentional spotlight have revealed that the position of the spatial attention oscillates, at the same frequency as the behavioral performance rhythmic fluctuations. Therefore, machine learning tools have been useful to describe the neural bases accounting for the behavioral attentional sampling fluctuations (Gaillard et al., 2020).

The aim of this review is to explore how advanced machine learning methods, beyond their applicability in neuroprosthetics, can be used as a powerful tool to better characterize brain functions, with a specific focus on covert functions. First, we will introduce the concept of *decoding* of brain activity and its application into multiple cognitive brain functions, and we will discuss the different methods of brain decoding. Second, we will provide an overview of the specific methodology that machine learning offers to neuroscientists to describe the relationship between brain activity and cognitive brain functions. Last, we will show how we can use this knowledge to develop accurate cognitive brain–computer interface tools based on neurofeedback and learning.

Decoding brain information

Input cortical signals

The aim of machine learning in neuroscience is to extract reliable information associated to a specific cognitive function subtended by brain activity. These computational methods are based on neural decoding, which consists in the ability of an algorithm to predict or reconstruct the information that has been encoded and represented in the activity of a specific brain region or network. In this section we will discuss the different methods for extracting brain activity and the different advantages and disadvantages related to decoding of information.

Broadly speaking, brain activity can be recorded either using invasive or non-invasive methods (**Figure 1**). Invasive

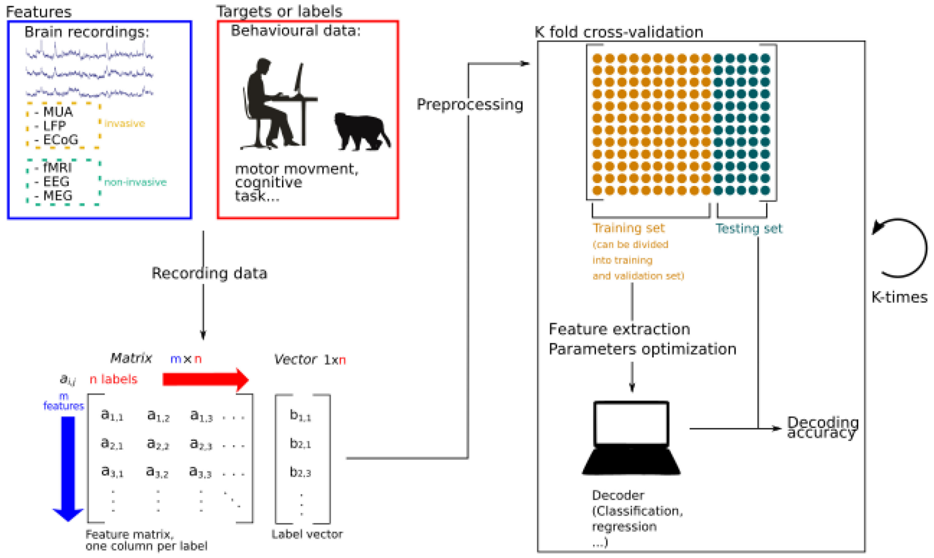


FIGURE 1
Schematic representation of machine learning procedures applied to brain decoding. Brain activity is recorded while participants are performing a specific behavioral task. The recorded data is stored as a matrix and data is pre-processed to extract relevant information and increase the signal to noise ratio. Extracted information is called features. Then the data is split into two sets, a training, and a testing set. The training set can further be divided into a training set and a validation set, used to perform feature extraction and parameters optimizations. If the algorithm is supervised, features and labels are fed to the decoder. In unsupervised learning, only the features are used to train the algorithm. The performance of the decoder is then estimated by testing its accuracy thanks to the remaining testing set. Splitting the data into training and testing sets can be done by splitting the data into *N* sets (*N* is a number to be defined) allowing to generalize the performance of the algorithm by calculating an average decoding accuracy.

electrophysiological recordings allow to record single-unit (SUA) or multi-unit activity (MUA), reflecting the activity of a few neurons, and corresponding to discrete action potentials, which can be seen as bits of neuronal information. They also allow to record the activity of a larger neuronal population, including their synaptic inputs and outputs, called local field potentials (LFP). To note, the size of the corresponding neuronal population reflected in the LFP depends on the physical impedance of the implanted electrodes (the lower the impedance, the larger the neuronal population that the LFP represents). These signals are characterized both by a very high temporal resolution and signal-to-noise ratio allowing a single trial level decoding. Indeed, this signal is recorded using electrodes that are purposefully implanted in cortical regions that are specifically involved in processing the function of interest. However, these methods also have a quite low spatial resolution as they sample only a few cortical sites, even when dense multi-unit recordings are performed, and therefore they are restricted to specific brain sources. Mostly used in non-human primate experiments, this type of invasive approach has led to massive advances in the field of motor neuroprosthetics. Previous studies have shown in non-human primates the possibility to drive robotic arms (Velliste et al., 2008) or virtual effectors (Golub et al., 2014) using a direct control over the activity of their motor cortex. These approaches have also been applied to covert cognitive functions, that is to say, to the decoding of the content of the cognitive processes rather than the observable associated behavior (Astrand et al., 2014a). In this context, important advances have been obtained in tracking spatial attention (irrespective of eye position) with both a high spatial (in the order of 0.5° visual degree) (Astrand et al., 2016; De Sousa et al., 2021) and a high temporal resolution (in the order of 50 ms) (Gaillard et al., 2020). Invasive experimental set ups have also been developed in human studies, whether using electrocorticography (ECoG), or intracranial electroencephalography (iEEG), but strictly restricted to specific clinical demands. Some studies have included tetraplegic patients applying decoding methods to their cortical activity in order to restore their motor functions (Hochberg et al., 2012; Bouton et al., 2016; Ajiboye et al., 2017). However, and in spite of the refined and precise information that they provide, the application of invasive methods of decoding brain activity in humans remains rare as they involve an invasive surgery and they come with strong ethical limitations (Nicolas-Alonso and Gomez-Gil, 2012).

Fortunately, brain activity can also be recorded using non-invasive techniques such as scalp electroencephalography (EEG) which provides recording activity at a very high temporal but low spatial resolution, and a very low signal-to-noise ratio; magnetoencephalography (MEG), which has a very good temporal and cortical spatial resolution, in spite of a variable signal-to-noise ratio; and functional magnetic resonance imaging (fMRI), which allows to record whole-brain activity with a high spatial resolution, but is limited temporally by the time of acquisition of the different brain slices and the huge delay (in the order of few seconds) of the hemodynamic response (BOLD signal) relative to the actual neuronal response, as well as by a relatively low signal-to-noise ratio (Nicolas-Alonso and Gomez-Gil, 2012). Another example of non-invasive methods of brain recording used for decoding information is the use of functional-near infrared spectroscopy (fNIRS), which is characterized by a higher spatial resolution than the non-invasive electrophysiological recordings (EEG/MEG) and much better temporal resolution than the fMRI, as well as by a higher portability than the abovementioned non-invasive methods (Wilcox and Biondi, 2015).

Despite these limitations, non-invasive methods of brain activity recording have shown quite successful decoding capacities. For instance, the lateralization of the locus of spatial attention has been decoded in humans using event related brain potentials (ERPs) extracted from scalp EEG recordings in humans in response to visual presentations (Trachel et al., 2015; Thiery et al., 2016). In fMRI, covert visuospatial attention can be decoded with a high level of accuracy when four positions are encoded (80% of accuracy) (Andersson et al., 2012) and 40% accuracy when eight positions are encoded (Loriette et al., 2021). Machine learning has also been applied to train predictive speech models using fNIRS, obtaining a 75% of accuracy in classifying long speech segments from brain activity (Liu and Ayaz, 2018).

Up-to-know, the current decoding of information capacities from non-invasive brain signal shows lower accuracies when compared with those obtained with invasive techniques. However, this limitation is overcome by the fact that they are readily accessible, easy to manipulate and "discomfort-free" by the user. Therefore, an intense effort is invested in the field to improve the decoding capacities using such methods to develop more successful brain–machine interface and other therapeutic applications.

Methodology of brain decoding

Machine learning methods rely on the development of algorithms to map recorded brain activity onto encoded information. These algorithms can be classified into *supervised* or *unsupervised* learning algorithms. Supervised algorithms learn the mapping between an input to an output by using input-output pairs from a set of *training* examples. From such learning process, an inferred function is provided that can be used to map new examples onto the different possible outputs (**Figure 1**). For instance, a function can be trained to identify whether a movement will be performed using the left or the right upper limb by providing multiple datasets of recorded brain activity from the left and right motor cortex concomitant with the movement of both arms. Once the algorithm has learned this association, it will be able to predict which arm is being moved by merely using novel brain recordings from both hemispheres. In contrast, *unsupervised* learning involves feeding a model with brain activity without giving any explicit information about the corresponding output, and therefore letting the algorithm estimate the number of possible outputs from a classification of the activity of the data based on its multidimensional statistical structure. New observations are then associated with these statistically defined classes. There exists a very large range of supervised and unsupervised classification algorithms which we will not discuss here as this would require an independent review in itself.

Some considerations must be considered when selecting which learning algorithm is more suitable to use. Unsupervised learning algorithms are quite sensitive to the size of the data (dimensionality and number of trials). It is well known that complex models such as deep neural network (networks with multiple layers) based learning algorithms require large data sets to obtain a good model estimation. In the absence of large enough data sets, these models risk overfitting, which consists in the over-learning of the training data structure, producing a low generalization of the prediction capacities when novel data sets are used. The risk of overfitting is precisely the reason why deep neural networks are often restricted to overt visual or motor data which can be collected in large amounts in a short time. Other lower complexity models such as support vector machine (SVM), linear discriminant analysis (LDA) or regression trees are applied to the decoding of covert cognitive functions, as these require more demanding and longer behavioral tasks to collect training data (Lemm et al., 2011; Abraham et al., 2014; Taghizadeh-Sarabi et al., 2015).

When considering supervised learning algorithms, specific good practices should be followed regarding training and testing procedures. The aim of these algorithms is to estimate the weights of a more or less complex function which minimizes the prediction error of the training set (i.e., a model which minimizes the error between the actual outputs of the training set and the outputs predicted by the model from the brain activity associated with the actual outputs of the training set). However, the prediction capabilities of the model (the decoding accuracy) are generally measured by using a different set of novel inputs (testing set) and evaluating the prediction error in these new examples. The standard methodology used to achieve a reliable decoding accuracy is the cross-validation, which consist in dividing or splitting the data set iteratively in two training and testing subsets. The decoding accuracy in these dataset will be defined as the average of the accuracies obtained in each iteration from different training and testing sets (Glaser et al., 2020) (**Figure 1**).

Last, most of the models are characterized by tuning parameters that can be changed to refine the model. As is the case for training, these parameters cannot be optimized on the whole data set as this will inflate the accuracy scores. Most often, the dataset is divided into three sets, one set for parameter optimization (called the validation set), one set for training data on the optimized model and one set for testing and estimating an unbiased decoding accuracy. As described from training/testing, cross-validation approaches can also be applied to model parameter optimization and to select the main features of the data prior to model training (Glaser et al., 2020).

Exploring cognitive brain function using decoding methods

Decoding accuracy as a window into cognitive brain processes

Decoding information related with covert cognitive functions such as attention, intention, and decisions are still in its early days. Indeed, decoding covert information often leads to lower decoding accuracies as compared to the decoding of sensory or motor functions, despite consistent effort invested to improve this decoding. This stems from multiple reasons. The first reason pertains to the fact that cognitive information (say spatial attention) is implemented in associative cortices, mixed with multiple other sources of cognitive information (e.g., working memory, temporal expectation, planning etc.). Such multiplexing of cognitive information has been theorized as a strategy to enhance the coding dimensionality of primate brains (Rigotti et al., 2013). In the absence of an appropriate pre-processing of the neuronal data, this results in a low signal-to-noise (SNR) when trying to extract a given dimension in isolation. In other words, signal corresponding to the specific cognitive process of interest might be too low compared to other sources of information contained in that same signal. Such effect has been well described in specific brain regions, such as the prefrontal cortex, in which different sources of information are simultaneously encoded, a property of neuronal populations known as mixed selectivity (Rigotti et al., 2013). With appropriate pre-processing, such as dimensionality reduction or de-mixed dimensionality reduction approaches (Kobak et al., 2016), SNR is enhanced, as it becomes possible to assign overall signal variance to the process of interest as well as to the cognitive processes of non-interest. Applying such dimensionality reduction approaches to MUA recordings form the FEF allows to better decode the spatial orientation of attention irrespective of whether the subjects are engaged in the task of not, thus dissociating between attention orientation and task engagement (Amengual and Ben Hamed, 2021; Amengual et al., 2022).

The second reason why decoding cognitive information does not reach high accuracies is due to the fact that, in contrast with motor or sensory responses which can be precisely timed relative to movement initiation or sensory stimulation onset, cognitive information is generated endogenously by the subjects, such that onset time can only be approximated from external events. For example, when producing an arm movement in response to a visual cue, subjects will produce a range of motor response from fast to slow. When instructed to orient their spatial attention following a visual cue, subjects will also do so more or less rapidly. However, there is no objective way of quantifying this precisely. Likewise, while an awkwardly organized arm movement can directly be observed, a sluggish attentional orientation can only be inferred from success in the ongoing task. This thus results in uncontrolled for sources of noise. A good example to illustrate this is the strategy that our research group has implemented to decode spatial attention orienting. Astrand et al. (2016) decode the ($x-$,$y-$) position of attention by training an algorithm on correct trials only, under the very strong assumption that on such trials, attention is precisely oriented at the cued location. The decoding output on the test trials makes it clear that this assumption is false not only on error trials (where one expects attention to be miss allocated) but also on correct trials, such that attention can be either close or far to the cued location, although on average, attention is positioned on the cued location (Astrand et al., 2016, 2020). These findings have been confirmed in humans using fMRI recordings (Loriette et al., 2021). Indeed, when predicting the spatial orienting of attention from BOLD activity in the striate and extrastriate cortex, while maximum decoding accuracy is achieved for the spatial location that the subjects are requested to attend to, on a significant proportion of the trials, attention is actually localized around the instructed location, thus indicating that attention is not always anchored at the cued location. Importantly, decoding output still strongly accounts for behavioral performance, such that the closer the decoded attention to the cued location at time of target presentation, the lower the probability that the monkeys produce a miss (so the higher the probability of a correct detection, **Figures 2A,B**) (Astrand et al., 2016, 2020). Based on recent advances in machine learning (Lemm et al., 2011; Abraham et al., 2014; Savage, 2019; Glaser et al., 2020; Iturrate et al., 2020), we thus reasoned that not all correct trials shared the same degree of attention-related information and we trained a second decoder selecting only those correct trials that predicted attention closest to the cued location from the initial decoding step (De Sousa et al., 2021). This remarkably enhanced the degree of correlation between the distance between decoded attention and the cued location on the one hand and success of the subjects in the task on the other hand. This was true whether attention was decoded from MUAs, or from LFPs (Esghaei and Daliri, 2014; Seif and Daliri, 2015; De Sousa et al., 2021), suggesting that such achievements could also be expected from ECoG, EEG or MEG signals. Indeed, pioneer studies have obtained remarkable accuracy thresholds in decoding attention

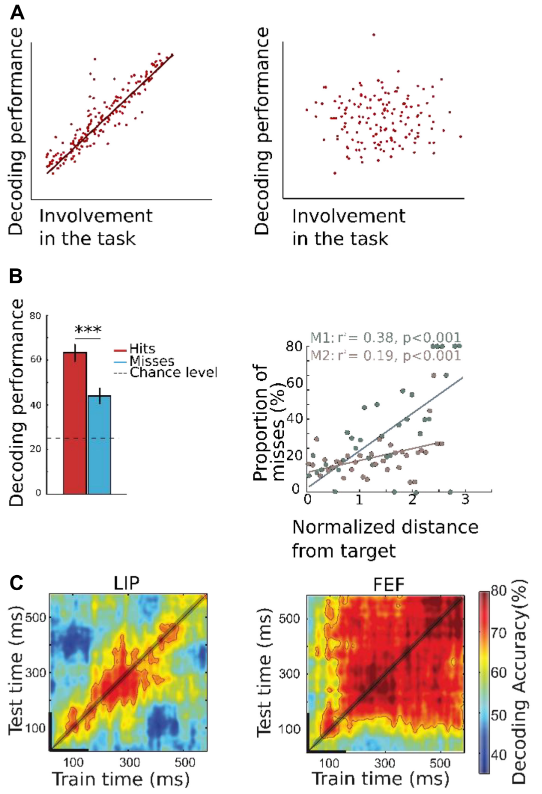


FIGURE 2

(A) Schematic representation of theoretical relationship between decoded information and behavioral proxy of decoded internal brain state. Left: Decoded information is highly correlated with the underlying cognitive brain state, and the behavioral proxy of this decoded brain information (hit, misses, reaction time) is strongly correlated with the decoded information. Right: If the decoder is not a good estimator of covert brain information, there is no correlation between decoded information and the behavioral proxy to the decoded cognitive function. Please note that this type of approach is more robust when the output of the decoder is continuous and non-categorial. **(B)** Decoded spatial locus of attention correlates with behavioral performance. Decoding was performed using MUA activity recorded from two rhesus monkeys, while the animals played a cued target detection task. On each trial, a cue was presented to the monkeys who were trained to keep their gaze at the center of the screen while orienting their covert attention to the cued location (one amongst four possible locations), to detect. The trained algorithm aimed at decoding the quadrant to which attention was oriented (left) or the actual (*x*-,*y*-) location of the attentional spotlight (right). Left: Decoding accuracy as a function of behavioral performance. Decoding accuracy is higher for correct detection trials than for miss trials, in which the monkeys did not properly identify the target, indicating that the decoding algorithm is able to capture internal attentional orientation information. Right: Proportion of misses as function of distance between the target and the decoded position of attention in *X*-*Y*- coordinates. Farther is the decoded position from the target at time of presentation, higher is the probability of missing the target (adapted from Astrand et al., 2016).
(C) Temporal generalization of two algorithms trained with MUA activity from two different brain regions known to have a role in covert attention: lateral intraparietal area LIP (left) and frontal eye field FEF (right). The algorithm is trained on successive times intervals and for each training interval, tested on all available

(Continued)

FIGURE 2 **(Continued)**

time intervals (cross-temporal decoding). This highlights two distinct neuronal population coding schemas: a dynamic coding in LIP whereby the neuronal code for attention orientation at a given time interval does not generalize to other time intervals, and a static coding in FEF, whereby the code identified at a given time generalizes at all times following cue presentation (Astrand et al., 2015). ***$p < 0.001$.

using non-invasive recordings such as EEG (Treder et al., 2011; O'Sullivan et al., 2015) and MEG (Battistoni et al., 2018; Desantis et al., 2020). The above studies rely on attentional tasks such that decoding attention accuracy can be confronted with the resultant behavioral attentional bias. Quite interestingly, decoding of attention can be achieved even when attention does not bias behavior, i.e., when subjects are engaged in cognitive tasks that do not explicitly monitor the attentional function (Westendorff et al., 2016), suggesting that decoding attention can be performed outside of a controlled laboratory setup.

A third reason why decoding of cognitive variables might be suboptimal is our proper understanding of how this function interacts with other cognitive functions or a proper understanding of its spatial and temporal properties. For example, it has been very intriguing to us to observe that error trials could still be produced even when spatial attention was decoded close to the cued location (Astrand et al., 2016, 2020; De Sousa et al., 2021). De-mixed dimensionality reduction approaches (Kobak et al., 2016), allowed us to demonstrate that spatial attention orientation organizes in the prefrontal cortex distinctly from engagement in the task, such that target miss-detections could arise from both an inappropriate allocation of attention or an inappropriate engagement in the task (Amengual and Ben Hamed, 2021; Amengual et al., 2022). This biologically-inferred decoding schema further enhances our accuracy at tracking the actual spatial spotlight of attention and better account for its contribution to overt behavior.

Another example is the recent understanding we gained on the dynamical structure of decoded spatial attention exploration and exploitation (Gaillard and Ben Hamed, 2020; Gaillard et al., 2020), corroborating a large field of behavioral (VanRullen et al., 2007; Landau and Fries, 2012; Fiebelkorn et al., 2013, 2018; Dugué et al., 2015, 2016; VanRullen, 2016, 2018) and electrophysiological (Fiebelkorn et al., 2018) body of research. Specifically, we show that rather than being stable at a given location in space, the attentional spotlight explores space rhythmically, at 8 Hz, alternating between epochs of exploration away from the task relevant locations, and epochs of exploitation, at task relevant locations. Considering this dynamic nature of prefrontal attentional information results in variations by up to 10% of decoding accuracy and enhances the predictive power of whether the subjects will correctly respond to a target, miss the target or else miss-respond to a distractor.

What cross-temporal decoding tells us about the brain

Cognitive processes in the brain are non-stationary and they evolve in time. In this context, neural decoding has shown to be a powerful tool to describe the temporal dynamics of coding information, describing the evolution of information decoding performance in time. For example, Bae (2020) studied the different time course differences between face identification (recognition) and face expression (emotion) processes using EEG recordings. In this study, participants performed a working memory task in which they were asked to remember a face image presented in the screen and, after a short delay, they were asked to report either the identity or the expression of the face. Importantly, participants had no *a priori* knowledge about the information they would be asked to report, thus, they had to extract and maintain both types of information during each trial using whole scalp raw data. In order to decode these two types of information, Bae and Luck (2018) used a combination of support vector machine and error-correcting output coded (Dietterich and Bakiri, 1994) to classify either facial expression of facial identification by using the scalp distribution of the phase-locked ERP voltage in the alpha-band activity (8–12 Hz). The decoding performance of both types of information was tracked in time, showing that these types of information exhibited a dissociated temporal dynamic. More specifically, decoding of the identification of the face was more prominent only during the time interval corresponding to the perception of the image, while the decoding of the face expression was more prominent during working memory maintenance, thus, along the whole trial duration. This result suggests that the neural representation of face identity and face expressions were, at least partially, independent. In addition, Bae and Luck (2018) uses the decoding performance as a tool to understand the temporal dynamics of encoded information not only within the same trial, but also between trials. Interestingly, they succeeded in decoding the identification of the face in the current trial using the information encoded from the previous trial. This did not apply to the decoding of face expression information. Therefore, this result suggested that neural decoding might be a useful tool to study how information encoded in the past can be reactivated regardless of its relevance for the current goal of the task.

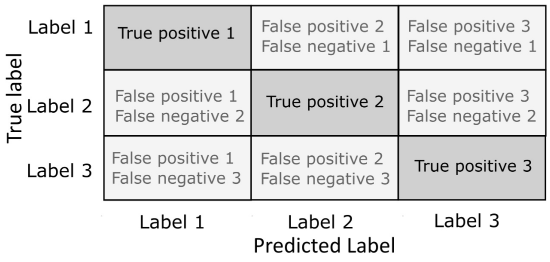
Other studies have used neural decoding to study the dynamics of the hippocampal replay. Davidson et al. (2009) recorded multiple single unit activity in the hippocampal area CA1 in rodents while those were exploring a track. Using a probabilistic neural decoding strategy to estimate the animal's position on the track from the spike trains, they evaluated whether recorded cells replayed spatial memory sequences (Brown et al., 1998; Zhang et al., 1998). Therefore, they conceived a neural decoding approach specific for replay detection. Interestingly, when rodents stopped exploring the environment, they showed signatures of time-compressed forward and reverse hippocampal replay of long behavioral sequences that, in turn, were associated with trains of ripple events. In addition, they found that replay was neither limited to locations associated with the reward, nor to those locations tied to the animal's current location. Other studies have shown evidence of replay using non-invasive brain recordings in humans. Kurth-Nelson et al. (2016) recorded MEG activity from a cohort of human subjects while they were performing a non-spatial reasoning task. The aim of this study consisted in finding sequences of neural representations associated with learning and online planning, similarly to those found by replay, but in a non-spatial context. Indeed, the task required selecting paths through a set of six visual objects. They trained pattern classifiers on the MEG activity evoked by the direct presentation of the objects alone. Posteriorly, they tested these classifiers using the activity obtained during periods when no object was presented. They show that brain activity encodes the representation of at least four objects that were presented sequentially, following backward trajectories along the paths in the task. This was one of the first studies showing clear signatures of replay using non-invasive methods of brain recordings.

Decoding can also be used to investigate brain networks dynamics and coding regime. In this context, several studies have used cross-temporal decoding (King and Dehaene, 2014; Astrand et al., 2015; Amengual and Ben Hamed, 2021), which consists in training models at a given time in the trial and testing these models all throughout the trial time. This method allows discriminating different computational properties of the neural population, between stable coding regime (whereby the code identified at a given time generalizes at all times, indicating a stable coding schema by the underlying neuronal population, **Figure 2C**), and dynamic computational/cognitive processes (whereby the neuronal code identified at a given time does not generalize at other times, indicating a recurrent dynamic coding schema by the underlying neuronal population; **Figure 2C**). Astrand et al. (2015) studied the different dynamics of population coding during a spatial attention task using parietal and prefrontal electrophysiology recordings in macaques. Cross-temporal decoding matrices were used to determine whether tow brain regions, the frontal eye field (FEF) and the lateral intraparietal area (LIP) were presenting stable or dynamic coding. They show that the spatial attention code identified in the FEF at any given time can generalize to other times, thus indicating a stable coding schema. Indeed, this coding regime characterizes regions with activation dynamics mimicking those observed in artificial recurrent neural networks (Buonomano and Maass, 2009). Conversely, population activity in LIP showed a dynamic coding regime when decoding spatial attention, variable from one time to the next within the same trial. Therefore, these results show that neural decoding is a very suitable tool to study how different neural populations encode the same type of information by using different coding regimes.

In this line, other studies have addressed the question on how decision making might exploit cognitive flexibility to adapt behavior. Using time-resolved population-level neural pattern analysis from intracranial recordings from the prefrontal cortex in macaques, Stokes et al. (2013) investigated how context is encoded and maintained in the neural population in order to be exploited behaviorally. Specifically, they show that an instruction cue indicating the context of the trial in the task triggers a dynamic coding at trial onset, while the same information during the delay period prior to the decision is coded by a stable low energy state (Stokes et al., 2013). This method allowed the extraction of hidden patterns of activity in the data characterized by high and low energy states that tuned the prefrontal cortex according with the task demands. This observation extends to human brain function. This tuning mechanisms by hidden population states has been also explored in humans using non-invasive electrophysiology. In a series of studies, Wolff et al. (2015, 2017) ask participants to perform a working memory task while they record concomitant brain activity using EEG and MEG. In some of the trials, they presented a non-informative image (called impulse) during the period of working memory maintenance. They trained a decoder to predict the memorized item using activity previous to the impulse presentation (pure memory activity), or on the activity elicited by the task-irrelevant probe The authors show that the impulse generates a dynamic coding of the memory item which does not generalize to the other testing times. Taken together, these studies support the theory of hidden coding low energy states, whereby working memory information is encoded dynamically and reactivated by task-relevant but also by task irrelevant items (Wolff et al., 2015, 2017).

What confusion matrices tell us about the brain

Until know, we have discussed how decoding performance and their variation in time provides unique information about the cognitive processes underlying behavior. The question we will tackle in this section is whether we can extract genuine knowledge about cognition by studying how decoding algorithms fail in classifying information on which they have been trained. Indeed, studying how a decoder fails may also be very informative about cognitive brain processes. In this context, confusion matrixes are used to quantify the miss-classifications produced by a given decoder (Kriegeskorte and Kreiman, 2011). Confusion matrices allow a fine-grained analysis of the performance of the decoding algorithms in terms of hit rates and errors. Specifically, it permits to visualize the number of correct and incorrect prediction for each label (**Figure 3**). In a confusion matrix, each row represents one of the labels to be predicted, and each column represents the actual output of the classifier. In other words, for any given label presented along the rows, the confusion matrix presents the count of classification outputs assigned to that specific label but also to each of the other labels. A label correctly classified is considered as a true positive classification. When a label is not correctly classified, this is considered as a false negative for the targeted label, and as a false negative for the predicted label. For example, the study of confusion matrices when decoding covert attention using fMRI shows clear attentional biases toward the lower visual field or along the horizontal and vertical meridians such that decoding accuracy is up to 10% higher at these locations relative to other locations in the visual field, thus confirming attentional biased observed behaviorally (Zenon et al., 2008; Zénon et al., 2009; Loriette et al., 2021). In another fMRI study, Kim et al. (2019) studied how the brain visually encodes tactile intensities. To do this, they used an associative learning method to decode the representation of tactile intensities (roughness) evoked either by tactile exploration only, or by visual observation of the tactile exploration. In this work, they show that the behavioral data obtained while evaluating roughness during tactile exploration or visuo-tactile exploration correlates with the confusion matrices obtained when they decoded the roughness based on fMRI brain activity. In particular, this correlation was specific of the supramarginal gyrus, suggesting its role in tactile discrimination (Kim et al., 2019). Confusion matrices have also provided insightful information for the development of categorical models of emotions. Saarimäki et al. (2016) acquired fMRI data while participants participated in a task consisting in the identification of different fine emotions such as "joyful," "amazed," or "nervous," embedded into basic emotion groups like "happiness," "surprise," or" fear." Using multivariate pattern analysis (MVPA), they classified emotions from the activation in different brain areas including medial and inferior lateral prefrontal cortices, frontal pole, precentral and postcentral gyri, precuneus and posterior cingulate cortex. Importantly, participants behaviorally tended



	Predicted Label		
True label	Label 1	Label 2	Label 3
Label 1	True positive 1	False positive 2 False negative 1	False positive 3 False negative 1
Label 2	False positive 1 False negative 2	True positive 2	False positive 3 False negative 2
Label 3	False positive 1 False negative 3	False positive 2 False negative 3	True positive 3

FIGURE 3
Confusion matrices. The instances of correct classification (true positives) and misclassifications (false negatives or false positives) are represented in a matrix with each row representing the actual true labels to be classified and each column representing the predicted label by the outcome of the classifier. Incorrect classification can be defined against the actual true label as a false negative or against the predicted label as a false positive.

to misclassify emotions which are in the same group and these misclassifications were similar to those obtained by decoding emotions from fMRI signal, suggesting a link between the activity in these regions and the emotional perception (Saarimäki et al., 2016). Electrophysiological studies have been benefited as well from the use of confusion matrices to evaluate the decoding capacities of the brain signal. Chen et al. (2019) developed a LFP-based close-loop deep brain stimulation strategy in Parkinson patients implanted with electrodes in the subthalamic nucleus. The aim of this method was to adapt the stimulation interval based on the sleep stage (Amara et al., 2016). Therefore, machine learning methods were applied to decode in real time the sleep stage of the patient from the LFP recorded from the DBS leads. Based on different combinations of signal features extracted within the time domain, frequency domain and variance of the signal, they succeeded in decode the sleep stage from the LFP in this region using a SVM-based algorithm. However, confusion matrix revealed that not all sleep stages were decoded with the same accuracy: while the decoder was able to accurately classify wakefulness against sleep and N2 phase against REM, phase N1 and REM showed a high degree of confusion, although still above chance level. Therefore, this study showed that subthalamic nucleus LFP-based activity encode information about the brain state activity (but only some of these states). Other studies have addressed the question of the regional- and modality-specificity of the decoding capacities of the cortex. Enander et al. (2019) delivered a set of electrical spatiotemporal tactile afferent activation patterns to the skin of the contralateral second digit of the forepaw in anesthetized rats. Concomitantly to this stimulation, there performed *in vivo* single neuron recordings in the right hemisphere, more specifically in the primary somatosensory cortex, but also in other cortical locations within and out of the somatosensory cortex. They used the recorded brain activity of these neurons to decode the different stimulation patterns delivered by the electrical stimulation. Interestingly, they found that neurons from the primary sensorimotor cortex showed similar decoding performance compared with neurons from out of this area. Indeed, confusion matrix revealed that neurons from regions of the visual cortex showed less miss-classification rates than those in the sensorimotor cortex. This study provided direct evidence on how the tactile information could be propagated globally across the neocortex, presumably *via* cortico-cortical but also cortico-thalamo-cortical pathways (Lübke and Feldmeyer, 2007; Frostig et al., 2008).

Exploring shared functional substrates amongst different cognitive tasks

Another possible application of decoding methods is to evaluate the decoding performance of a decoder trained to decode a specific type of information (e.g., position) of a cognitive process (e.g., memory) and testing this neural decoding on a different information (e.g., color) or cognitive modality (e.g., attention). By using this method, it is possible to identify informational communalities between both sources of information or between two cognitive processes. We selected two different fMRI studies that exemplify this point (Albers et al., 2013; Dijkstra et al., 2019). In these studies, participants were presented either with a visual stimulus, or asked to mentally imagine these stimuli. Interestingly, both studies succeeded in decoding each of the two cognitive processes (mental imagery and visual perception) by using a decoder trained on either mental imagery brain activities or visual perception related brain activities, indicating that mental imagery coding and visual perception share similar cortical representation.

Exploring model parameters to model the brain

Until now, we have shown that neural decoding accuracy provides information about a broad series of cognitive processes and their neural underpinnings. However, other elements associated with the neural decoding algorithms, such as the parameters obtained from the decoding model (e.g., the weights that are fit in linear regression), can be exploited to study cognitive brain functions. One of these parameters is the number of extracted features. This number is obtained by a process consisting in extracting a small number of features that maximizes the information about the statistical structure of the data. Numerous methods are available in order to perform feature selection, ranging from statistical test as ANOVA, to principal component analysis (PCA), mutual information maximization, searchlight and others (Pedregosa et al., 2011; Abraham et al., 2014; Allefeld and Haynes, 2014; Cunningham and Yu, 2014; Padmanaban et al., 2018). For example, in fMRI, an ANOVA-based feature selection precisely reveals the cortical topography for covert visual attention to guide single trial fMRI-based spatial decoding of attention (Loriette et al., 2021). Another type of feature selection, called searchlight, aims at searching for the most informative features and select them to train the decoder, by looking at how each feature separately contributes to improving the classification accuracy. Stokes et al. (2009) investigated mental imagery in an fMRI study with human participants. Authors compared the classification accuracy in different fMRI regions while participants imagined different letters. This method allowed to identify high informative areas in mental imagery [including inferior occipital gyrus (IOG), middle occipital gyrus (MOG), fusiform gyrus (FG), middle temporal gyrus (MTG) and temporal gyrus/Heschl's gyrus (STG/HG)]. When comparing these areas to the ones obtained from a searchlight procedure on brain activity elicited by visual stimuli, the authors further identify cortical regions that are involved in mental imagery and not in pure visual perception (STG/HG) (Stokes et al., 2009). Another study compared searchlight method to 3D classical fMRI analysis, showing that the searchlight method is more spatially specific compared to classical methods (Chen et al., 2011). This result indicates that this method outperforms classical statistical tests and reveal very precise local functional selectivity patterns in brain areas that were initially thought as functionally homogenous. Other studies have succeeded in using EEG recordings to decode semantic information. One of the relevant questions here is what kind of signal (or pattern) might provide more exploitable semantic information to be used for decoding. For instance, Jafakesh et al. (2016) used different features of the recorded EEG signal to decode the semantic category of different visual stimuli. Specifically, they used several within electrode cross-frequency coupling (CFC) measures such as amplitude-amplitude coupling (AAC), phase-amplitude coupling (PAC), and phase-phase coupling (PPC) within each electrode and used them as input to SVM classifier. They found a higher decoding performance using PPC than using the other two measures, specifically in the alpha and gamma frequency bands. In addition, they tested whether using CFC-based measures to classify semantic information outperformed the decoding with respect of using wavelet transform of the EEG signal. In this context, they obtained a higher decoding performance using PAC relative to using wavelet coefficient. Therefore, CFC measures provide information regarding semantics that is not available in the time-frequency components.

Although there is not a clear consensus on their interpretation, the weights of the decoder are often used to analyze informative cortical voxels, under some conditions. Generally speaking, weights represent the accountability of the information content of each feature which is fed into the decoding algorithm (Kriegeskorte and Bandettini, 2007; Haufe et al., 2014; Kia et al., 2017; Hebart and Baker, 2018). As an example, weights extracted from Linear Discriminant Analysis (LDA) model in EEG while performing gesture classification at the single trial level reveals the frequency and the channels which are the most informative during walk preparation (Velu and de Sa, 2013). In contrast, weights extracted using a linear regression analysis on intra-cortical neuronal data to decode spatial attention from the single trial activity of a prefrontal or a parietal neuronal population can only be interpreted when normalized by the average response of each neuron (Astrand et al., 2015). As a result, high weights associated with weak average neuronal responses might turn out less informative than smaller weights associated with high average neuronal responses. In other types of decoders yet, weights cannot be readily interpreted (Haufe et al., 2014; Kriegeskorte and Douglas, 2019).

Interfering with brain activity with neurofeedback or learning

Brain–machine interfaces

Although the scope of this review is to present the neural decoding as a tool not only used to develop strategies for neuroprosthetics but also to understand cognitive function, we found it necessary to discuss some examples on how decoding of cognitive information has been used to develop brain–computer interfaces (BCI) in turn producing knowledge on brain organization and plasticity. BCIs are direct or indirect communication interfaces between the brain and a computer. These methods rely on closed-loop systems, which refers to the fact of providing to the subject a direct (e.g., MUA or BOLD activation level in a specific cortical region) or an indirect (e.g., decoded information across a set of MUA, EEG or BOLD signals, or signal coherence across multiple EEG channels) feedback extracted from brain activity (Chaudhary et al., 2016). One example of BCIs is the neuroprosthetic BCIs, aiming at replacing a deficient brain function. For example, motor neuroprosthetics have been developed to allow tetraplegic patients to control a robotic arm thanks to the real-time decoding of ECoG recordings (Hochberg et al., 2012; Bensmaia and Miller, 2014). Sensory neuroprostheses have also been developed to restore tactile sensation, for example injecting complex microstimulation patterns into the somatosensory cortex of macaque monkeys to generate artificial sensations guided by touch sensors implemented in a robot arm (O'Doherty et al., 2011) (**Figure 4**).

These decoding approaches allow not only to explore brain functions but also to train them, by providing a feedback to the participant or the animal about their brain activity (raw or processed/decoded) for them to act on it to improve behavior. This closed-loop procedure is called neurofeedback (Sitaram et al., 2017). One example of these neurofeedback tools is the P300 speller, which is an EEG-based BCI. This method consists in spelling a word that the participant has in mind by attending to a target letter presented in the midst of other letters (all alphabet letters being covered sequentially) by flashing a sequence of letters alternating between rows and columns. When the letter selected by the participant is flashed, the evoked response in the brain is stronger and the decoder can exploit this difference to identify the selected letter amongst the presented letters (Guy et al., 2018). Performing with the P300 speller trains subjects to flexibly use their attentional resources and this results in an enhancement of their attentional performance beyond this specific task (Arvaneh et al., 2019). Neurofeedback has also been implemented in fMRI protocols on categorical attention in order to increase behavior as well as attention-related brain information as assessed by the decoder's accuracy (deBettencourt et al., 2015).

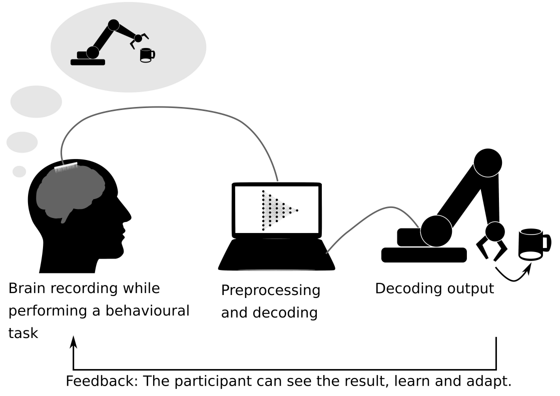


FIGURE 4
Schematic representation of a closed-loop brain–machine interface. The brain activity is recorded while the subject is performing a behavioral task. Here, the participant performs a motor-task and tries to grab a coffee cup with a robotic arm, while her/his brain activity is being recorded using ECoG sensors. The recorded data is pre-processed and fed into a machine learning algorithm previously trained to interpret the participant's motor intentions (Hochberg et al., 2012). The decoded movement is translated into a robotic arm movement. The actual movement of the robotic arm serves as a feedback for the participant on her/his brain activity, allowing her/him to learn control the device.

Importantly, neurofeedback has additionally been used in order to investigate the effects of interfering with brain functions. For example, an fMRI study shows that it is possible to selectively increase or decrease a participant's confidence in her/his performance during a behavioral task without interfering in her/his actual success rate in the task using neurofeedback specifically based on the decoded participant's confidence (Cortese et al., 2016). In another fMRI study, participants were able to learn to associate red or green color with different grating orientations. Specifically, after training a decoder to discriminate the color the participants were presented with (green or red), participants were trained to modulate their brain activity while a grating with a specific orientation was displayed. If the grating was vertical, participants had to modulate their brain activity in order to increase the likelihood of decoding red color in the visual cortex. If the grating was horizontal, participants were asked to increase the likelihood of decoding "green." As a result, after neurofeedback training, when presented with vertical or horizontal achromatic gratings, the participants tended to perceive them in the color they were trained to modulate their activity in association with the grating orientation. Surprisingly, these effects persisted up to 5 months (Amano et al., 2016).

There is an ever-growing number of neurofeedback studies and tools aiming at both training cognitive functions and understanding brain functions. In particular, this approach has been shown to potentiate brain plasticity, triggering not only behavioral long term effects, but also plastic changes in brain structure and anatomy, such as an increase in gray matter volume and white matter myelinization (Ghaziri et al., 2013; Marins et al., 2019; Loriette and Ziane, 2021).

An important field of application of brain decoding and neurofeedback relies on using brain–machine interfaces (BCIs) in order to restore an acute or chronic brain deficits for neurorehabilitation (Chaudhary et al., 2016; Lebedev and Nicolelis, 2017; Bockbrader et al., 2018). Impressive advances have been achieved in BCI-driven motor rehabilitation after stroke (Wang et al., 2018) or spinal cord injury (Pohlmeyer et al., 2009; Bensmaia and Miller, 2014; Ajiboye et al., 2017). As an example, Pohlmeyer et al. (2009) used a cortical controlled electrical stimulation of the forelimb of spinal cord injured macaques and managed to restore voluntary movement. Another study using EEG BCI coupled with electrical stimulation restored hand movements in patients after a stroke injury, this restoration persisting 6–12 months after the BCI training (Biasiucci et al., 2018).

Manifold coding and learning, a new approach to understand cognition

Recording brain activity with the most up-to-date recording methods result in very high dimensional data sets. This can be a limiting factor in decoding speed. Techniques such as principal component analysis (PCA) (Cunningham and Yu, 2014), independent component analysis (ICA) (Brown et al., 2001) and their derivatives have been developed to reduce the dimension of the dataset. Importantly, such dimensionality reduction methods, in addition to compressing dataset size provide a very unique insight in how the neural code is encoded in the brain. These low-dimensional spaces which allow to analyze patterns of activity and interaction between neurons are called neural manifold (Elsayed and Cunningham, 2017; Degenhart et al., 2020). These manifolds can be seen as low dimensional spaces that reflect neural modes (**Figure 5A**), that is to say, patterns of activity corresponding to cognitive actions which can be embedded into a surface. The response activity from one cognitive task (for example moving the arm) can be tuned differently depending on little variations of this task but the activity remains embedded in the same surface, while a second action (for example walking around) will be embedded in a different manifold. As an example, moving an arm in a certain direction can be seen as a trajectory in a three-dimensional surface. Each movement direction is embedded in this surface but can be separated visually and mathematically one from the other (Gallego et al., 2017, 2018) (**Figure 5B**). While dramatically enhancing our understanding of how the brain copes with the dimensionality of the information, decisions and cognitive functions it has to implement (Rigotti et al., 2013), it also allows to enhance brain computer interfaces techniques, accelerating computational speed as well as allowing

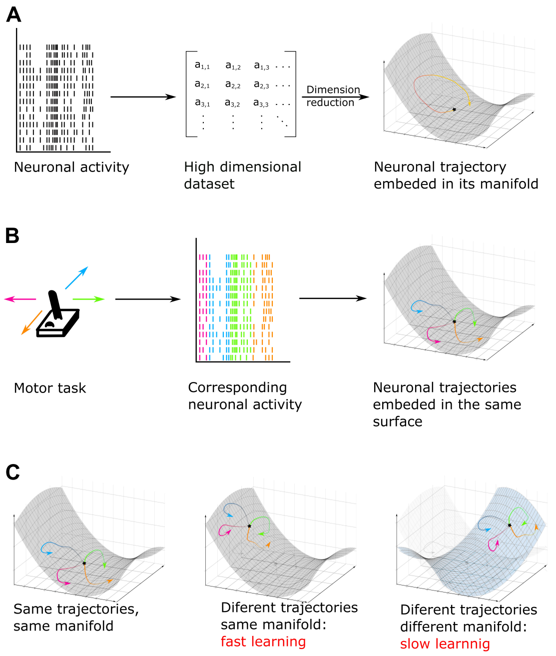


FIGURE 5
(A) Neuronal activity, here MUA recordings, taken as an example, is converted into a lower dimensional space called manifold. In this low dimensional surface, the neural code can be represented as data points or trajectories when the pattern of activity is analyzed over successive time epochs. **(B)** Example of a motor task. Here, neuronal activity is recorded while subjects move a joystick in one of four possible directions. On individual trials, each movement can be represented by a different pattern of activity or trajectory but all trajectories are embedded in the same neuronal low dimensional manifold neural space [based on Gallego et al. (2017, 2018)]. **(C)** Example of cognitive training using manifold perturbation. Participants or animals are given a feedback (for example, cursor velocity) and have to learn to control it by using the same neuronal pattern as recorded during previous tasks (left) (for example, during a motor task). It is also possible to train subjects by using a different pattern embed in the original manifold (center) or in a new manifold surface (right). The last option leads to a slower learning and an increased difficulty of the task (based on Sadtler et al., 2014; Golub et al., 2018; Oby et al., 2019).

to cope with neural instability by realigning recorded activity to its original manifold before performing decoding, thus maintaining constant decoding accuracy in time (Degenhart et al., 2020; Gallego et al., 2020).

The theory of neural patterns embedded into manifolds has been specifically applied in order to investigate learning, which by definition impacts neuronal computation as learning progresses. For example, researchers have used brain computer interfaces in order to train monkeys to control the velocity of a cursor by using the neuronal activity generated by their motor cortex. They first decoded arm velocity from the neuronal activity of the arm region of the primary motor cortex. Then, they trained the monkey to control the velocity of the cursor, not based on the pre-existing code assessed in the first step, but either using a new activity pattern embedded into the original manifold, or using a new activity pattern positioned outside of the original manifold (**Figure 5C**). They found that monkeys learn faster if the new pattern of activity they have to learn is embedded in the original manifold (Sadtler et al., 2014; Golub et al., 2018). Moreover, researchers found that forcing the monkey to learn a new neural code out of the original manifold is not only slower, but leads to the emergence of a new neural mode, in other words, a new manifold in which the neural code will be embedded (Oby et al., 2019). Overall, these studies result in a new understanding of how learning is implemented in the brain and how it can best be potentiated. This substantiates why learning complex and new tasks is slower than learning familiar tasks, as this requires creating new patterns of activity that do not interfere with prior learning (**Figure 5C**).

Deep learning: A window onto the complexity of brain functions

The vast majority of decoding algorithms used in the examples cited in this review are based on learning certain rules of inference that are used to estimate linear or non-linear prediction functions to map input activity onto a set of outputs. Importantly, these algorithms do not learn neural representations directly, but are trained to determine decision boundaries that are used to classify the inputs onto specific outputs (Amengual and Ben Hamed, 2021). However, some non-linear properties of these representation and their large dimensionality prevent the optimal performance of these classification algorithms and, thus, additional preprocessing methods based on reducing data dimensionality and feature extraction techniques are needed (Blum and Langley, 1997; Schölkopf et al., 2007; Abrol et al., 2021). The application of these preprocessing steps on input data imposes certain *a priori* assumptions and requires a certain high expertise of the users, which reduces part of the automatization of the knowledge extraction process. As a response to this need, deep learning approaches have been introduced for neural decoding purposes. Differently to standard machine learning approaches, deep learning characterizes patterns embedded in the raw data as a part of the training process. To do this, deep learning models are based on multiple layers artificial neural networks (ANN) that allow to progressively extract high-level features from input data. These models consist in a composition of components that are formed by linear and non-linear operations forming complex layered architectures. Some examples of these networks are the recurrent neural networks (RNN), which are suitable to model the temporal dynamic behavior of a given process, convolutional neural networks (CNN), most commonly applied to analyze visual imagery and long short-term memory (LSTM) networks, a particular case of RNN with feedback connections, suitable to model memory processes. In the following we will discuss some applications of the deep learning methods for neural decoding in vision.

Indeed, decoding visual stimuli, understood as the capacity to predict the identity (meaning) or physical attributes of visual stimuli by using brain activity, represents a major challenge in neuroscience. Seminal fMRI studies have shown that visual features such as orientation, motion direction and visual object categories can be decoded from BOLD signal recorded over the visual cortex and ventral parietal cortex (Haxby et al., 2001; Cox and Savoy, 2003; Kamitani and Tong, 2005, 2006). However, these studies used the activation from voxels in selected visual cortices to feed into the decoding algorithm and, thus, they did not take into account the internal relationship between the different visual areas. Indeed, it is well known that a higher level function such as object recognition requires the co-activation of different brain areas in a hierarchical manner along the ventral stream (Mishkin et al., 1983). Anatomical studies have found that connections between the different layers of the ventral stream are bidirectional (Bar, 2003). These forward and backward connections provide the anatomical substrate of the information flow in the visual cortex. In this context, visual information might flow from primary visual cortices toward high-level visual cortices to obtain high-level semantic understanding [bottom-up visual mechanisms, (Logothetis and Sheinberg, 1996)]. In turn, visual information feedbacks from high-level to low-level visual cortices, which is known the top-down visual mechanisms (McMains and Kastner, 2011). In order to succeed in neural decoding of object recognition, Qiao et al. (2019) conceived the use of a RNN in which they split the neurons into positive and negative directions and fed the activity of each voxel in each visual area while participants observed natural photographs. In this way, they did not model only the information of each visual area, but also the internal relationship between the different visual cortices, in the decoding method. Comparing the decoding accuracy of this method with the accuracy obtained using other classical classifiers such as decision trees and random forest, they found that this method of decoding improved the classical decoding methods by a 5% of accuracy, on average. In addition, they concluded that the representation in visual cortices were hierarchical, distributed and complementary, since the increment in decoding performance depended on the conception of the multiple layers simultaneously. Other studies have tried to use deep learning approaches to decode brain's responses to representation of natural video stimuli. To this end, Wen et al. (2018) acquired very long fMRI acquisitions of three human subjects watching 972 different video clips that included diverse scenes and actions. The aim of this study was to use a convolutional neural network (CNN) in order to reconstruct and categorize the visual stimuli based on the fMRI activity recorded from the dorsal and ventral streams in a dynamic condition. They found that the CNN was able to predict non-linear and complex patterns of responses in both dorsal and ventral streams, with a high decoding accuracy in category representation. Indeed, the CNN supported the reconstruction of decoded natural movies and direct semantic categorization. All in all, these studies exemplify how deep learning algorithms can decode visual information with a very high degree of specificity. It is expected that such methods can be generalized in the future to the read out of more complex cognitive functions.

Conclusion

Advances in brain activity recording and processing and in machine learning have led, in these recent years, to a new way of exploring brain function. In fact, it is now possible to access to a better understanding of brain function using recorded activity, as for example while trying to infer the spatial location of covert attention in real-time in macaques or humans. Decoding neuronal brain activity can be performed with both invasive and non-invasive techniques, each presenting its pros and cons. It permits to have a direct access to a part of brain information and build brain–machine interfaces as for example, controlling a robotic arm with motor cortex activity. But these major advances performed in this field in these last years did not only permit to perform "mind reading" of the brain. These methods have also generated robust statistical tools to better understand brain function and cognition. In this review, we have explored brain decoding approaches, not only from the perspective of inferring hidden brain states but also from the perspective of understanding brain functions. For example, exploring decoding accuracy allows to explore the temporality and the stability of brain processes (Astrand et al., 2015; Wolff et al., 2015, 2017) while searchlight methods or decoder weights analysis allows to extract a refined view of how the brain organizes information processing at a high spatial resolution (Chen et al., 2011; Haufe et al., 2014). The development of these exploratory methods has resulted in new hypotheses about how neural networks can encode a given function and even how this code can be modified by learning. As an example, numerous articles are now exploring the theory of manifolds embedding neural activity (Gallego et al., 2017, 2018).

These advances in decoding and computational neurosciences open the way of combining different brain activity modalities when exploring any given function (e.g., EEG and fMRI or MUA and LFP). Indeed, each recording technique can bring specific information. This is expected to enhance our understanding of brain functions and allow to explore differences in information content between data collected simultaneously in different modalities. One thinks of differences in temporal and spatial resolution, but other functional differences are also increasingly reported, as for example in the informational content of spikes versus LFP (Pesaran et al., 2002; Perel et al., 2013).

The road is still long before a full exploitation of all of the potentialities of this new research field which combines a mechanistic understanding of the brain with machine learning tools (Pedregosa et al., 2011; Savage, 2019; Glaser et al., 2020; Iturrate et al., 2020). Brain computer interfaces and neurofeedback protocols are still at their early days and will probably benefit from the continuous progresses observed in computational sciences.

Author contributions

CL and SBH: conception of the review. CL, SBH, and JA: write draft and corrections. All authors contributed to the article and approved the submitted version.

Funding

CL, JA, and SBH was supported by ERC BRAIN3.0 # 681978 and ANR-11-BSV4-0011 and ANR-14-ASTR-0011-01, LABEX CORTEX funding (ANR-11-LABX-0042) from the Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR) to SBH. CL was supported by a Ph.D. funding from the French Ministry of Research and higher Education.

Conflict of interest

Publisher's note

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., et al. (2014). Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* 8:14. doi: 10.3389/fninf.2014.00014

Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., et al. (2021). Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. *Nat. Commun.* 12:353. doi: 10.1038/s41467-020-20655-6

Adewole, D. O., Serruya, M. D., Harris, J. P., Burrell, J. C., Petrov, D., Chen, H. I., et al. (2016). The evolution of neuroprosthetic interfaces. *Crit. Rev. Biomed. Eng.* 44, 123–152. doi: 10.1615/CritRevBiomedEng.2016017198

Ajiboye, A. B., Willett, F. R., Young, D. R., Memberg, W. D., Murphy, B. A., Miller, J. P., et al. (2017). Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: A proof-of-concept demonstration. *Lancet Lond. Engl.* 389, 1821–1830. doi: 10.1016/S0140-6736(17)30601-3

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., and de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Curr. Biol.* 23, 1427–1431. doi: 10.1016/j.cub.2013.05.065

Allefeld, C., and Haynes, J.-D. (2014). Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. *Neuroimage* 89, 345–357. doi: 10.1016/j.neuroimage.2013.11.043

Amano, K., Shibata, K., Kawato, M., Sasaki, Y., and Watanabe, T. (2016). Learning to associate orientation with color in early visual areas by associative decoded fMRI neurofeedback. *Curr. Biol.* 26, 1861–1866. doi: 10.1016/j.cub.2016.05.014

Amara, A. W., Walker, H. C., Joop, A., Cutter, G., DeWolfe, J. L., Harding, S. M., et al. (2016). Effects of subthalamic nucleus deep brain stimulation on objective sleep outcomes in Parkinson's disease. *Mov. Disord. Clin. Pract.* 4, 183–190. doi: 10.1002/mdc3.12375

Amengual, J. L., and Ben Hamed, S. (2021). Revisiting persistent neuronal activity during covert spatial attention. *Front. Neural Circuits* 15:679796. doi: 10.3389/fncir.2021.679796

Amengual, J. L., Di Bello, F., Ben Hadj Hassen, S., and Ben Hamed, S. (2022). Distractibility and impulsivity neural states are distinct from selective attention and modulate the implementation of spatial attention. *Nat. Commun.* 13:4796. doi: 10.1038/s41467-022-32385-y

Andersson, P., Ramsey, N. F., Raemaekers, M., Viergever, M. A., and Pluim, J. P. W. (2012). Real-time decoding of the direction of covert visuospatial attention. *J. Neural Eng.* 9:045004. doi: 10.1088/1741-2560/9/4/045004

Arvaneh, M., Robertson, I. H., and Ward, T. E. (2019). A P300-based brain-computer interface for improving attention. *Front. Hum. Neurosci.* 12:524. doi: 10.3389/fnhum.2018.00524

Astrand, E., Enel, P., Ibos, G., Dominey, P. F., Baraduc, P., and Ben Hamed, S. (2014a). Comparison of classifiers for decoding sensory and cognitive information from prefrontal neuronal populations. *PLoS One* 9:e86314. doi: 10.1371/journal.pone.0086314

Astrand, E., Wardak, C., and Ben Hamed, S. (2014b). Selective visual attention to drive cognitive brain machine interfaces: From concepts to neurofeedback and rehabilitation applications. *Front. Syst. Neurosci.* 8:144. doi: 10.3389/fnsys.2014.00144

Astrand, E., Ibos, G., Duhamel, J.-R., and Ben Hamed, S. (2015). Differential dynamics of spatial attention, position, and color coding within the parietofrontal network. *J. Neurosci.* 35, 3174–3189. doi: 10.1523/JNEUROSCI.2370-14.2015

Astrand, E., Wardak, C., Baraduc, P., and Ben Hamed, S. (2016). Direct two-dimensional access to the spatial location of covert attention in macaque prefrontal cortex. *Curr. Biol.* 26, 1699–1704. doi: 10.1016/j.cub.2016.04.054

Astrand, E., Wardak, C., and Ben Hamed, S. (2020). Neuronal population correlates of target selection and distractor filtering. *Neuroimage* 209:116517. doi: 10.1016/j.neuroimage.2020.116517

Bae, G.-Y. (2020). The time course of face representations during perception and working memory maintenance. *Cereb. Cortex Commun.* 2:tgaa093. doi: 10.1093/texcom/tgaa093

Bae, G.-Y., and Luck, S. J. (2018). Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials. *J. Neurosci.* 38, 409–422. doi: 10.1523/JNEUROSCI.2860-17.2017

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609. doi: 10.1162/089892903321662976

Battistoni, E., Kaiser, D., Hickey, C., and Peelen, M. V. (2018). Spatial attention follows category-based attention during naturalistic visual search: Evidence from MEG decoding. *bioRxiv [Preprint]* doi: 10.1101/390807

Bensmaia, S. J., and Miller, L. E. (2014). Restoring sensorimotor function through intracortical interfaces: Progress and looming challenges. *Nat. Rev. Neurosci.* 15, 313–325. doi: 10.1038/nrn3724

Biasiucci, A., Leeb, R., Iturrate, I., Perdikis, S., Al-Khodairy, A., Corbet, T., et al. (2018). Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke. *Nat. Commun.* 9:2421. doi: 10.1038/s41467-018-04673-z

Bisley, J. W. (2011). The neural basis of visual attention. *J. Physiol.* 589, 49–57. doi: 10.1113/jphysiol.2010.192666

Blum, A., and Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artif. Intell.* 97, 245–271. doi: 10.1016/S0004-3702(97)00063-5

Bockbrader, M. A., Francisco, G., Lee, R., Olson, J., Solinsky, R., and Boninger, M. L. (2018). Brain computer interfaces in rehabilitation medicine. *PM R* 10, S233–S243. doi: 10.1016/j.pmrj.2018.05.028

Bogadhi, A. R., Bollimunta, A., Leopold, D. A., and Krauzlis, R. J. (2018). Brain regions modulated during covert visual attention in the macaque. *Sci. Rep.* 8:15237. doi: 10.1038/s41598-018-33567-9

Bouton, C. E., Shaikhouni, A., Annetta, N. V., Bockbrader, M. A., Friedenberg, D. A., Nielson, D. M., et al. (2016). Restoring cortical control of functional movement in a human with quadriplegia. *Nature* 533, 247–250. doi: 10.1038/nature17435

Branco, M. P., Freudenburg, Z. V., Aarnoutse, E. J., Bleichner, M. G., Vansteensel, M. J., and Ramsey, N. F. (2017). Decoding hand gestures from primary somatosensory cortex using high-density ECoG. *Neuroimage* 147, 130–142. doi: 10.1016/j.neuroimage.2016.12.004

Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., and Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J. Neurosci.* 18, 7411–7425. doi: 10.1523/JNEUROSCI.18-18-07411.1998

Brown, G. D., Yamada, S., and Sejnowski, T. J. (2001). Independent component analysis at the neural cocktail party. *Trends Neurosci.* 24, 54–63. doi: 10.1016/S0166-2236(00)01683-0

Buonomano, D. V., and Maass, W. (2009). State-dependent computations: Spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* 10, 113–125. doi: 10.1038/nrn2558

Buschman, T. J., and Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315, 1860–1862. doi: 10.1126/science.1138071

Chaudhary, U., Birbaumer, N., and Ramos-Murguialday, A. (2016). Brain–computer interfaces for communication and rehabilitation. *Nat. Rev. Neurol.* 12, 513–525. doi: 10.1038/nrneurol.2016.113

Chen, Y., Gong, C., Hao, H., Guo, Y., Xu, S., Zhang, Y., et al. (2019). Automatic sleep stage classification based on subthalamic local field potentials. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 118–128. doi: 10.1109/TNSRE.2018.2890272

Chen, Y., Namburi, P., Elliott, L. T., Heinzle, J., Soon, C. S., Chee, M. W. L., et al. (2011). Cortical surface-based searchlight decoding. *Neuroimage* 56, 582–592. doi: 10.1016/j.neuroimage.2010.07.035

Contini, E. W., Wardle, S. G., and Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia* 105, 165–176. doi: 10.1016/j.neuropsychologia.2017.02.013

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Cortese, A., Amano, K., Koizumi, A., Kawato, M., and Lau, H. (2016). Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. *Nat. Commun.* 7:13669. doi: 10.1038/ncomms13669

Cox, D. D., and Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* 19, 261–270. doi: 10.1016/S1053-8119(03)00049-1

Cunningham, J. P., and Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* 17, 1500–1509. doi: 10.1038/nn.3776

Davidson, T. J., Kloosterman, F., and Wilson, M. A. (2009). Hippocampal replay of extended experience. *Neuron* 63, 497–507. doi: 10.1016/j.neuron.2009.07.027

De Sousa, C., Gaillard, C., Di Bello, F., Ben Hadj Hassen, S., and Ben Hamed, S. (2021). Behavioral validation of novel high resolution attention decoding method from multi-units & local field potentials. *Neuroimage* 231:117853. doi: 10.1016/j.neuroimage.2021.117853

deBettencourt, M. T., Cohen, J. D., Lee, R. F., Norman, K. A., and Turk-Browne, N. B. (2015). Closed-loop training of attention with real-time brain imaging. *Nat. Neurosci.* 18, 470–475. doi: 10.1038/nn.3940

Degenhart, A. D., Bishop, W. E., Oby, E. R., Tyler-Kabara, E. C., Chase, S. M., Batista, A. P., et al. (2020). Stabilization of a brain–computer interface *via* the alignment of low-dimensional spaces of neural activity. *Nat. Biomed. Eng.* 4, 672–685. doi: 10.1038/s41551-020-0542-9

Desantis, A., Chan-Hon-Tong, A., Collins, T., Hogendoorn, H., and Cavanagh, P. (2020). Decoding the temporal dynamics of covert spatial attention using multivariate EEG analysis: Contributions of raw amplitude and alpha power. *Front. Hum. Neurosci.* 14:570419. doi: 10.3389/fnhum.2020.570419

Dietterich, T. G., and Bakiri, G. (1994). Solving multiclass learning problems *via* error-correcting output codes. *J. Artificial Intell. Res.* 2, 263–286.

Dijkstra, N., Bosch, S. E., and van Gerven, M. A. J. (2019). Shared neural mechanisms of visual perception and imagery. *Trends Cogn. Sci.* 23, 423–434. doi: 10.1016/j.tics.2019.02.004

Dugué, L., McLelland, D., Lajous, M., and VanRullen, R. (2015). Attention searches nonuniformly in space and in time. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15214–15219. doi: 10.1073/pnas.1511331112

Dugué, L., Roberts, M., and Carrasco, M. (2016). Attention reorients periodically. *Curr. Biol.* 26, 1595–1601. doi: 10.1016/j.cub.2016.04.046

Elsayed, G. F., and Cunningham, J. P. (2017). Structure in neural population recordings: An expected byproduct of simpler phenomena? *Nat. Neurosci.* 20, 1310–1318. doi: 10.1038/nn.4617

Esghaei, M., and Daliri, M. R. (2014). Decoding of visual attention from LFP signals of macaque MT. *PLoS One* 9:e100381. doi: 10.1371/journal.pone.0100381

Enander, J. M. D., Spanne, A., Mazzoni, A., Bengtsson, F., Oddo, C. M., and Jörntell, H. (2019). Ubiquitous neocortical decoding of tactile input patterns. *Front. Cell. Neurosci.* 13:140. doi: 10.3389/fncel.2019.00140

Fiebelkorn, I. C., Pinsk, M. A., and Kastner, S. (2018). A dynamic interplay within the frontoparietal network underlies rhythmic spatial attention. *Neuron* 99, 842–853.e8. doi: 10.1016/j.neuron.2018.07.038

Fiebelkorn, I. C., Saalmann, Y. B., and Kastner, S. (2013). Rhythmic sampling within and between objects despite sustained attention at a cued location. *Curr. Biol.* 23, 2553–2558. doi: 10.1016/j.cub.2013.10.063

Frostig, R. D., Xiong, Y., Chen-Bee, C. H., Kvašňák, E., and Stehberg, J. (2008). Large-scale organization of rat sensorimotor cortex based on a motif of large activation spreads. *J. Neurosci.* 28, 13274–13284. doi: 10.1523/JNEUROSCI.4074-08.2008

Gaillard, C., Ben Hadj Hassen, S., Di Bello, F., Bihan-Poudec, Y., VanRullen, R., and Ben Hamed, S. (2020). Prefrontal attentional saccades explore space rhythmically. *Nat. Commun.* 11:925. doi: 10.1038/s41467-020-14649-7

Gaillard, C., and Ben Hamed, S. (2020). The neural bases of spatial attention and perceptual rhythms. *Eur. J. Neurosci.* 55, 3209–3223. doi: 10.1111/ejn.15044

Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A., and Miller, L. E. (2020). Long-term stability of cortical population dynamics underlying consistent behavior. *Nat. Neurosci.* 23, 260–270. doi: 10.1038/s41593-019-0555-4

Gallego, J. A., Perich, M. G., Miller, L. E., and Solla, S. A. (2017). Neural manifolds for the control of movement. *Neuron* 94, 978–984. doi: 10.1016/j.neuron.2017.05.025

Gallego, J. A., Perich, M. G., Naufel, S. N., Ethier, C., Solla, S. A., and Miller, L. E. (2018). Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nat. Commun.* 9:4233. doi: 10.1038/s41467-018-06560-z

Ghaziri, J., Tucholka, A., Larue, V., Blanchette-Sylvestre, M., Reyburn, G., Gilbert, G., et al. (2013). Neurofeedback training induces changes in white and gray matter. *Clin. EEG Neurosci.* 44, 265–272. doi: 10.1177/1550059413476031

Glaser, J. I., Benjamin, A. S., Chowdhury, R. H., Perich, M. G., Miller, L. E., and Kording, K. P. (2020). Machine learning for neural decoding. *eNeuro* 7:ENEURO.0506-19.2020. doi: 10.1523/ENEURO.0506-19.2020

Golub, M. D., Sadtler, P. T., Oby, E. R., Quick, K. M., Ryu, S. I., Tyler-Kabara, E. C., et al. (2018). Learning by neural reassociation. *Nat. Neurosci.* 21, 607–616. doi: 10.1038/s41593-018-0095-3

Golub, M. D., Yu, B. M., Schwartz, A. B., and Chase, S. M. (2014). Motor cortical control of movement speed with implications for brain-machine interface control. *J. Neurophysiol.* 112, 411–429. doi: 10.1152/jn.00391.2013

Guy, V., Soriani, M.-H., Bruno, M., Papadopoulo, T., Desnuelle, C., and Clerc, M. (2018). Brain computer interface with the P300 speller: Usability for disabled people with amyotrophic lateral sclerosis. *Ann. Phys. Rehabil. Med.* 61, 5–11. doi: 10.1016/j.rehab.2017.09.004

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., et al. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87, 96–110. doi: 10.1016/j.neuroimage.2013.10.067

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430. doi: 10.1126/science.1063736

Hebart, M. N., and Baker, C. I. (2018). Deconstructing multivariate decoding for the study of brain function. *Neuroimage* 180, 4–18. doi: 10.1016/j.neuroimage.2017.08.005

Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., et al. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* 485, 372–375. doi: 10.1038/nature11076

Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., et al. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* 442, 164–171. doi: 10.1038/nature04970

Huang, W., Yan, H., Wang, C., Li, J., Zuo, Z., Zhang, J., et al. (2020). Perception-to-image: Reconstructing natural images from the brain activity of visual perception. *Ann. Biomed. Eng.* 48, 2323–2332. doi: 10.1007/s10439-020-02502-3

Iturrate, I., Chavarriaga, R., and Millán, J. D. R. (2020). General principles of machine learning for brain-computer interfacing. *Handb. Clin. Neurol.* 168, 311–328. doi: 10.1016/B978-0-444-63934-9.00023-8

Jafakesh, S., Jahromy, F. Z., and Daliri, M. R. (2016). Decoding of object categories from brain signals using cross frequency coupling methods. *Biomed. Signal Process. Control* 27, 60–67. doi: 10.1016/j.bspc.2016.01.013

Johnson, K. O. (2000). Neural coding. *Neuron* 26, 563–566. doi: 10.1016/s0896-6273(00)81193-9

Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8, 679–685. doi: 10.1038/nn1444

Kamitani, Y., and Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Curr. Biol.* 16, 1096–1102. doi: 10.1016/j.cub.2006.04.003

Kashefi, M., and Daliri, M. R. (2021). A stack LSTM structure for decoding continuous force from local field potential signal of primary motor cortex (M1). *BMC Bioinformatics* 22:26. doi: 10.1186/s12859-020-03953-0

Kia, S. M., Vega Pons, S., Weisz, N., and Passerini, A. (2017). Interpretability of multivariate brain maps in linear brain decoding: Definition, and heuristic quantification in multivariate analysis of MEG time-locked effects. *Front. Neurosci.* 10:619. doi: 10.3389/fnins.2016.00619

Kim, J., Bülthoff, I., Kim, S.-P., and Bülthoff, H. H. (2019). Shared neural representations of tactile roughness intensities by somatosensation and touch observation using an associative learning method. *Sci. Rep.* 9:77. doi: 10.1038/s41598-018-37378-w

King, J.-R., and Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends Cogn. Sci.* 18, 203–210. doi: 10.1016/j.tics.2014.01.002

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., et al. (2016). Demixed principal component analysis of neural population data. *eLife* 5:e10989. doi: 10.7554/eLife.10989

Kriegeskorte, N., and Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage* 38, 649–662. doi: 10.1016/j.neuroimage.2007.02.022

Kriegeskorte, N., and Douglas, P. K. (2019). Interpreting encoding and decoding models. *Curr. Opin. Neurobiol.* 55, 167–179. doi: 10.1016/j.conb.2019.04.002

Kriegeskorte, N., and Kreiman, G. (eds). (2011). *Visual population codes: Toward a common multivariate framework for cell recording and functional imaging, computational neuroscience series*. Cambridge, MA: MIT Press.

Kurth-Nelson, Z., Economides, M., Dolan, R. J., and Dayan, P. (2016). Fast sequences of non-spatial state representations in humans. *Neuron* 91, 194–204. doi: 10.1016/j.neuron.2016.05.028

Landau, A. N., and Fries, P. (2012). Attention samples stimuli rhythmically. *Curr. Biol.* 22, 1000–1004. doi: 10.1016/j.cub.2012.03.054

Lebedev, M. A., and Nicolelis, M. A. L. (2017). Brain-machine interfaces: From basic science to neuroprostheses and neurorehabilitation. *Physiol. Rev.* 97, 767–837. doi: 10.1152/physrev.00027.2016

Lemm, S., Blankertz, B., Dickhaus, T., and Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *Neuroimage* 56, 387–399. doi: 10.1016/j.neuroimage.2010.11.004

Liu, Y., and Ayaz, H. (2018). Speech recognition *via* fNIRS based brain signals. *Front. Neurosci.* 12:695. doi: 10.3389/fnins.2018.00695

Logothetis, N. K., and Sheinberg, D. L. (1996). Visual object recognition. *Annu. Rev. Neurosci.* 19, 577–621. doi: 10.1146/annurev.ne.19.030196.003045

Loriette, C., De Sousa Ferreira, C., Clavagnier, S., Lamberton, F., Ibarolla, D., and Ben Hamed, S. (2021). Non-invasive real-time access to spatial attention information from 3T fMRI BOLD signals. *bioRixv [Preprint]* doi: 10.1101/2021.11.24.469873

Loriette, C., and Ziane, C. (2021). Neurofeedback for cognitive enhancement and intervention and brain plasticity. *Rev. Neurol.* 177, 1133–1144.

Lübke, J., and Feldmeyer, D. (2007). Excitatory signal flow and connectivity in a cortical column: Focus on barrel cortex. *Brain Struct. Funct.* 212, 3–17. doi: 10.1007/s00429-007-0144-2

Marins, T., Rodrigues, E. C., Bortolini, T., Melo, B., Moll, J., and Tovar-Moll, F. (2019). Structural and functional connectivity changes in response to short-term neurofeedback training with motor imagery. *Neuroimage* 194, 283–290. doi: 10.1016/j.neuroimage.2019.03.027

McMains, S., and Kastner, S. (2011). Interactions of top-down and bottom-up mechanisms in human visual cortex. *J. Neurosci.* 31, 587–597. doi: 10.1523/JNEUROSCI.3766-10.2011

Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends Neurosci.* 6, 414–417. doi: 10.1016/0166-2236(83)90190-X

Nazari, M. R., Nasrabadi, A. M., and Daliri, M. R. (2021). Single-trial decoding of motion direction during visual attention from local field potential signals. *IEEE Access* 9, 66450–66461. doi: 10.1109/ACCESS.2021.3076865

Nicolas-Alonso, L. F., and Gomez-Gil, J. (2012). Brain computer interfaces, a review. *Sensors* 12, 1211–1279. doi: 10.3390/s120201211

Oby, E. R., Golub, M. D., Hennig, J. A., Degenhart, A. D., Tyler-Kabara, E. C., Yu, B. M., et al. (2019). New neural activity patterns emerge with long-term learning. *Proc. Natl. Acad. Sci. U.S.A.* 116, 15210–15215. doi: 10.1073/pnas.1820296116

O'Doherty, J. E., Lebedev, M. A., Ifft, P. J., Zhuang, K. Z., Shokur, S., Bleuler, H., et al. (2011). Active tactile exploration enabled by a brain-machine-brain interface. *Nature* 479, 228–231. doi: 10.1038/nature10489

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355

Padmanaban, S., Baker, J., and Greger, B. (2018). Feature selection methods for robust decoding of finger movements in a non-human primate. *Front. Neurosci.* 12:22. doi: 10.3389/fnins.2018.00022

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.

Perel, S., Sadtler, P. T., Godlove, J. M., Ryu, S. I., Wang, W., Batista, A. P., et al. (2013). Direction and speed tuning of motor-cortex multi-unit activity and local field potentials during reaching movements. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2013, 299–302. doi: 10.1109/EMBC.2013.6609496

Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., and Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.* 5, 805–811. doi: 10.1038/nn890

Pohlmeyer, E. A., Oby, E. R., Perreault, E. J., Solla, S. A., Kilgore, K. L., Kirsch, R. F., et al. (2009). Toward the restoration of hand use to a paralyzed monkey: Brain-controlled functional electrical stimulation of forearm muscles. *PLoS One* 4:e5924. doi: 10.1371/journal.pone.0005924

Qiao, K., Chen, J., Wang, L., Zhang, C., Zeng, L., Tong, L., et al. (2019). Category decoding of visual stimuli from human brain activity using a bidirectional recurrent neural network to simulate bidirectional information flows in human visual cortices. *Front. Neurosci.* 13:692. doi: 10.3389/fnins.2019.00692

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., et al. (2019). A deep learning framework for neuroscience. *Nat. Neurosci.* 22, 1761–1770. doi: 10.1038/s41593-019-0520-2

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., et al. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585–590. doi: 10.1038/nature12160

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., et al. (2016). Discrete neural signatures of basic emotions. *Cereb. Cortex* 26, 2563–2573. doi: 10.1093/cercor/bhv086

Sadtler, P. T., Quick, K. M., Golub, M. D., Chase, S. M., Ryu, S. I., Tyler-Kabara, E. C., et al. (2014). Neural constraints on learning. *Nature* 512, 423–426. doi: 10.1038/nature13665

Savage, N. (2019). How AI and neuroscience drive each other forwards. *Nature* 571, S15–S17. doi: 10.1038/d41586-019-02212-4

Schölkopf, B., Platt, J., and Hofmann, T. (2007). *Advances in neural information processing systems 19: Proceedings of the 2006 conference*. Cambridge, MA: MIT Press.

Seif, Z., and Daliri, M. R. (2015). Evaluation of local field potential signals in decoding of visual attention. *Cogn. Neurodyn.* 9, 509–522. doi: 10.1007/s11571-015-9336-2

Shen, G., Dwivedi, K., Majima, K., Horikawa, T., and Kamitani, Y. (2019). End-to-end deep image reconstruction from human brain activity. *Front. Comput. Neurosci.* 13:21. doi: 10.3389/fncom.2019.00021

Shulman, G. L., Ollinger, J. M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Petersen, S. E., et al. (1999). Areas involved in encoding and applying directional expectations to moving objects. *J. Neurosci.* 19, 9480–9496. doi: 10.1523/JNEUROSCI.19-21-09480.1999

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., et al. (2017). Closed-loop brain training: The science of neurofeedback. *Nat. Rev. Neurosci.* 18, 86–100. doi: 10.1038/nrn.2016.164

Stokes, M., Thompson, R., Cusack, R., and Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *J. Neurosci.* 29, 1565–1572. doi: 10.1523/JNEUROSCI.4657-08.2009

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78, 364–375. doi: 10.1016/j.neuron.2013.01.039

Taghizadeh-Sarabi, M., Daliri, M. R., and Niksirat, K. S. (2015). Decoding objects of basic categories from electroencephalographic signals using wavelet transform and support vector machines. *Brain Topogr.* 28, 33–46. doi: 10.1007/s10548-014-0371-9

Tam, W.-K., Wu, T., Zhao, Q., Keefer, E., and Yang, Z. (2019). Human motor decoding from neural signals: A review. *BMC Biomed. Eng.* 1:22. doi: 10.1186/s42490-019-0022-z

Thiery, T., Lajnef, T., Jerbi, K., Arguin, M., Aubin, M., and Jolicoeur, P. (2016). Decoding the locus of covert visuospatial attention from EEG signals. *PLoS One* 11:e0160304. doi: 10.1371/journal.pone.0160304

Trachel, R. E., Clerc, M., and Brochier, T. G. (2015). Decoding covert shifts of attention induced by ambiguous visuospatial cues. *Front. Hum. Neurosci.* 9:358. doi: 10.3389/fnhum.2015.00358

Treder, M. S., Bahramisharif, A., Schmidt, N. M., van Gerven, M. A. J., and Blankertz, B. (2011). Brain-computer interfacing using modulations of alpha activity induced by covert shifts of attention. *J. Neuroeng. Rehabil.* 8:24. doi: 10.1186/1743-0003-8-24

VanRullen, R. (2016). Perceptual cycles. *Trends Cogn. Sci.* 20, 723–735.

VanRullen, R. (2018). Attention cycles. *Neuron* 99, 632–634. doi: 10.1016/j.neuron.2018.08.006

VanRullen, R., Carlson, T., and Cavanagh, P. (2007). The blinking spotlight of attention. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19204–19209. doi: 10.1073/pnas.0707316104

Velliste, M., Perel, S., Spalding, M., Whitford, A. S., and Schwartz, A. (2008). Cortical control of a prosthetic arm for self-feeding. *Nature* 453, 1098–1101. doi: 10.1038/nature06996

Velu, P. D., and de Sa, V. R. (2013). Single-trial classification of gait and point movement preparation from human EEG. *Front. Neurosci.* 7:84. doi: 10.3389/fnins.2013.00084

Wang, T., Mantini, D., and Gillebert, C. R. (2018). The potential of real-time fMRI neurofeedback for stroke rehabilitation: A systematic review. *Cortex* 107, 148–165. doi: 10.1016/j.cortex.2017.09.006

Wen, H., Shi, J., Zhang, Y., Lu, K.-H., Cao, J., and Liu, Z. (2018). Neural encoding and decoding with deep learning for dynamic natural vision. *Cereb. Cortex* 1991, 4136–4160. doi: 10.1093/cercor/bhx268

Westendorff, S., Kaping, D., Everling, S., and Womelsdorf, T. (2016). Prefrontal and anterior cingulate cortex neurons encode attentional targets even when they do not apparently bias behavior. *J. Neurophysiol.* 116, 796–811. doi: 10.1152/jn.00027.2016

Wilcox, T., and Biondi, M. (2015). fNIRS in the developmental sciences. *Wiley Interdiscip. Rev. Cogn. Sci.* 6, 263–283. doi: 10.1002/wcs.1343

Wolff, M., Ding, J., Myers, N., and Stokes, M. (2015). Revealing hidden states in visual working memory using electroencephalography. *Front. Syst. Neurosci.* 9:123. doi: 10.3389/fnsys.2015.00123

Wolff, M. J., Jochim, J., Akyürek, E. G., and Stokes, M. G. (2017). Dynamic hidden states underlying working memory guided behaviour. *Nat. Neurosci.* 20, 864–871. doi: 10.1038/nn.4546

Zafar, R., Malik, A. S., Kamel, N., Dass, S. C., Abdullah, J. M., Reza, F., et al. (2015). Decoding of visual information from human brain activity: A review of fMRI and EEG studies. *J. Integr. Neurosci.* 14, 155–168. doi: 10.1142/S0219635215500089

Zenon, A., Ben Hamed, S., Duhamel, J.-R., and Olivier, E. (2008). Spatial and temporal dynamics of attentional guidance during inefficient visual search. *PLoS One* 3:e2219. doi: 10.1371/journal.pone.0002219

Zénon, A., Ben Hamed, S., Duhamel, J.-R., and Olivier, E. (2009). Attentional guidance relies on a winner-take-all mechanism. *Vision Res.* 49, 1522–1531. doi: 10.1016/j.visres.2009.03.010

Zhang, K., Ginzburg, I., McNaughton, B. L., and Sejnowski, T. J. (1998). Interpreting neuronal population activity by reconstruction: Unified framework with application to hippocampal place cells. *J. Neurophysiol.* 79, 1017–1044. doi: 10.1152/jn.1998.79.2.1017

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 08 September 2022
DOI 10.3389/fpsyg.2022.881269

OPEN ACCESS

EDITED BY
Hong Xu,
Nanyang Technological University,
Singapore

REVIEWED BY
Benjamin Taylor Files,
DEVCOM Army Research Laboratory,
Human Research and Engineering
Directorate, United States
Stefan Pollmann,
Otto von Guericke University
Magdeburg, Germany

*CORRESPONDENCE
Olga Lukashova-Sanz
olga.lukashova@uni-tuebingen.com

SPECIALTY SECTION
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

RECEIVED 22 February 2022
ACCEPTED 19 August 2022
PUBLISHED 08 September 2022

CITATION
Lukashova-Sanz O, Agarwala R and
Wahl S (2022) Context matters during
pick-and-place in VR: Impact on
search and transport phases.
Front. Psychol. 13:881269.
doi: 10.3389/fpsyg.2022.881269

Context matters during pick-and-place in VR: Impact on search and transport phases

Olga Lukashova-Sanz[1,2]*, Rajat Agarwala[1] and Siegfried Wahl[1,2]

[1]Zeiss Vision Science Lab, Institute for Ophthalmic Research, University of Tübingen, Tübingen, Germany, [2]Carl Zeiss Vision International Gesellschaft mit beschränkter Haftung (GmbH), Aalen, Germany

When considering external assistive systems for people with motor impairments, gaze has been shown to be a powerful tool as it is anticipatory to motor actions and is promising for understanding intentions of an individual even before the action. Up until now, the vast majority of studies investigating the coordinated eye and hand movement in a grasping task focused on single objects manipulation without placing them in a meaningful scene. Very little is known about the impact of the scene context on how we manipulate objects in an interactive task. In the present study, it was investigated how the scene context affects human object manipulation in a pick-and-place task in a realistic scenario implemented in VR. During the experiment, participants were instructed to find the target object in a room, pick it up, and transport it to a predefined final location. Thereafter, the impact of the scene context on different stages of the task was examined using head and hand movement, as well as eye tracking. As the main result, the scene context had a significant effect on the search and transport phases, but not on the reach phase of the task. The present work provides insights into the development of potential supporting intention predicting systems, revealing the dynamics of the pick-and-place task behavior once it is realized in a realistic context-rich scenario.

KEYWORDS

pick-and-place task, scene context, object manipulation, virtual reality, realistic scene

1. Introduction

Over the last several decades, the development of external assistive systems, such as prosthetic arms and exoskeletons, received much attention due to their strong potential to complement and improve the lives of people with sensory-motor impairments (Lazarou et al., 2018). One of the grand challenges for these assistive means is non-intuitive complicated mutual communication between the device and the patient, and thus, their inability to adapt to the individual patient's needs (Lazarou et al., 2018; Sensinger and Dosen, 2020). The evolution of the eye-tracking technology opened a possibility to ensure more intuitive communication between the assistive system and the user (Shafti et al., 2019; de Brouwer et al., 2021; Subramanian et al., 2021). Developing an intuitive algorithm that uses eye movements to support the user in object manipulation requires the knowledge of how our eyes and hand move when grasping an object. Several studies investigated the dynamics of object manipulation. They described the coordinated eye and hand movements in a grasping task for intact people as well as people with motor disabilities (Ballard et al., 1992; Johansson et al., 2001; Lavoie et al., 2018; Gregori et al., 2019; de Brouwer et al., 2021). Even so, the vast majority of the research on eye-hand interaction in a grasping task mainly focused on single objects manipulation without placing them in a meaningful scene (e.g., Lavoie et al., 2018; Gregori et al., 2019). In real life, however, we perceive the world and interact with the surrounding objects in a context-rich environment rather than isolated objects (Võ and Wolfe, 2013; Võ, 2021). While the existing research provides valuable information about eye and hand movement when performing a grasp of an object and transporting it from one location to another, it is still unclear whether the eye and hand movement will remain in a more realistic scene context the same. It is, therefore, essential to investigate human object manipulation once the object is placed into a context-rich natural setting.

The importance of contextual information in various visual tasks has been shown in multiple studies (e.g., Chun and Jiang, 1998; Marek and Pollmann, 2020). Furthermore, the visual search research provides strong evidence that the scene context affects where and how often we gaze within a given scene (Brockmole et al., 2006; Torralba et al., 2006; Võ and Wolfe, 2013; Wolfe, 2020; Võ, 2021). As such, if we are to find a laptop in the room, we are likely to gaze mostly at the surfaces, particularly the working table (Pereira and Castelhano, 2019; Võ, 2021). Interestingly, this holds also when the search target is not present in the scene, underlining the predictive nature of human search strategy (Biederman et al., 1982; Bar, 2004, 2009; Võ et al., 2019). Scene context also has been shown to facilitate action recognition (Wurm and Schubotz, 2017). Moreover, it has been shown that when searching for an object, people tend to rely on relevant anchors in the scene, that is, larger characteristic objects which typically are associated with the target object location (Draschkow and Võ, 2017; Võ, 2021). Until now, most of the studies in the field have been implemented in a non-interactive manner, where the observer had to perform a visual search task without manipulating the target object. In real-world scenarios, however, we often do not just observe the environment but perceive the world in terms of affordances, i.e., opportunities to interact with the surrounding environment (Gibson, 2014). This interaction is tightly bound to grasping and manipulating objects around us. In a grasping task, it is yet unknown whether different phases of object manipulation, such as reaching the object or transporting it from one location to another, are affected by the scene context.

One recent study looked into the effect of the scene context consistency on interaction with objects, where participants had to construct environments from a set of virtual objects either in agreement with their semantic expectations or against them (Draschkow and Võ, 2017). Among other results, the authors showed an increased grasping time of the object when it did not match the scene context. A possible explanation for this effect is increased decision time on where to put the object when it doesn't fit the scene context. There are countless possible locations in contrast to limited locations when the object is congruent with the scene context. Furthermore, previous studies demonstrated that reaching an object before grasping it can be affected by various factors. As an example, motor inhibition when approaching dangerous objects has been shown due to the emergence of aversive affordances (Mustile et al., 2021). These studies provide evidence that the way we look at and grasp objects around us might differ depending on the meaning of that object and the semantic context they are placed.

In daily life, we often are confronted with a combination of a visual search task and subsequent grasping of an object, such as when we are looking for the keys before leaving the house. How is the behavior in such a scenario affected by the scene context? Would there be a difference when the object fits or doesn't fit the scene context or when there is no context at all? In the present study, we addressed object manipulation in a pick-and-place task when performing it in a context-rich environment and when no meaningful context is present. In particular, the present study is intended to investigate whether the scene context primarily impacts only the searching, as is suggested in the visual search literature where the object is searched longer if it is incongruent with the scene context compared to when it fits the scene. Alternatively, when the object does not match the scene context, does it take people longer to reach or transport it in addition to a more prolonged search? Moreover, consider comparing scenarios where the target object is placed in a semantically congruent context against the case when the object is isolated, i.e., in a context-poor environment. Would the presence of additional visual stimuli and possible obstacles in the context-rich environment serve as a distractor and lead to a more prolonged search and further interactions with the object even when the object matches the context, or would the scene context facilitate the search and object manipulation? To answer these questions and advance in understanding how humans manipulate objects in realistic scenes, it is vital to systematically address the effect of the scene context on different phases of our interaction with objects.

When studying interactive object manipulation, it is important to develop the paradigm realistically. Specifically, unconstrained head and hand movement and free eye movement are essential for natural behavior. The rapid development of modern technologies such as Virtual Reality (VR) and VR eye-tracking enabled researchers to study human interaction with the surrounding environment in more realistic 3D settings (Boettcher et al., 2018; Olk et al., 2018). Furthermore, VR provides a possibility to simulate various real-world scenarios in a yet controlled laboratory environment at an efficient cost. To explore the interactive domain of object manipulation, it is thus, convenient to develop the experimental paradigm in VR. When studying grasping using VR, it can be, however, challenging to reproduce natural grasping behavior in a virtual scene due to the mismatch of the visual response and lack of haptic feedback (Levin et al., 2015; Furmanek et al., 2019). In VR experiments, typically, the interaction with the virtual environment is realized *via* controllers instead of a real hand (e.g., Draschkow and Võ, 2017), where depending on the virtual hand representation, differences in hand movement between the real and virtual settings might emerge (Viau et al., 2004; Cai et al., 2021). Nonetheless, recent studies demonstrated that a virtual representation of a hand-looking object when interacting in VR could be a reasonable approach to the imitation of a realistic grasping as it enables the strongest sense of ownership (Lougiakis et al., 2020; Cai et al., 2021; Lavoie and Chapman, 2021). The realism can be enhanced when showing grasping an object by presenting a grasping hand pose holding the object (Tian et al., 2019; Lavoie and Chapman, 2021). Combining these findings with the advantages of VR for the experimental design mentioned above, in the present study, we chose VR as the tool to study object manipulation, where a virtual glove represented the hand, and respective grasping poses were generated.

The current study investigated the effect of the scene context on human object manipulation in an interactive task in a realistic scenario. Using the head and hand movement, as well as eye-tracking, the impact of the scene context on different stages of a pick-and-place task was examined. Specifically, pick-and-place task performance was evaluated while placing the objects of interest into typical everyday visual scenes implemented in VR.

2. Materials and methods

2.1. Participants

Thirteen naïve participants (5 female and 8 male), with normal or corrected to normal vision were tested. Participants were aged between 19 and 31 years old. No formal power analysis for the sample size calculation was performed. All procedures conformed to Standard 8 of the American Psychological Association's "Ethical Principles of Psychologists and Code of Conduct (2010)". The study was approved by the ethics committee of the Faculty of Medicine at the University of Tübingen with a corresponding ethical approval identification code 986/2020BO2. Signed informed consent was obtained from each participant before the measurements. All data were stored and analyzed in full compliance with the principles of the Data Protection Act GDPR 2016/679 of the European Union.

2.2. Experimental setup

2.2.1. Hardware specifications

The visual content was displayed to the participants using HTC Vive Pro Eye (HTC Corporation, Taoyuan, Taiwan) virtual reality headset running on a Windows 10 PC with NVIDIA GeForce GTX 1070 graphics card (NVIDIA Corporation, Santa Clara, California, USA). The field of view of the headset and the refresh rate reported by the manufacturer are 110° and 90 Hz, respectively. The participant interacted with the environment *via* the HTC Vive controller held in the right hand of the person. The position and rotation of the headset and the controller were tracked *via* four HTC base stations 2.0. The complete size of the tracked area was approximately three by three meters. The eye-tracking data were collected using the built-in eye tracker at a frequency of 120 Hz. During the experiment, the participant was in a standing position and could freely move within the working space. The experimental setup is schematically shown in Figure 1.

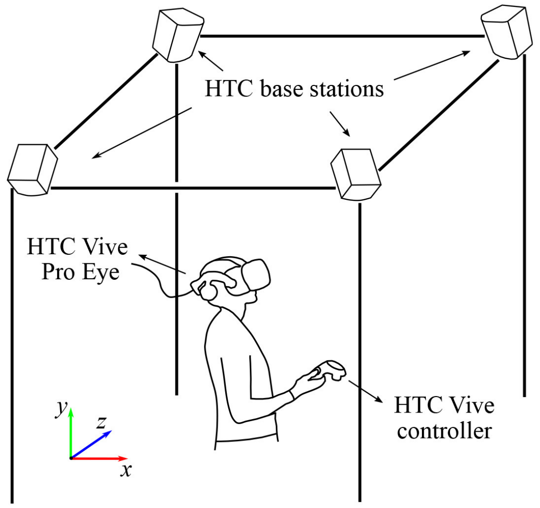


FIGURE 1
Schematic representation of the experimental setup. The visual scenes were displayed using HTC Vive Pro Eye virtual reality headset. The position and rotation of the headset and the controller were tracked *via* four HTC base stations 2.0 paced in the four corners of the working space. The interaction with the environment was realized using the HTC controller held in the right hand of the participant. The colored axes represent Unity left-hand coordinate system. For details, see the main text.

2.2.2. Software specifications

The experimental paradigm was generated using the Unity Game engine (Unity Technologies, 2019), Unity version 2019.4.0.f1. The eye movement data were collected using Unity package SRanipal version 1.3.3.0. Recording of the eye movement data at a maximum sampling rate 120 Hz was realized by using a separate thread parallel to the main script execution. The data analysis was performed using Python 3.6 packages NumPy (Van Der Walt et al., 2011) version 1.19.1, SciPy (Virtanen et al., 2020) version 1.5.2

and Pandas (McKinney, 2010) version 1.1.3. The statistical analysis was conducted using R version 3.6.1, in particular, package nlme (Pinheiro et al., 2022). The data visualization was performed using Python packages Matplotlib (Hunter, 2007) version 3.3.1 and Seaborn (Waskom et al., 2017) version 0.11.0.

2.3. Virtual environment and stimuli

2.3.1. Realistic man-made VR scenes

The virtual environment was composed of realistic man-made indoor scenes. Specifically, three different habitual scene contexts were selected: kitchen, bathroom, and office. These contexts are commonly met in daily life and were previously used in the scene guidance literature (e.g., Wurm and Schubotz, 2017; Boettcher et al., 2018; Beitner et al., 2021). Two different variations of each scene context were designed, resulting in six distinct context-rich scenes. The two variations of each context were introduced to maintain the variety of the environments and unique configuration of each trial. One requirement for the scenes was an equal set of anchors (see Section 2.3.3). That is, there was a computer in each office, a microwave in each kitchen, and a sink in each bathroom. Otherwise, the room design was arbitrary. The complete set of implemented scenes can be found in Supplementary material.

Furthermore, six empty virtual rooms were created with no scene context present: the rooms only contained a set of shelves replicating the spatial configurations of each of the context-rich environments, respectively. Doing so, the "empty" experimental condition was implemented (for details, see Section 2.4.3). The size of the scenes was set to two by three meters. An example of one scene and its empty match is shown in Figure 2. All context-rich scenes were created using a set of open-source 3D assets. The complete set of implemented "empty" scenes can also be found in Supplementary material.

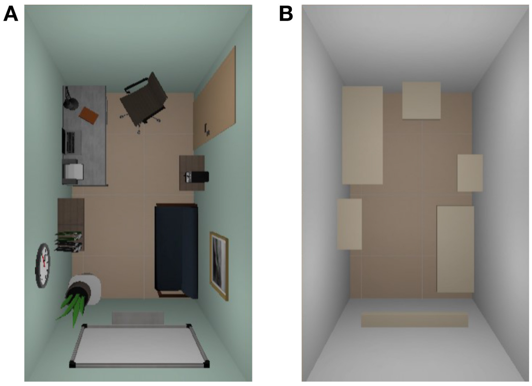


FIGURE 2
The top view of one of the designed VR scenes; **(A)** context-rich scene, here: office; **(B)** empty scene matching the spatial configuration of the respective context-rich scene. The size of the rooms was set to two by three meters. Note, the scenes are shown without the end location which was represented by a column 92 cm of height, and was always in the center of each room. The complete set of six implemented scenes and their respective empty equivalents can be found in Supplementary material.

2.3.2. Target objects and distractors

To maximally separate the scene context effect from such biases as the size or the shape of the object, and for comparability, we chose to use cubes with images projected on their faces instead of actual virtual objects. This way, it was intended that the participants focus more on the semantic meaning of the objects. Moreover, there is a practical advantage to using a cubic shape for a more accurate gaze evaluation. In particular, in Unity, the gaze point is detected on the object once the gaze ray hits the collider around that object (see more details in Section 2.5.3). Therefore, the collider should ideally have the same shape as the object to avoid a mismatch between detected and actual gaze points. Creating complex mesh colliders for different VR objects is a tedious task and slows down the display of the visual content in the VR headset. On the other hand, the cubic shape is one of the basic collider shapes in Unity and can be efficiently used.

A set of target objects was generated, where each object was represented by a cube with an image projection on its faces. The size of the cubes was set to 0.084 Unity-meters. For each scene context, a set of seven images was used, resulting in a total of 21 objects. All images were selected and adjusted from different open-source pictures on the Internet. The grayscale of the images was selected to avoid a pop-out effect due to the color-based saliency of some images compared to others. In Figure 3A, an example of the target object is shown. For the objects, we chose the cubic shape instead of arbitrary shapes of actual 3D objects to universalize the target objects across trials and make them comparable. Specifically, the homogeneous shape ensures no size or shape bias when estimating gaze locations on the target object. Furthermore, for the gaze evaluation, the cubic form allows a more accurate estimation of the gaze point on the object in Unity due to the simplicity of the box collider around the object.

Similarly, a set of distractors was generated for each scene context. The distractors were designed in the same manner as the target objects: grayscale images projected on the facets of cubes. The target-object-alike distractors were introduced to ensure the relative complexity of the task and prevent the participant from searching for the only cube existing in the scene. For each scene context, a set of 10 open-source images was used, resulting in a total of 10 distractors per scene. An example of a distractor is illustrated in Figure 3B.

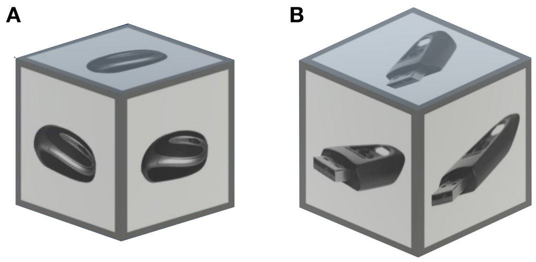


FIGURE 3
An example of **(A)** target object and **(B)** distractor used in the experiment. The objects were represented by grayscale images projected on the cube facets. The size of the cubes was set to 0.084 Unity-meters. Here, both the target object and the distractor, belong to the same scene context: office. For details, see Section 2.3.2.

FIGURE 4
An example of a VR scene from the participant's perspective. The cubic grayscale objects are the distractors and the target object. In the lower central part of the image, a part of the end location is shown, which was represented by a column in the center of the room. The controller was represented by a virtual SteamVR glove. Note, that a grasping pose for each object was designed and generated and shown to the participant upon grasping a virtual object. For more details, see the main text.

2.3.3. Anchors

For each of the context-rich scenes, a set of anchors was selected. Although there have been recent attempts to formally define anchors (Boettcher et al., 2018), up until now, no validated database was developed. Therefore, in the present study, the anchors were selected arbitrarily following an intuitive description: an anchor is considered to be a static larger object which is typically not easily moved with one hand (Võ, 2021). The complete set of the anchors and their respective target objects can be found in Supplementary material.

The predefined position for each target object was set to a location in the proximity of its anchor, where depending on the anchor, the object could appear either next to or on top of it. As such, a toothbrush would appear next to the sink in the bathroom, whereas a pan would be placed on top of the kitchen stove in the kitchen.

2.4. Experimental procedure

2.4.1. General procedure

In each trial, the participant's task was to find a specific object, pick it up using the controller, and transport the object to the final predefined location. The VR controller was represented by a SteamVR virtual glove. Note that a grasping pose for each object was designed and generated in advance and shown to the participant upon grasping a virtual object. In other words, the virtual hand did not disappear when the participant pressed the trigger to grasp the object but instead was still visible in a grasping pose (see Figure 4). The end location was represented by a column 92 cm of height and was always in the center of every room. The time of each trial was not limited, and participants were asked to perform the task at a normal pace. The participant was shown a gray background between each trial where the target object for the upcoming trial was displayed. The participant was requested to start each trial from a specific position in the room indicated by a round target which switched its color from blue to green once the participant was inside the target. To start the trial, the participant pressed a button on the controller. Thereafter, a small pause of 1 s was introduced where the participant was asked to gaze at a fixation target on the gray background before the virtual room appeared. Doing so ensured that all participants started the scene exploration initially, gazing in the same direction. Once the virtual room appeared, participants performed the task. After placing the target object in the final location, the trial was finished upon a button press, followed by the subsequent trial.

In Figure 4, an example of the scene view from the participant's perspective is shown.

2.4.2. Training session and main experiment

During the experimental session, participants performed the training session followed by the main experiment. A short 3-to-5-min break was introduced between the two sessions, during which participants removed the VR headset and rested. At the beginning of each session, the eye tracker calibration procedure was performed. First, participants completed a training session. During 32 trials, participants executed the task. During the training session, in contrast to the main experiment, the set of target objects was compiled by images of objects typically found outdoors (e.g., a traffic lights, a park bench). However, the virtual rooms were identical to those in the main experiment (kitchens, bathrooms, offices, and empty rooms with shelves). The goal of the training session was for the participants to get familiar with the environments and become acquainted with the dynamics of the trials, picking up and transporting virtual objects and switching between the trials. After a break, participants proceeded to the main experiment. The procedure was identical to the training phase, except the target objects were context-dependent.

Each trial was unique in terms of the combination of a specific room type and a target object. In doing so, we intended to prevent learning of specific spatial configurations. The total amount of trials was then composed of seven unique target objects for each of three scene contexts designed in two variations and used in three experimental conditions (see Section 2.4.3). This resulted in a total of 126 trials for each participant. The order of trials was randomized, where all trials, including the training session, were performed within one 1-h appointment.

2.4.3. Experimental conditions

To evaluate the impact of the scene context on the performance of the task, three experimental conditions were implemented. In the "congruent" condition, the target object matched the scene context (e.g., a toothbrush in a bathroom). In the "incongruent" condition, the object did not fit the scene context (e.g., a toothbrush in a kitchen). Specifically, the objects belonging to the remaining two contexts were randomly selected to form the incongruent condition. Finally, in the "empty" condition, no scene context was present.

Each room had a set of specific locations where the target object could appear. The possible locations were always on top of surfaces. The locations were determined by the proximity of each target object to its corresponding anchor (see Section 2.3.3). In the incongruent condition, the set of possible target locations was identical to that in the congruent condition, where throughout the trials each of the seven potential locations was occupied by one of the objects that did not belong to the context. Finally, in the empty condition, the configurations of the target objects, the distractors, and the virtual rooms were the same as those in the context-rich environments. This way, the possible spatial locations of the target objects were replicated in each of the three conditions, enabling the comparison across the conditions. In each trial, only one of the possible target objects was present.

Besides the target object, each room included a set of 10 distractors, which were located in specific positions in the context-rich and empty environments (Section 2.3.2).

2.5. Analysis

2.5.1. Eye movements data pre-processing

To evaluate the task performance, first, the eye movement data were analyzed and fixations were detected. The eye movement data were recorded at a frequency 120 Hz. The gaze position data was accessed using a customized written Unity script utilizing the HTC SRanipal SDK package functions. The eye data processing flow was adapted from our previous work (Lukashova-Sanz and Wahl, 2021). In Table 1 the main, recorded variables are described. All variables were recorded for left and right eyes.

TABLE 1 Main eye- and head-movement-related raw variables recorded during the experiment.

Variable	Units	Meaning
Time stamp	An integer number	The time in ms at the moment of sample recording.
Eye data validity bit mask	An integer from 0 to 31	Represents the validity of the data. A value of 31 indicates the highest validity of the recorded data. This parameter is used to filter the raw data where the eye tracker lost the pupil, including filtering blinks.
Gaze normalized direction vector	A three-coordinates vector *(x, y, z)* with each coordinate ranging from −1 to 1	A gaze vector indicating the direction of gaze in the headset right-hand coordinate system.
Head rotation	A rotation quaternion *(x, y, z, w)* of head	A quaternion describing the rotation of the headset in Unity world coordinates.

To prepare the data for further processing, first, similar to (Imaoka et al., 2020), the raw data were filtered based on the eye data validity bitmask value, which represents the bits containing all validity for the current frame. After the filtering, only the data where the eye data validity bit mask had value 31 for both eyes Table 1, were selected. Doing so, the data where the eye tracker partly or completely lost the pupil (including blinks) was filtered out. Next, for subsequent fixation detection (see Section 2.5.2), the gaze position was calculated in spherical coordinates. In particular, the polar ϕ and azimuthal θ angles were computed using Equations (1) and (2). In Unity, the z-axis corresponds to the depth dimension.

$$\phi = \arctan\frac{x}{z}, \tag{1}$$

$$\theta = \arctan2\,(y, \sqrt{x^2 + z^2}), \tag{2}$$

where (x, y, z) are coordinates of normalized gaze directional vector in headset coordinates. Note that SRanipal returns the gaze direction vector in the right-handed coordinate system. To convert the coordinates in Unity world coordinate system, which is a left-hand coordinate system, the x-coordinate was multiplied by -1. To compute the gaze position in Unity world coordinate system, the gaze position in the headset coordinate system was multiplied by the head rotation quaternion.

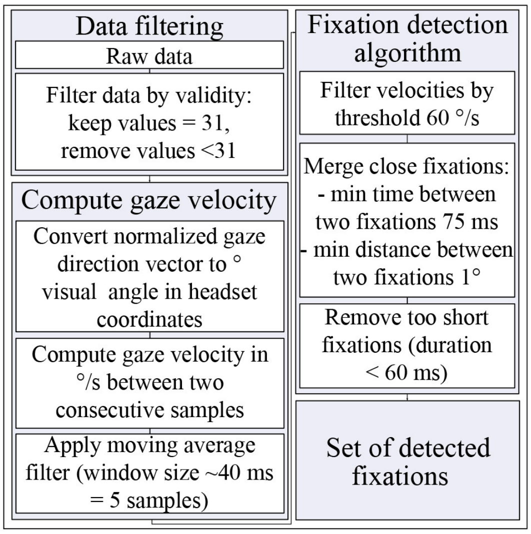


FIGURE 5
Eye movement data processing algorithm. For details, see Section 2.5.2.

2.5.2. Fixation detection algorithm — I-VT

Fixations were identified using *velocity threshold algorithm for fixation identification* (I-VT) (Salvucci and Goldberg, 2000). The algorithm was implemented following the description in (Kübler, 2020) and (Olsen, 2012). The gaze velocity v was computed in °/s between each two consecutive samples (Equation 3).

$$v = \frac{\sqrt{(\phi_i - \phi_{i-1})^2 + (\theta_i - \theta_{i-1})^2}}{t_i - t_{i-1}}, \tag{3}$$

where (ϕ_i, θ_i) and $(\phi_{i-1}, \theta_{i-1})$ are consecutive gaze positions in degrees of visual angle in headset coordinates, and t_i and t_{i-1} are respective time stamps. To reduce the noise level of the data, a running average filter was applied with the window size of five samples, which is ~ 40 ms. An eye movement was considered to be a fixation if the gaze velocity did not exceed a threshold of 60 °/s (Leube et al., 2017). Two fixations were merged in a single fixation if the time between them was under 75 ms (Komogortsev et al., 2010), and the angular distance was under 1° (Over et al., 2007; Komogortsev et al., 2010). Too short fixations with a duration under 60 ms were filtered out (Over et al., 2007; Komogortsev et al., 2010). In Figure 5 the eye movement data processing algorithm is summarized in a flow chart.

2.5.3. Determining gaze position on a virtual object

To determine the gaze position on the virtual object, a hit point of the gaze ray and the 3D object collider were recorded for each frame. Using this approach, Unity returns a set of three coordinates of a specific spatial point on the collider surface which was crossed by the gaze ray. Furthermore, the name of the hit collider was continuously recorded. Doing so, it was tracked which object was gazed at, in which frame, and for how long.

2.5.4. Task phases: search, reach, and transport

In each trial, the data was segmented into three phases: the search phase, the reach phase, and the transport phase.

The *search phase* is the period of time between the beginning of the trial and the first fixation on the target object. The first fixation on the object was defined as the first fixation during which the object collider was hit by the gaze ray. The *reach phase* was defined as the period of time between the first fixation on the target object and the moment of picking up the object, which is determined by the virtual hand attachment to the target object. Finally, the *transport phase* is the period of time starting from picking up the object until releasing it from the virtual hand when placing the object to the final location.

In Figure 6, an example of velocities for one participant in a single trial is shown. Different curves represent velocities of the head, the hand, the target object, and the end target location. The colored areas correspond to the search, reach, and transport phases.

2.5.5. Behavioral metrics

As mentioned in Section 2.2.2, the statistical analysis was conducted using R version 3.6.1, in particular, package nlme (Pinheiro et al., 2022). As described in Section 2.4.2, a diverse range of the target objects was used for the experiment, where each trial was unique in terms of the combination of the target object and the corresponding scene. The models did not consider the specific objects or the scenes as an additional factor due to very limited amount of trials for each particular object and scene. The temporal data was normalized using log function, whereas the proportional data was not normalized. All models were fitted using REML (reduced maximum likelihood) method. Each dependent variable was fitted with a separate model. No multiple comparison correction was performed. More details on specific linear mixed models for each variable can be found in Supplementary material.

Task duration

To evaluate the impact of the scene context on task performance, first, the *task duration* was evaluated. The effect of the scene context was estimated by fitting a linear mixed model to the data, where *task duration* is a dependent variable, *condition* is a fixed effect, and *participant* is a random factor.

Search, reach, and transport duration

Next, the duration of different task phases: search, reach, and transport, was examined and compared across different experimental conditions. The impact of the scene context was

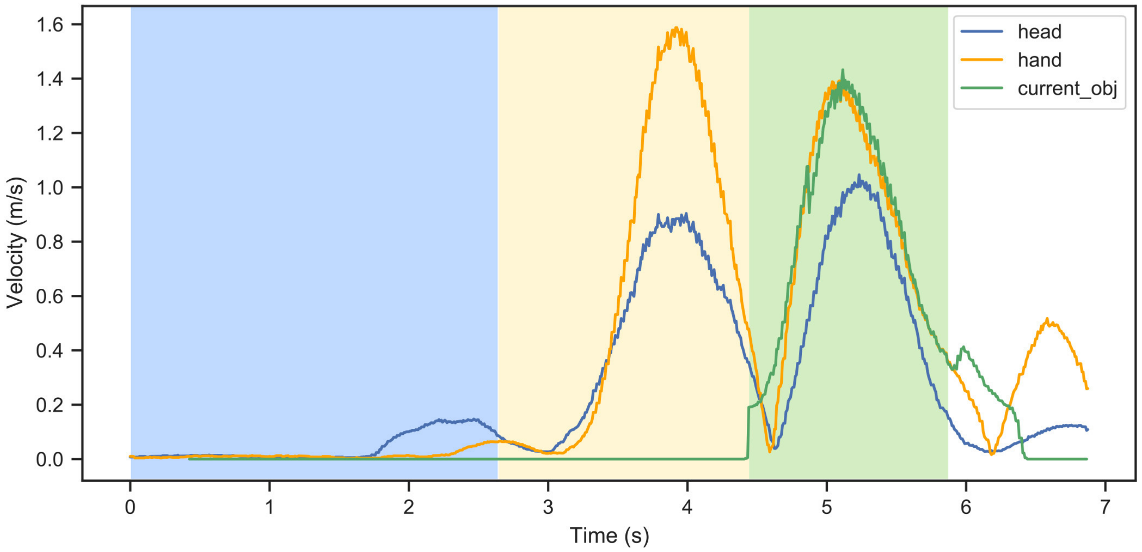


FIGURE 6
An example of velocities as a function of the trial duration for one participant in one trial: different curves represent the head, the hand, the target object, and the end target, respectively. The colored sectors correspond to the search (blue), reach (yellow), and transport (green) phases. In this example, the task was completed in approximately 7 s.

estimated by fitting linear mixed models to the data, fitting a separate model for each of the metrics, where *search duration*, *reach duration*, and *transport duration* are dependent variables, *condition* is a fixed effect, and *participant* is a random factor.

For the search phase, it is important to mention, that the search time is naturally expected to depend on the initial gaze-object distance at the beginning of the trial. In such, if the target object was originally behind the participant upon the trial start, it is likely to take longer to find it, compared to when the object was initially in front of the participant. Therefore, for the search duration, the *initial gaze-object distance* was set as an additional fixed factor.

Scene coverage

Another parameter that is expected to be affected by the scene context is *scene coverage* – a common metric indicating the proportion of an area covered during the trial. To compute the scene coverage in each trial, first, a 2D histogram of gaze points was plotted for the scene where the whole span of 360 °va and 180 °va in horizontal and vertical directions, respectively, was considered. The size of the histogram bins was set to 2 °va which approximately corresponds to the eye tracker accuracy. The histogram was then transformed into a binary image, with black pixels representing the area in which some gaze points fell. Finally, the scene coverage was computed as the proportion of the black pixels to the total amount of pixels of the scene. Similarly to the task and search duration, the impact of scene context was estimated by fitting the linear mixed models to the data, where *scene coverage* is the dependent variable, *condition* is a fixed effect, and *participant* is a random factor.

Proportion of gaze on target object and anchor

The *proportion of gaze on target object* was defined as the proportion of the number of frames gazing on the target object out of the total amount of recorded frames of the trial. In line with other metrics, this parameter indicated whether the scene context facilitated the task performance, which would be implicitly demonstrated by a larger proportion of gaze on the target object.

The *proportion of gaze on the anchor* was computed similarly as the proportion of the number of frames gazing on the anchor relative to the total number of frames in the trial. Note, that this metric was computed only for the context-rich conditions as in the empty condition no context and, therefore, no anchors were present. Furthermore, in the congruent condition, the anchor was relevant to the target object, whereas in the incongruent condition, even though in the same spatial configuration, semantically it was irrelevant. This metric enabled an implicit evaluation of the context facilitation, namely, a larger proportion of gaze on the anchor would demonstrate the importance of the relevant anchor for the task performance.

For both metrics, the impact of the scene context was evaluated by fitting the linear mixed models to the data, where a separate model was fitted to each metric. In the model, the *proportion of gaze on object and anchor* are dependent variables, *condition* is a fixed effect, and *participant* is a random factor. Note, that due to the realistic nature of the scenes, the anchors varied in size as well as in their relative position to the target object (see Section 4.4).

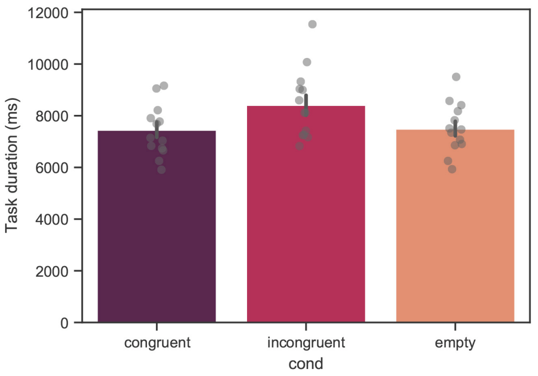


FIGURE 7
Mean task duration across all trials for all measured participants. Each bar corresponds to a separate experimental condition. The error bars indicate confidence intervals of 95% computed using bootstrapping (see main text). The individual points correspond to the mean value for each individual subject. The difference between the congruent and incongruent conditions was significant with $p < 0.001$. The difference between the congruent and empty conditions was not significant.

Anchor-object transition

Finally, the *anchor-object transition* was evaluated. The anchor-object transition is the time between the first fixation on the anchor and the first fixation on the target object. A positive value corresponds to when the anchor was fixated before the target object, whereas negative values show that the target object was fixated before the anchor. A shorter anchor-object transition would indicate the importance of the anchor. Note, that this metric was computed only for the context-rich conditions, as in the empty condition, no anchor was present. A linear mixed model was fitted to the data with *anchor-object transition* as dependent variable, *condition* as fixed effect, and *participant* as random factor.

3. Results

In this section, we report the mean values of each variable of interest together with their standard deviation. Using the nlme package (Pinheiro et al., 2022), the model output summary returns the fixed effects estimates, their approximate standard errors, the denominator degrees of freedom, the ratios between the estimates and their standard errors, and the associated p-values from a t-distribution. In the present work, the significance of the scene context effect on various dependent variables was evaluated based on the computed p-values returned in the output of the respective models. The output of each model can be found in the Supplementary material.

During a few trials, the connection between Unity and the headset was lost. After the data curing, for each participant, a total of maximum four trials was excluded from further analysis.

Most of the variables are visualized as bar plots with an overlay of individual subject values. The error bars correspond to the 95% confidence intervals. The default Seaborn (Waskom et al., 2017) setting was used to compute the confidence intervals, namely, through bootstrapping by sampling 1000 samples uniformly with replacement from the original data.

3.1. Task duration

In Figure 7, the mean task duration estimated across all participants is shown. The mean values of the task duration were 7433 ± 2820, 8395 ± 3822, and $7487 \pm 2767 ms$ for the congruent, incongruent, and empty conditions, respectively. From linear mixed model analysis, over the course of all trials, a significant effect of condition was found with $p < 0.001$. The difference between the congruent and empty conditions was not significant.

3.2. Search, reach, and transport duration

In Figure 8A, the mean search phase duration across all participants is shown. The mean values of the search duration were 4232 ± 2266, 5073 ± 3040, and $4499 \pm 2377 ms$ for the congruent, incongruent, and empty conditions, respectively. From linear mixed model analysis, over the course of all trials, a significant effect of condition was found with $p < 0.001$. Figure 8B demonstrates the search duration as a function of the initial gaze-object distance. As expected, there was found a significant effect of the initial distance, with $p < 0.001$ which is indicated by a positive slope of the linear fit for all the conditions. The difference between the congruent and empty conditions was not significant.

Figures 9A,B show the reach and transport duration across all trials for all measured participants. For the reach phase, no significant effect of the context condition was found. For the transport phase, the duration in the empty condition was significantly shorter than in the congruent condition with $p < 0.01$. The mean values for the reach phase duration were 1992 ± 1367, 2140 ± 1744, and $1895 \pm 1046 ms$ for the congruent, incongruent, and empty conditions, respectively. The mean values for the transport phase duration were 1753 ± 657, 1666 ± 1057, and $1578 \pm 475 ms$ for the congruent, incongruent, and empty conditions, respectively.

3.3. Scene coverage

In Figure 10, the mean scene coverage across all trials for all participants is shown. The mean values of the scene coverage were 0.022 ± 0.010, 0.026 ± 0.014, and 0.023 ± 0.012 for the congruent, incongruent, and empty conditions, respectively.

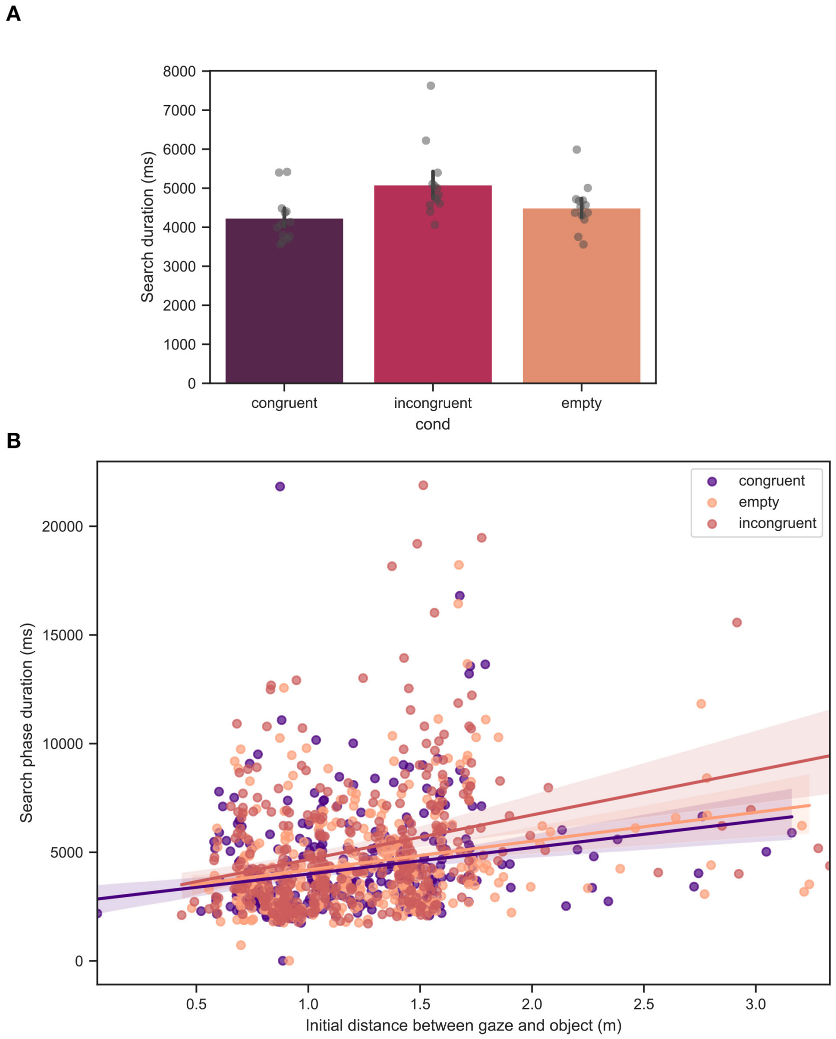


FIGURE 8
(A) Mean search duration across all trials for all measured participants. Each bar corresponds to a separate experimental condition. The error bars indicate confidence intervals of 95% computed using bootstrapping (see main text). The individual points correspond to the mean value for each individual participant. The difference between the congruent and incongruent conditions was significant with $p < 0.001$. The difference between the congruent and empty conditions was not significant. **(B)** Mean search duration as a function of the initial gaze-object distance, defined as the distance between the looked at point and the target fixation at the beginning of the trial. The straight lines are linear fits to the data. The effect of the initial gaze-object distance was significant with $p < 0.001$.

From linear mixed model analysis, over the course of all trials, a significant effect of condition was found with $p < 0.001$. The difference between the congruent and empty conditions was not significant.

3.4. Proportion of gaze on target object and anchor

In Figure 11, the proportion of gazing on the target object and the relevant anchor averaged over all participants is shown. Note, that Figure 11B demonstrates data only for the two context-rich conditions as in the empty condition no context and, thus, no anchor was present. The mean values of the proportion of gazing on the target object were 0.31±0.11, 0.29±0.11, and 0.30±0.10 for the congruent, incongruent, and empty conditions, respectively. The mean values of the proportion of gazing on the relevant anchor were 0.10±0.07, and 0.09±0.06 for the congruent, and incongruent conditions, respectively.

From linear mixed model analysis, over the course of all trials, a significant effect of condition for the proportion of gazing on the target object with a significant difference between

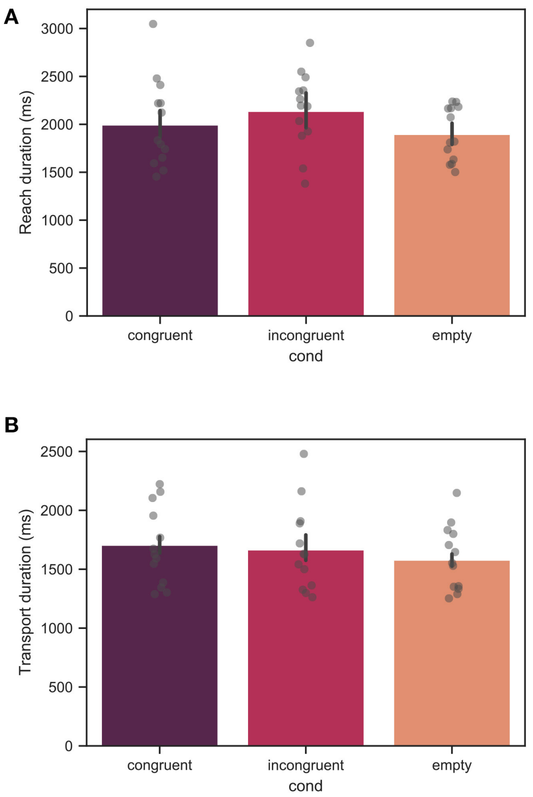


FIGURE 9
Mean duration of the **(A)** reach phase, and **(B)** transport phase across all trials for all measured participants. Each bar corresponds to a separate experimental condition. The error bars indicate confidence intervals of 95% computed using bootstrapping (see main text). The individual points correspond to the mean value for each individual participant. For the reach phase, the effect of condition was not significant. For the transport phase, the difference between the congruent and empty conditions was significant with $p < 0.01$, whereas no significant difference was found between the congruent and incongruent conditions.

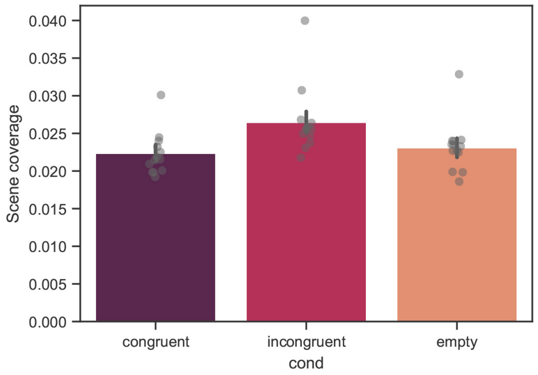


FIGURE 10
Mean scene coverage across all trials for all measured participants. The variable is computed as a proportion, thus, ranges between 0 and 1. Each bar corresponds to a separate experimental condition. The error bars indicate confidence intervals of 95% computed using bootstrapping (see main text). The individual points correspond to the mean value for each individual participant. The difference between the congruent and incongruent conditions was significant with $p < 0.001$. The difference between the congruent and empty conditions was not significant.

the congruent and incongruent conditions with $p < 0.01$ was found. The difference between the congruent and empty conditions was not significant. Moreover, a significant effect of the context condition was found for the proportion of gazing on the relevant anchor with $p < 0.05$.

3.5. Anchor-object transition

Figure 12 shows the anchor-object transition represented by boxen (letter-value) plots. For a large data set, this advanced boxplot type offers an advantage for visualizing the data distribution as it prevents a visual overload by the outliers. For more details on the boxen plots see Seaborn documentation (Waskom et al., 2017). The mean values across all participants with the respective confidence intervals, as well as the individual average values for each participant are shown in the inset plot in Figure 12.

Note, that the positive values indicate that the object was gazed at after the anchor, whereas the negative values correspond to the case when the object was gazed at before the anchor. Furthermore, the data is shown only for the context-rich conditions, as in the empty condition no anchor was present. From linear mixed model analysis, over the course of all trials, a significant difference between the congruent and incongruent conditions was found with $p < 0.05$.

4. Discussion

Only a few studies looked into the scene context impact on human behavior in an interactive task until now. The present study investigated how the scene context affects human object manipulation in a pick-and-place task in a realistic scenario. This study examined whether object manipulation in an isolated setting differs once the object of interest is brought into a scene context. Using a psychophysics approach implemented in a VR environment, we evaluated behavior during three phases: the search, the reaching and picking of the object, and then transporting it to a predefined final location. Specifically, the performance was evaluated in three different conditions: when the object matched the scene context, when it did not fit the context, and when no context was present.

Overall, the experimental paradigm captured well an interactive task in a realistic 3D environment. The possibility

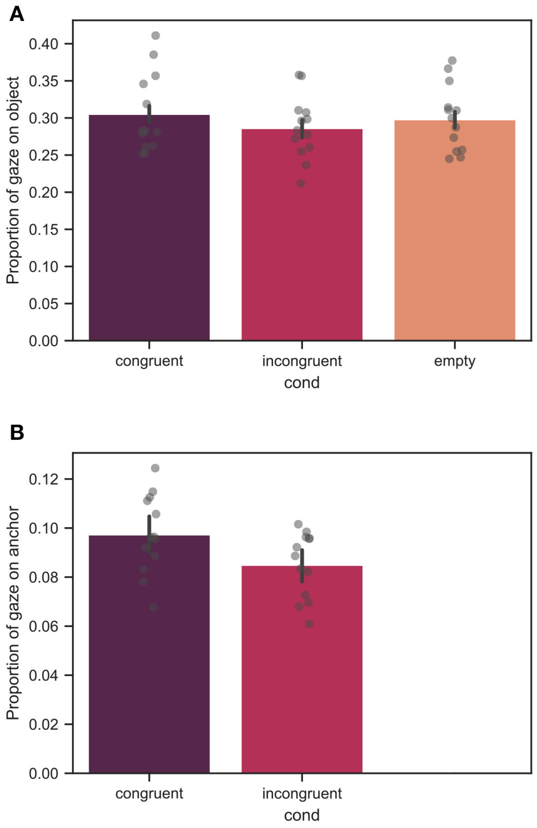


FIGURE 11
The proportion of gazing on the **(A)** target object and **(B)** relevant anchor across all trials, averaged over all measured participants. Each bar corresponds to a separate experimental condition. The error bars indicate confidence intervals of 95% computed using bootstrapping (see main text). The individual points correspond to the mean value for each individual participant. The difference in **(A)** between the congruent and incongruent conditions was significant with $p < 0.01$, whereas no significant difference was found between the congruent and empty conditions. Note, that in **(B)** the proportion of gazing on the relevant anchor is shown only for the context-rich conditions as in the empty condition no anchor was present. There was a significant difference for the proportion of gazing on the relevant anchor with $p < 0.05$.

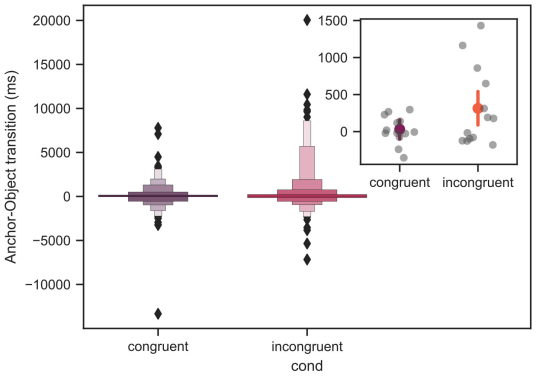


FIGURE 12
Mean anchor-object transition across all trials for all measured participants represented by boxen plots. The black diamond markers correspond to the outliers. For more details on the boxen plot structure see main text. The inset in the upper right location demonstrates the mean values with the respective 95% confidence intervals computed using bootstrapping (see main text). The gray individual points in the inset plot correspond to the mean value for each individual subject. The axes labels of the inset plot are identical to those of the main plot. There was a significant difference between the congruent and incongruent conditions with $p < 0.05$.

of freely moving head and gaze, as well as unconfined hand movement, brought the controlled experimental setting closer to a real-life scenario compared to traditional screen-based paradigms.

4.1. Task duration, search duration, scene coverage

The significant increase of almost a second, which was found for the task duration in the incongruent condition compared to the congruent condition, shows general facilitation of the task performance by the scene context, which is also apparent from previous studies (Biederman et al., 1982; Võ and Wolfe, 2013). In particular, the performance improvement in the present study was caused mainly by the search duration difference, indicated by a significantly longer search time for the incongruent condition. These results are in line with the previous research on the impact of the scene context on visual search (Boettcher et al., 2018; Võ, 2021). Likewise, the eye movement data examination showed a corresponding increase in the proportion of the scene covered by the gaze for the incongruent condition, indicating that when the object did not match the scene context, participants had to gaze around more compared to when the object semantically matched the scene. One explanation of this behavior has been previously proposed in the literature (Biederman et al., 1982; Bar, 2004, 2009). In particular, it was suggested that context (also referred to as the scene grammar) facilitation of individual object recognition in a scene originates in the generation of specific predictions by the observer, which can be later used to find the object. Such, when the object recognition cannot be rapidly resolved based solely on the physical features of the target object, contextual information can contribute more to the efficient recognition of that object than its physical attributes (Bar, 2004, 2009).

Notably, no significant differences were found for the task duration, the search duration, or the scene coverage when comparing the congruent and empty conditions. These results indicate that although there is substantially more visual content in a context-rich environment, it does not seem to distract the user from the task performance. And instead, it enables an efficient search process. Nonetheless, when the target object is placed in the same scenes and the same spatial configurations but does not match the context, the performance decreases, underlining the role of the scene context. In line with previous research, these findings suggest that in the context-poor "empty" environment, the participants based their search primarily on the physical attributes of the target. When the same target was placed in the incongruent scene context, that is, the contextual information was non-informative, a substantially larger number of visual stimuli served as a distractor for the searcher and led to longer search times. In contrast, when the contextual information was relevant, it compensated for a large number of visual stimuli. It is suggested, that the context maintained the search time comparable to that in the "empty" scenario.

It is worth mentioning that the present interpretation of the results is based on the assumption that the search in the empty condition is relatively complex. During the study development, as mentioned in Section 2.3.2, the number of distractors for all scenes was selected rather large to ensure a considerable search time even for the empty condition, here, at least 4 s on average. However, no detailed analysis of the search complexity in the empty condition at a given number of distractors was performed. To strengthen the present results and their interpretation, in future studies, it is recommended to deliberately analyze the search complexity in the empty condition, for example, through systematic modulation of the number of distractors and evaluation of the respective search times.

4.2. Reach and transport duration

Until now, few studies attempted to describe the effect of the scene context in an interactive task rather than a search display. In one recent study, participants were asked to construct environments from a set of various available virtual objects, either according to the contextual scene expectations of the observer or against them (Draschkow and Võ, 2017). Among other findings, the authors demonstrated that participants held objects for a longer time in the context-incongruent condition. In the present study, we similarly evaluated whether the scene context affected the transport duration of the target object. In contrast to (Draschkow and Võ, 2017), in the present study, the final location was predefined and kept constant throughout all trials. Therefore, participants did not need additional time to decide where to put the object. Nonetheless, it took participants less time to transport the object to its final location in the empty condition. Further analysis did not reveal any differences in the number of gaze points on the target object during the transport phase across the conditions. We, therefore, speculate that since the final location was predefined and the target object was not looked at more in any of the conditions during its transport, no additional processing was necessary. However, in context-rich environments, participants had to overcome more obstacles when transporting the target object compared to empty rooms with shelves, thus, leading to a slightly longer transport duration. Beyond the scope of the current work, further studies should address whether the uncertainty of the final location would introduce a variation in transport duration in different scene context conditions, as well as systematically explore the impact of the obstacles.

Considering the reach phase duration, previous studies have demonstrated motor inhibition when approaching dangerous objects due to the emergence of aversive affordances (Mustile et al., 2021). Furthermore, from visual search research, it is known that violations of one's scene grammar lead to longer and more fixations on the critical objects, which is typically attributed to more extended processing of those objects (Henderson et al., 1999; Cornelissen and Võ, 2017; Draschkow and Võ, 2017). In the present study, no significant differences in the reach duration were revealed across the conditions, meaning no effect of the scene context was found on the reach duration. It is suggested that if the exact target object is known before the trial start, it is still harder to find it in a semantically not matching context. However, once it is located, no extra processing is necessary. Therefore, no significant elongation of the reach phase emerged in the incongruent condition. Further studies are required to evaluate the impact of the scene context on the reach phase when the target object is unfamiliar to the observer before the task.

4.3. Target object and anchor

A small but significant difference in the gaze proportion on the target object correlated with the notion of task facilitation by the scene context. As expected, participants spent less time and effort to find the object in the congruent condition, resulting in proportionally more gazing at the object. More interesting, however, was the significant difference in gazing on the relevant anchor across the conditions. In particular, the anchor was gazed on more often when it could potentially be helpful to perform the task, meaning, in the congruent condition. The role of the anchors became more apparent from the recent studies (Boettcher et al., 2018; Võ, 2021). In natural contexts, people seem to be able to exploit the knowledge about the scene configurations when looking for an object (Võ and Wolfe, 2013; Draschkow and Võ, 2017). Furthermore, people tend to rely on a rather global context than local information, and, thus, the larger scene-typical objects—anchors—appear to have more influence on the search facilitation. The present study results confirm this notion, where the anchor in the incongruent condition appeared to be less supportive of the task performance in contrast to the congruent condition.

Considering the previously suggested scene processing (Võ, 2021) where the small objects are located after larger global objects, we generally expected the anchors to be fixated before the target objects in most of the trials. In contrast, the distribution of the anchor-object transition did not confirm our expectations. From literature, it is known that the search strategy can be composed of several processes, including feature guidance as well as scene guidance (Võ and Wolfe, 2013). In the current study, it is speculated that participants utilized scene guidance to facilitate the search process throughout the trials, which is reflected in better task performance. However, the unique appearance of the target objects and distractors in a given scene defined a set of specific features which could enforce the feature guidance (grayscale same-size cubes). Therefore, the anchors were not always fixated before the target object. Nonetheless, the anchor-object distribution appeared to be significantly narrower in the congruent condition, which underlines the role of the anchors even if it was fixated after the target object was already located. Further studies are necessary to investigate the dynamics of the mutual object and anchor fixations. Moreover, in the future studies it would be interesting to further address the dynamic nature of fixating the target object and the anchor throughout the search.

4.4. Limitations

Although the present work captured the effect of the scene context on the performance in an interactive pick-and-place task, it is important to comment on its limitations. First, even though there are recent attempts to formally define what anchors are (Draschkow and Võ, 2017), to the best of our knowledge, there is no existing database. Therefore, the anchors used in this study were selected intuitively based on a common understanding of the typical scenes. This, in turn, could inflate the individual differences in perceiving the intended anchors as such and reduce the effect of the scene context on the task performance. Furthermore, due to the natural scene contexts, it was not possible to design the anchors of uniform size as well as set the target object to the same position relative to the anchors. Thus, although the target objects were always in the proximity of the anchors, in some configurations, they were next to the anchors, whereas, in others, they were directly on top of them. This could, for example, influence the anchor-object transition. Another challenge for the paradigm design was variability in the similarities between the target objects and the distractors. As such, a toothbrush is more likely to be at first confused with a fork than with a soap dispenser due to the shape similarity, which would possibly increase the total search time of the target object. It is not a straightforward task to avoid this limitation due to the realistic nature of the objects and the scenes. Nonetheless, it could be advantageous to do a more systematic generation of the target objects and distractors sets in the future. Furthermore, in the present study, the effect of specific objects and scenes was not tested due to a very limited number of trials per object and scene. In future studies, it would be interesting to address how specific objects and scenes impact the behavior in a pick-and-place task. Finally, in the present study, grasping was implemented using the VR controller. While VR offers an excellent opportunity to simulate realistic scenarios in a lab environment and accurately track hand motion, a more natural solution would be to implement the grasping using only the participant's hand without a controller. This, however, significantly increases the complexity of the setup when it comes to reliable controller-free object manipulation. When manipulating objects in VR, some recent studies demonstrated a strong sense of ownership when the virtual hand is represented by a hand-like object which we also used in the present study. Nonetheless, the direct transformability of the current study results in a real-world grasping scenario should be a focus of future studies. For example, it would be interesting to compare the dynamics of a pick-and-place task in the real scenario and its replica in a virtual environment.

5. Conclusion

To conclude, this work evaluated the impact of the scene context on the performance of an interactive task, precisely, the pick-and-place task where the object had to be found, picked, and transported to a predefined location. In line with visual search literature, we found a disadvantage in search time when the object does not belong to the scene context compared to the context-congruent condition. When comparing the congruent and no-context conditions, the search performance was similar. This finding supports the notion that when the object fits the scene, the other objects and the context-rich environment itself seem to not introduce an additional distraction for the searcher and keep the search efficient.

The reach phase duration was not affected by the scene context. A small difference was found in the transport phase duration between the empty condition and both context-rich conditions. However, as discussed in Section 4.2, the elongation seems to be originating from the need to overcome some obstacles in the context-rich environments and not due to additional processing. Although this suggestion requires further systematic testing, at least in the present configuration where the final location for the object was known, the semantic congruency of the object and the scene context does not seem to affect the interactive phases of the pick-and-place task. This strengthens the validity of transferring eye and hand movement knowledge in a grasping task performed in an isolated setting to a realistic scenario within a context-rich environment.

The present study contributes to a better understanding of the dynamics of the pick-and-place task once the target object is placed in a realistic context-rich scene. Keeping the possible applications in mind, the findings of this work provide insights into the potential development of supporting intention predicting systems. In particular, the information about the object's semantic congruency with the scene context could potentially be used as an additional input parameter to train and calibrate future assistive algorithms for the support system. On a broader scope, the findings of the present study can be relevant for designing intention prediction-based assistive systems helping, e.g., visually impaired with intelligent tunable lenses, or to control prosthetics like robotic arms, wheelchairs, or exoskeletons.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: osf.io/grm45.

Ethics statement

The studies involving human participants were reviewed and approved by Faculty of Medicine at the University of Tübingen with a corresponding ethical approval identification code 986/2020BO2. The patients/participants provided their written informed consent to participate in this study.

Author contributions

OL-S, RA, and SW developed the research idea, designed the experiment, edited and finally approved the manuscript. OL-S programmed the experiment, collected the data, conducted the analysis, and wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The present research was supported by the funding from the European Union's Horizon 2020 research and innovation program under Grant agreement No. 951910, and by the German Research Foundation (DFG): SFB 1233, Robust Vision: Inference Principles and Neural Mechanisms, TP TRA, project number: 276693517.

Acknowledgments

We acknowledge support by Open Access Publishing Fund of University of Tübingen.

Conflict of interest

We declare that OL-S and SW are scientists at the University of Tübingen and employees of Carl Zeiss Vision International GmbH, as detailed in the affiliations.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.881269/full#supplementary-material

References

Ballard, D. H., Hayhoe, M. M., Li, F., and Whitehead, S. D. (1992). Hand-eye coordination during sequential tasks. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 337, 1281. doi: 10.1098/rstb.1992.0111

Bar, M. (2004). Visual objects in context. *Nat. Rev. Neurosci.* 5, 617–629. doi: 10.1038/nrn1476

Bar, M. (2009). The proactive brain: memory for predictions. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 364, 1235–1243. doi: 10.1098/rstb.2008.0310

Beitner, J., Helbing, J., Draschkow, D., and Võ, M. L. (2021). Get your guidance going: Investigating the activation of spatial priors for efficient search in virtual reality. *Brain Sci.* 11, 1–17. doi: 10.3390/brainsci11010044

Biederman, I., Mezzanotte, R. J., and Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cogn. Psychol.* 14, 143–177. doi: 10.1016/0010-0285(82)90007-X

Boettcher, S. E., Draschkow, D., Dienhart, E., and Võ, M. L. (2018). Anchoring visual search in scenes: assessing the role of anchor objects on eye movements during visual search. *J. Vision* 18, 1–13. doi: 10.1167/18.13.11

Brockmole, J. R., Castelhano, M. S., and Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: global and local contexts. *J. Exp. Psychol. Learn. Memory Cogn.* 32, 699–706. doi: 10.1037/0278-7393.32.4.699

Cai, Q., Li, J., and Long, J. (2021). Effect of physical and virtual feedback on reach-to-grasp movements in virtual environments. *IEEE Trans. Cogn. Dev. Syst.* 14, 708–714. doi: 10.1109/TCDS.2021.3066618

Chun, M. M., and Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cogn. Psychol.* 36, 28–71. doi: 10.1006/cogp.1998.0681

Cornelissen, T. H., and Võ, M. L. (2017). Stuck on semantics: processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Attent. Percept. Psychophys.* 79, 154–168. doi: 10.3758/s13414-016-1203-7

de Brouwer, A. J., Flanagan, J. R., and Spering, M. (2021). Functional use of eye movements for an acting system. *Trends Cogn. Sci.* 25, 252–263. doi: 10.1016/j.tics.2020.12.006

Draschkow, D., and Võ, M. L.-H. (2017). Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Sci. Rep.* 7, 16471. doi: 10.1038/s41598-017-16739-x

Furmanek, M. P., Schettino, L. F., Yarossi, M., Kirkman, S., Adamovich, S. V., and Tunik, E. (2019). Coordination of reach-to-grasp in physical and haptic-free virtual environments. *J. Neuroeng. Rehabil.* 16, 78. doi: 10.1186/s12984-019-0525-9

Gibson, J. J. (2014). *The Ecological Approach to Visual Perception*. New York: Psychology Press. doi: 10.4324/9781315740218

Gregori, V., Cognolato, M., Saetta, G., Atzori, M., and Gijsberts, A. (2019). On the visuomotor behavior of amputees and able-bodied people during grasping. *Front. Bioeng. Biotechnol.* 7, 316. doi: 10.3389/fbioe.2019.00316

Henderson, J. M., Weeks, P. A., and Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 210–228. doi: 10.1037/0096-1523.25.1.210

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 99–104. doi: 10.1109/MCSE.2007.55

Imaoka, Y., Flury, A., and de Bruin, E. D. (2020). Assessing saccadic eye movements with head-mounted display virtual reality technology. *Front. Psychiatry* 11, 572938. doi: 10.3389/fpsyt.2020.572938

Johansson, R. S., Westling, G., Bäckström, A., and Randall Flanagan, J. (2001). Eye-hand coordination in object manipulation. *J. Neurosci.* 21, 6917–6932. doi: 10.1523/JNEUROSCI.21-17-06917.2001

Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D. H., and Gowda, S. M. (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Trans. Biomed. Eng.* 57, 2635–2645. doi: 10.1109/TBME.2010.2057429

Kübler, T. C. (2020). "The perception engineer's toolkit for eye-tracking data analysis," in *Eye Tracking Research and Applications Symposium (ETRA)* (New York, NY: Association for Computing Machinery), 1–4. doi: 10.1145/3379156.3391366

Lavoie, E., and Chapman, C. S. (2021). What's limbs got to do with it? Real-world movement correlates with feelings of ownership over virtual arms during object interactions in virtual reality. *Neurosci. Conscious*. doi: 10.1093/nc/niaa027

Lavoie, E. B., Valevicius, A. M., Boser, Q. A., Kovic, O., Vette, A. H., Pilarski, P. M., et al. (2018). Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during objectinteraction tasks. *J. Vision* 18, 1–20. doi: 10.1167/18.6.18

Lazarou, I., Nikolopoulos, S., Petrantonakis, P. C., Kompatsiaris, I., and Tsolaki, M. (2018). EEG-based brain-computer interfaces for communication and rehabilitation of people with motor impairment: a novel approach of the 21st century. *Front. Hum. Neurosci.* 12, 14. doi: 10.3389/fnhum.2018.00014

Leube, A., Rifai, K., and Wahl, S. (2017). Sampling rate influences saccade detection in mobile eye tracking of a reading task. *Journal of Eye Movement Research* 10(3). doi: 10.16910/jemr.10.3.3

Levin, M. F., Magdalon, E. C., Michaelsen, S. M., and Quevedo, A. A. (2015). Quality of grasping and the role of haptics in a 3-D immersive virtual reality environment in individuals with stroke. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 1047–1055. doi: 10.1109/TNSRE.2014.2387412

Lougiakis, C., Katifori, A., Roussou, M., and Ioannidis, I.-P. (2020). "Effects of virtual hand representation on interaction and embodiment in HMD-based virtual environments using controllers," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Atlanta, GA), 510–518. doi: 10.1109/VR46266.2020.1581086151885

Lukashova-Sanz, O., and Wahl, S. (2021). Saliency-aware subtle augmentation improves human visual search performance in VR. *Brain Sci.* 11, 283. doi: 10.3390/brainsci11030283

Marek, N., and Pollmann, S. (2020). Contextual-cueing beyond the initial field of view-a virtual reality experiment. *Brain Sci.* 10, 446. doi: 10.3390/brainsci1007 0446

McKinney, W. (2010). "Data structures for statistical computing in Python," in *Proceedings of the 9th Python in Science Conference T (SCIPY 2010)* (Austin, TX), 56–61. doi: 10.25080/Majora-92bf1922-00a

Mustile, M., Giocondo, F., Caligiore, D., Borghi, A. M., and Kourtis, D. (2021). Motor inhibition to dangerous objects: electrophysiological evidence for task-dependent aversive affordances. *J. Cogn. Neurosci.* 33, 826–839. doi: 10.1162/jocn_a_01690

Olk, B., Dinu, A., Zielinski, D. J., and Kopper, R. (2018). Measuring visual search and distraction in immersive virtual reality. *R. Soc. Open Sci.* 5, 172331. doi: 10.1098/rsos.172331

Olsen, A. (2012). *The Tobii I-VT Fixation Filter Algorithm Description*. Stockhom.

Over, E. A., Hooge, I. T., Vlaskamp, B. N., and Erkelens, C. J. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision Res.* 47, 2272–2280. doi: 10.1016/j.visres.2007.05.002

Pereira, E. J., and Castelhano, M. S. (2019). Attentional capture is contingent on scene region: using surface guidance framework to explore attentional mechanisms during search. *Psychon. Bull. Rev.* 26, 1273–1281. doi: 10.3758/s13423-019-01610-z

Pinheiro, J., Bates, D., and R Core Team (2022). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3. Montreal, QC, Canada, 1–157.

Salvucci, D. D., and Goldberg, J. H. (2000). "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the Eye Tracking Research and Applications Symposium 2000* (New York, NY: Association for Computing Machinery), 71–78. doi: 10.1145/355017.355028

Sensinger, J. W., and Dosen, S. (2020). A review of sensory feedback in upper-limb prostheses from the perspective of human motor control. *Front. Neurosci.* 14, 345. doi: 10.3389/fnins.2020.00345

Shafti, A., Orlov, P., and Faisal, A. A. (2019). "Gaze-based, context-aware robotic system for assisted reaching and grasping," in *Proceedings - IEEE International Conference on Robotics and Automation* (Italy). doi: 10.1109/ICRA.2019.8793804

Subramanian, M., Park, S., Orlov, P., Shafti, A., and Faisal, A. A. (2021). "Gaze-contingent decoding of human navigation intention on an autonomous wheelchair platform," in *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*. doi: 10.1109/NER49283.2021.9441218

Tian, H., Wang, C., Manocha, D., and Zhang, X. (2019). Realtime hand-object interaction using learned grasp space for virtual environments. *IEEE Trans. Visual. Comput. Graph.* 25, 2623–2635. doi: 10.1109/TVCG.2018.2849381

Torralba, A., Oliva, A., Castelhano, M. S., and Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol. Rev.* 113, 766–786. doi: 10.1037/0033-295X.113.4.766

Unity Technologies (2019). *Unity*.

Van Der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng.* 13, 22–30. doi: 10.1109/MCSE.2011.37

Viau, A., Feldman, A. G., McFadyen, B. J., and Levin, M. F. (2004). Reaching in reality and virtual reality: a comparison of movement kinematics in healthy subjects and in adults with hemiparesis. *J. Neuroeng. Rehabil.* 1, 10. doi: 10.1186/1743-0003-1-11

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-019-0686-2

Võ, M. L., and Wolfe, J. M. (2013). The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition* 126, 198–212. doi: 10.1016/j.cognition.2012.09.017

Võ, M. L. H. (2021). The meaning and structure of scenes. *Vision Res.* 181, 10–20. doi: 10.1016/j.visres.2020.11.003

Võ, M. L. H., Boettcher, S. E., and Draschkow, D. (2019). Reading scenes: how scene grammar guides attention and aids perception in real-world environments. *Curr. Opin. Psychol.* 29, 205–210. doi: 10.1016/j.copsyc.2019.03.009

Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., et al. (2017). *mwaskom/seaborn: v0.8.1*.

Wolfe, J. M. (2020). Visual search: how do we find what we are looking for? *Annu. Rev. Vision Sci.* 6, 2.1–2.24. doi: 10.1146/annurev-vision-091718-015048

Wurm, M. F., and Schubotz, R. I. (2017). What's she doing in the kitchen? Context helps when actions are hard to recognize. *Psychon. Bull. Rev.* 24, 503–509. doi: 10.3758/s13423-016-1108-4

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 27 September 2022
DOI 10.3389/fpsyg.2022.904132

OPEN ACCESS

EDITED BY
Tom Carr,
Michigan State University,
United States

REVIEWED BY
Tahamina Begum,
Universiti Sains Malaysia (USM),
Malaysia
Grainne McLoughlin,
University of California, San Diego,
United States

*CORRESPONDENCE
Hong Shi
hzs0032@auburn.edu

SPECIALTY SECTION
This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

RECEIVED 25 March 2022
ACCEPTED 05 September 2022
PUBLISHED 27 September 2022

CITATION
Shi H (2022) Inhibition modulated by self-efficacy: An event-related potential study.
Front. Psychol. 13:904132.
doi: 10.3389/fpsyg.2022.904132

Inhibition modulated by self-efficacy: An event-related potential study

Hong Shi*

Department of English Language and Literature, School of Foreign Languages, China University of Petroleum, Beijing, China

Inhibition, associated with self-efficacy, enables people to control thought and action and inhibit disturbing stimulus and impulsion and has certain evolutionary significance. This study analyzed the neural correlates of inhibition modulated by self-efficacy. Self-efficacy was assessed by using the survey adapted from the Motivated Strategies for Learning Questionnaire. Fifty college students divided into low and high self-efficacy groups participated in the experiments. Their ability to conduct inhibitory control was studied through Go/No-Go tasks. During the tasks, we recorded students' brain activity, focusing on N2 and P3 components in the event-related potential (ERP). Larger No-Go N2 amplitudes for the high self-efficacy group were found compared with the low self-efficacy group. Conflict detection as represented by N2 was modulated by self-efficacy, whereas conflict inhibition as represented by P3 was not modulated by self-efficacy. The highly self-efficacious students were more capable of detecting conflicts but not necessarily more capable of inhibiting action given that conflict was detected. Taken together, these findings offer neurophysiological evidence of the important regulatory role of self-efficacy in inhibitory control ability development.

KEYWORDS

inhibition, self-efficacy, Go/No-Go task, neural correlate, ERP

Introduction

Inhibition or inhibitory control, referred to as response inhibition, is an executive function defined as the ability to deliberately withhold or override a dominant, prepotent (habitual), or automatic response to resist distraction or temptation and to achieve the desired goal (Nigg, 2000; Gagne, 2017; Kloo and Sodian, 2017). Inhibition in response to a stimulus is associated with self-efficacy (Bembenutty, 2011; Mcauley et al., 2011). Self-efficacy refers to self-perceptions or beliefs of the capability to learn or perform tasks at designated levels (Bandura, 1997). Students with high self-efficacy may have stronger inhibitory control ability and would continue working even when a task-irrelevant temptation to stop calls for attention, and they may be better able to exit from an ongoing action sequence in response to a task-relevant signal to do so.

However, students with low self-efficacy beliefs may have weaker inhibitory control ability and tend to succumb to temptation and let disruptive thoughts interfere with performance or miss a task-relevant signal to interrupt an ongoing action (Bembenutty, 2011; Mcauley et al., 2011; Wang, 2018). Previous studies have documented individuals' behavioral performance on measures of inhibition, and for example, there is evidence that inhibitory control is related to academic skills (e.g., Cragg and Nation, 2008; Allan et al., 2014; Dekker et al., 2017; Morgan et al., 2019; Litkowski et al., 2020). Furthermore, previous studies identified a modulation role for self-efficacy in inhibition (e.g., Mcauley et al., 2011; Liew, 2012; Gärtner et al., 2018), but few studies have examined neural underpinnings of inhibition modulated by self-efficacy. The present study investigates the underlying brain neural correlates of inhibition modulated by self-efficacy using the event-related potential (ERP) method. ERPs can provide evidence of the brain mechanisms of response inhibition, and the ERP literature has examined N2 and P3 components to analyze the inhibition process (e.g., Smith et al., 2006; Enriquez-Geppert et al., 2014; Wang, 2018). Therefore, we want to clarify the potential mechanisms underlying inhibition modulated by self-efficacy through the analysis of N2 and P3 components.

The common way to examine the neural correlates of inhibition is using Go/No-Go tasks (Rueda et al., 2005; Wiebe et al., 2012; Smith et al., 2013). In a typical Go/No-Go task, individuals respond to a frequently occurring stimulus type (Go trials), and inhibit their response when a less frequently occurring stimulus is presented (No-Go trials). The extent to which participants are able to withhold a response on No-Go trials serves as a measure of their inhibitory control abilities. The bulk of the Go/No-Go ERP studies focusing on inhibition have examined the No-Go N2-P3 complex (e.g., Lahat et al., 2010; Hoyniak, 2017). Some current views claimed that the conflict detection operation, which is a part of the inhibition process, is associated with the N2, whereas inhibition of the action operation is associated with the P3 (Smith et al., 2008; Albert et al., 2013; Enriquez-Geppert et al., 2014). The N2, an increased frontal scalp negativity, is a wave within 150–400 ms after the onset of the stimulus, which peaks at approximately 300 ms post-stimulus onset. The N2 component can be used to index conflict detection between Go and No-Go response tendencies, and its amplitude is largest when the response conflict is high. Randall and Smith (2011) called it the conflict detection hypothesis of the N2. In many experiments, the N2 components are greater in the No-Go trials compared with the Go trials (e.g., Jodo and Kayama, 1992; Hoyniak, 2017; Lahat et al., 2010; Waldvogel et al., 2000). Studies revealed that the N2 was the unlikely equivalent of proper motor inhibition, and the N2 was evoked when stimulus constellations were associated with conflicts in information processing even though a response has to be executed (Huster et al., 2013). Enriquez-Geppert et al. (2010) suggested an association of the N2 with conflict-related effects with less frequently occurring trial types. Frequent responses are prepotent, eventually leading to conflicts at the response representation level when infrequent responses have to be made (Braver et al., 2001; Jones et al., 2002). Regarding N2, although almost no research examined the neurocognitive correlates of inhibition modulated by self-efficacy, researchers have investigated inhibition regulated by emotional induction or anxiety. A person's affective states and actions are closely associated with self-efficacy (Bandura, 1997), and they could result in improved inhibitory control and increases in associated aspects of brain activity (Farbiash and Berger, 2015). Students with increased attention and higher effortful brain activation tended to view the execution of tasks as under their control. For example, Farbiash and Berger (2015) reported inhibition of No-Go trials was associated with larger N2 amplitudes during negative emotional induction for children aged 5–6. Hum et al. (2013) also found significantly larger No-Go N2 amplitudes for 8–12 year olds with anxiety than those without anxiety. The other manifestation of the inhibition mechanism is P3. It is the positive component that appears in the frontal center. Its manifestation is roughly a positive wave in the 300–600 ms range. The P3 amplitude differences have been found in response to Go versus No-Go trials (Ramautar et al., 2004; Dimoska et al., 2006). The P3 amplitude is larger in the No-Go trials (e.g., Bokura et al., 2001; Ciesielski et al., 2004). It seems that the recent predominant literature on Go/No-Go tasks identified that the P3 was directly related to the suppression of overt motor response (Huster et al., 2013). The majority of analyses indicate that P3 originates from multiple brain regions including frontal and temporo-parietal areas (Polich, 2007). Wang (2018) examined the brain inhibitory effect of self-efficacy of college students to English biological and non-biological vocabulary stimuli and identified that the No-Go P3 amplitude in the high self-efficacy group was larger than that in the low self-efficacy group, and indicated that students with high self-efficacy had better inhibitory control ability. Rosen (2010) adopted a different task—a flanker task (a task that varies task difficulty without changing the nature of the task due to its use of congruent and incongruent flanking stimuli, and the incongruent task requires greater interference control to inhibit task-irrelevant stimuli and execute the correct response) and found self-efficacy was related to enhanced stimulus processing, as evidenced by larger P3 amplitudes. Themanson and Rosen (2014) also found a positive relationship between self-efficacy and P3 amplitude during the completion of a flanker task. But these studies represented specific research designs and did not provide enough evidence for inhibition mechanisms modulated by self-efficacy. Furthermore, although some related studies adopted Go/No-Go tasks, both No-Go N2 and No-Go P3 are not consistently found associated with self-efficacy. The present study thus intends to use Go/No-Go tasks to clarify the inhibition process modulated by self-efficacy through the analysis of N2 and P3 components so that we can learn more about people's control over thought and action to allow them to reduce interference and maintain goal-oriented actions. The experimental hypothesis is that participants with high levels of self-efficacy have stronger inhibitory control ability on stimulus and interference, and thus have larger No-Go N2 and No-Go P3 in terms of test indicators.

Materials and methods

The ability to conduct inhibitory control is studied through Go/No-Go tasks. The experimental indicators are No-Go N2 and No-Go P3.

Participants

This study was conducted at a public university of science and engineering in the city of Beijing, China. We used convenience sampling for this study. There were 61 students who had previously been tested for self-efficacy and appropriate samples were selected based on their pretest scores. Twenty-four college students with high self-efficacy were selected, and 26 college students with low self-efficacy were selected. There were 12 females and 12 males, 11 undergraduate students and 13 graduate students in the high self-efficacy group; and 14 females and 12 males, 12 undergraduate students and 14 graduate students in the low self-efficacy group. The experimental participants were right-handed, with normal or corrected vision, and they had no history of mental illness. At the request of the Academic Ethics Committee of the university, they signed informed consent to participate in this study. **Table 1** (see **Appendix B**) shows the demographic information of the participants.

Self-reported instrument

Self-efficacy was assessed before Go/No-Go tasks by using the survey adapted from the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al., 1993). The MSLQ has been validated and used in many studies (e.g., Pintrich et al., 1991, 1993; Pintrich, 2003). This questionnaire is a self-report instrument designed to assess college students' motivational orientations and self-regulated learning, and the self-efficacy subscale in MSLQ is designed particularly to measure the self-efficacy beliefs of students (Pintrich et al., 1991). This study mainly used the self-efficacy subscale in MSLQ to measure the self-efficacy of participants (see **Appendix A**). The value of Cronbach's alpha for the self-efficacy scale was 0.902. Students rated themselves on a 9-point Likert scale, from 1 (not at all true of me) to 9 (very true of me). It was a median split (the score for high self-efficacy is ≥ 5, and for low self-efficacy is < 5). The mean score for the high self-efficacy group is 7.06 (range of scores is 5.38–8.79), and the mean score for the low self-efficacy group is 3.58 (range of scores 2.31–4.83).

Stimuli and procedure

Before the experiment, the experimental procedure was described and the students participated in the experiment in a relaxed state. Pictures including single triangle and double triangle were selected as the stimulus. The participants were instructed to respond by pressing the button "/" (right hand) or "z" (left hand) as quickly as they could whenever the "Go" stimulus (double triangle) was presented and not to press the button when the "No-Go" stimulus (single triangle) was presented. A fixation cross to orient attention to the middle of the screen was presented for 500 ms. Stimuli were presented for 50 ms, and participants could respond anytime within the onset of the stimulus and the interstimulus interval (950 ms). A practice phase of 20 trials with feedback was given. Each condition consisted of 150 test trials. Go stimuli were presented for most of the trials (80%). All experimental tasks were presented using the E-prime software 3.0.

Electroencephalogram recording and preprocessing

Electroencephalogram (EEG) signals were continuously recorded with the NeuroLab digital amplifier system (Yiran Sunny Technology Co., Ltd., Beijing, China), using NeuCap with Ag/AgCl electrodes at 32 sites according to the extended international 10–20 system. The reference electrode was placed on the nose tip. The nose-tip reference was converted into bilateral mastoid for reference in offline data analysis. Vertical and horizontal electrooculography (EOG) signals were recorded with two electrodes placed above and below the right eye and with two electrodes at the right and left outer canthi of the eyes, respectively. Data were recorded continuously at a sampling rate of 1,000 Hz and filtered offline with a bandpass of 0.01–100 Hz. Electrode impedance was maintained below 5 kΩ throughout the experiment.

The EEGLab software[1] was used to analyze EEG data. Blinks were corrected using an ICA procedure. The plotted average, condition-specific activation had to account for the N2/P3 time course. The EOG components were identified and selected according to the topographical maps that had to show a fronto-central scalp distribution, which is usually seen with the N2 and P3. One of the EOG components was removed on average. They were performed on all subjects. Remaining artifacts exceeding $\pm\ 100\mu V$ in amplitude or containing a

---

[1] https://sccn.ucsd.edu/eeglab/index.php

change of over 100μV within a period of 50 ms were rejected. After performing EOG correction and visual inspection, only artifact-free trials were considered (rejected epochs, 3%).

The EEG was segmented in epochs of 700 ms, time-locked to picture onset, and included a 100 ms pre-stimulus baseline. Trials contaminated by amplifier clipping, bursts of EMG activity, or peak-to-peak deflection exceeding ± 100 μV were excluded from averaging (excluded 5%). The averaged ERP waveforms were low-pass filtered at 30 Hz. Filter frequencies correspond to the half-amplitude cut-off (24 dB/octave rolloff).

Data analysis

For response times, only Go trials with correct responses were included. There were no response times for correct response No-Go trials, as no response was made if participants inhibited the response successfully. Means for all the conditions are presented in **Table 2**. Since the accuracy is high, the ERP analyses are restricted to correct-response trials.

The ERP literature has examined Fz, Cz, and Pz to analyze the inhibition process or self-efficacy effects (e.g., Themanson et al., 2008; Zhao et al., 2019). The Fz, Cz, and Pz electrodes in the electroencephalogram record were selected for the statistical analysis of mean amplitude over the time window of 200–400 ms and 400–600 ms. Repeated-measures ANOVA was conducted with Go/No-Go trials and electrode sites (Fz, Cz, and Pz) as within-subject factors and self-efficacy level (high vs. low) as between-subject factors. The Greenhouse-Geisser correction was applied where sphericity was violated. When a main effect or interaction was significant, *post hoc* comparisons were performed with Bonferroni correction. Analyses were conducted using the SPSS software (Version 26, SPSS Inc.).

Results

N2 amplitude

While repeated-measures ANOVA (see **Table 3** in **Appendix B**) revealed the main effect of Go/No-Go trials ($F(1,48) = 4.48$, $p = 0.04$, partial $\eta^2 = 0.09$) and the main effect of the site ($F(2,96) = 6.21$, $p = 0.003$, partial $\eta^2 = 0.11$), with the Go trials (−2.59 μV) exhibiting smaller amplitude than the No-Go trials (−2.83 μV), and the N2 amplitude at the Cz (- 3.45 μV) significantly larger than those at the Fz (- 1.19 μV) and Pz (- 2.77 μV), there was no interaction among self-efficacy level, Go/No-Go trials, and electrode sites, between Go/No-Go trials and electrode sites, and between self-efficacy and electrode sites. But the main effect of self-efficacy was significant ($F(1,48) = 7.12$, $p = 0.01$, partial $\eta^2 = 0.13$). There is interaction between self-efficacy and Go/No-Go trials ($F(1,48) = 4.76$, $p = 0.03$, partial $\eta^2 = 0.09$). The amplitudes of No-Go N2 were significantly larger in the high self-efficacy group (- 3.42 μV) than in the low self-efficacy group (- 0.83 μV, $p < 0.03$).

P3 amplitude

Repeated-measures ANOVA (see **Table 4** in **Appendix B**) revealed no main effect of self-efficacy and Go/No-Go trials, but there was a main effect of the site ($F(2,96) = 6.74$, $p = 0.002$, partial $\eta^2 = 0.12$), with the P3 amplitude at the Pz (0.76 μV) significantly larger than those at the Cz (- 0.47 μV) and Fz (- 0.70 μV). A significant interaction effect was found between Go/No-Go trials and electrode sites ($F(2,96) = 18.77$, $p < 0.001$, partial $\eta^2 = 0.28$), with no other interaction effect observed. When we explored each site, self-efficacy level and Go/No-Go trials significantly interacted at only Pz ($F(1,48) = 5.02$, $p = 0.01$, partial $\eta^2 = 0.12$), and an enhanced P3 amplitude on No-Go trials in high self-efficacy group (- 0.30 μV) compared with low self-efficacy group (- 0.61 μV) was identified, but this difference was not significant ($p > 0.1$).

Figure 1 shows the ERP waveform of electrode sites Fz, Cz, and Pz of high self-efficacy group (**Figure 1A**) and low self-efficacy group (**Figure 1B**). We can see that the overall wave amplitude development trend of the N2 and P3 waveforms was obvious. In different task conditions, the amplitudes of No-Go N2 and No-Go P3 were generally larger than that of Go N2 and Go P3 respectively. **Figure 2** shows that compared with the low self-efficacy group the amplitudes of No-Go N2 in the high self-efficacy group were generally larger than that of Go N2, especially at the site Cz. **Figure 3** shows the topographical map of the high self-efficacy group (**Figure 3A**) and the low self-efficacy group (**Figure 3B**). From the location of the brain region, the maximum amplitude of P3 occurred in the frontal center area.

Discussions

This study investigated the neural correlates of inhibition modulated by self-efficacy. Two groups (low self-efficacy vs. high self-efficacy) of 50 college students participated in the experiments. We adopted Go/No-Go tasks and selected two ERP components associated with inhibitions—N2 and P3. The amplitudes of No-Go N2 were significantly larger in the high self-efficacy group compared with the low self-efficacy group. But there was no significant difference in P3 amplitudes on No-Go trials between the high self-efficacy group and the low self-efficacy group.

The finding of larger No-Go N2 amplitudes in the high self-efficacy group indicates that a person with a higher self-efficacy level is more likely to have a stronger ability to detect a conflict. According to Randall and Smith (2011)'s conflict detection hypothesis of the N2, it was used as a neural index of the ability

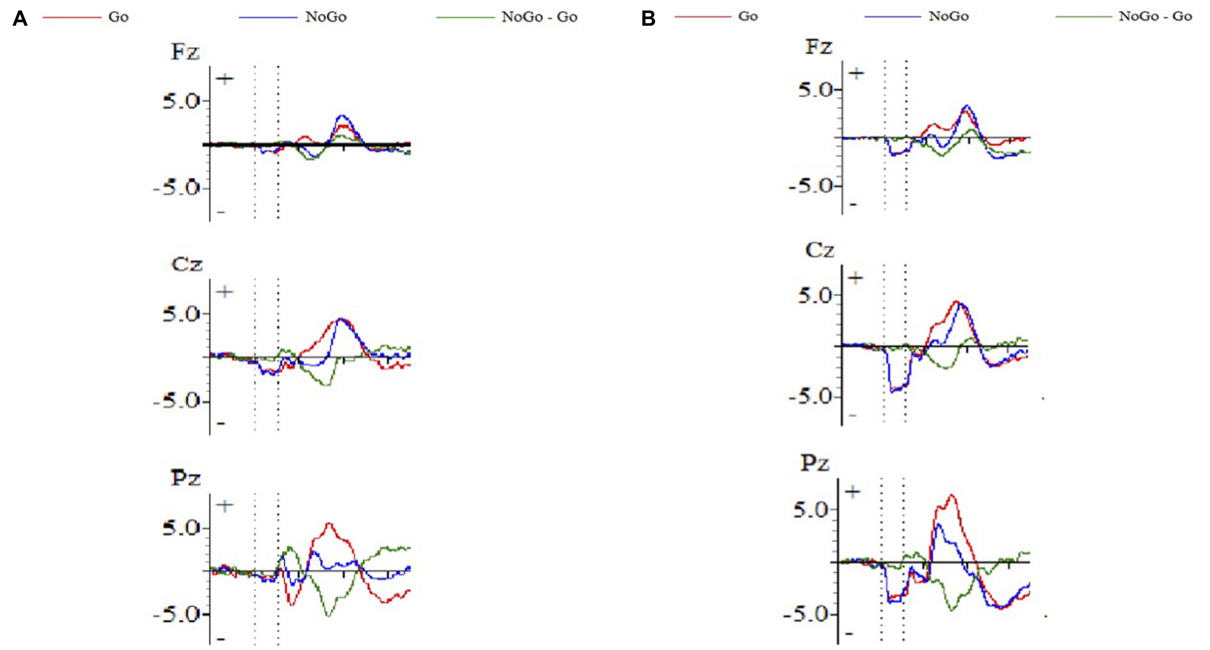


FIGURE 1
The ERP waveform diagram of Fz, Cz, and Pz. **(A)** High self-efficacy group; **(B)** low self-efficacy group. The red line represents the waveform result of the Go trials, and the blue line represents the waveform result of the No-Go trials.

to detect conflict between Go and No-Go response tendencies. Students with higher self-efficacy proved more sensitive to stimuli that presented a temptation to respond in accord with the prepotent action in an ongoing series of actions but in fact required a different response. Thus, it appears that higher self-efficacious students tend to regulate their behaviors in response to conflicts or challenges by "amping up" cortical activities that have become more efficient with the level of self-efficacy. According to Bandura (1977, 1986), high self-efficacy improves cognitive performance since it enhances on-task attention. This increased attention leads to greater focusing, which improves the ability to detect and monitor conflicts or irrelevant stimuli (Farbiash and Berger, 2015). High self-efficacy also enhances motivational aspects (Bandura, 1977, 1986), which facilitates conflict monitoring during a Go/No-Go task (Leue et al., 2012). A person's affective states and actions are closely associated with self-efficacy (Bandura, 1997). The finding of the present study is, to some extent, in line with a study conducted by Lewis et al. (2006) that found the N2 on No-Go trials was greater during conditions of negative emotion induction. They explained that children's performance was better when they were emotionally distressed, and the larger N2 amplitudes reflected higher effortful brain activation during the negative emotional experience (Lewis and Stieben, 2004). Negative emotion can result in improved inhibitory control and increases in associated aspects of brain activity (Farbiash and Berger, 2015). It seems that negative emotion, inhibitory control, and the greater N2 on No-Go trials might have some chain effects. The larger N2 amplitudes reflected higher effortful brain activation of students with high self-efficacy, which is also supported by Caraway et al. (2003) and they stated, "self-efficacy determines the aspect of task engagement including which tasks individuals choose to take on, the amount effort, persistence, and perseverance they demonstrate with regard to the task, and their feelings related to the task" (p. 423). Consistent with these points of view, in the present study, it may be the case that students' performance (detect interference and regulate their behaviors in response to conflicts) was better when they had a higher level of self-efficacy, and these students with enhanced motivation, increased attention, and higher effortful brain activation tended to view the execution of tasks as under their control. Lewis et al. (2006)'s study focused on children of 5–16 years of age and examined the effects of negative emotion on mechanisms of response inhibition, whereas in the present study, undergraduate and graduate students engaged in the experiments and inhibition mechanisms in relation to self-efficacy were examined. Previous studies mainly focused on children (e.g., Lewis et al., 2006; Farbiash and Berger, 2015), and few of them explored the effects of cognitive factors on college students' inhibition mechanisms from the perspective of cognitive neurology. This study provides new insights for related experimental research or theory development. Additionally, this finding is consistent with the study of Enriquez-Geppert et al. (2010), which suggested an association of the N2 with conflict-related effects with less frequently occurring trial types. In this study, No-Go is the less frequently occurring trial type, and Go stimuli were presented for most of the trials (80%). Frequent responses are prepotent, eventually leading to conflicts at the response representation

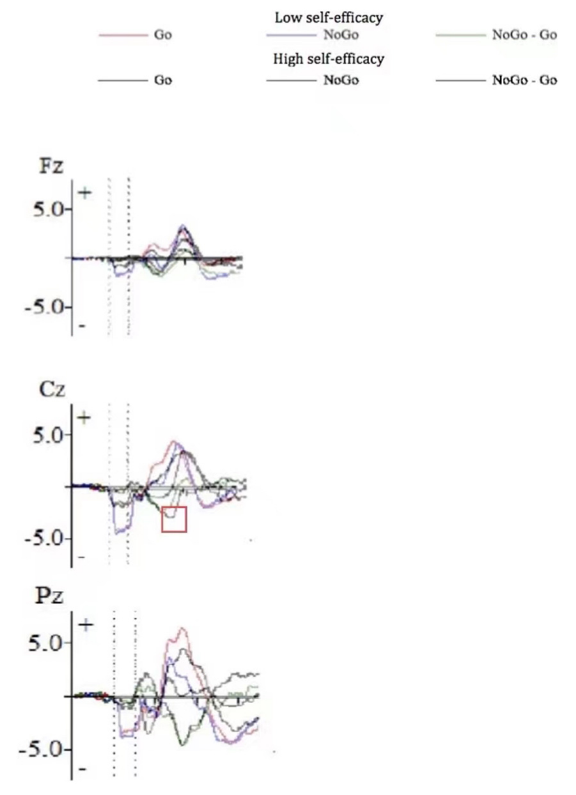


FIGURE 2
The ERP waveform diagram of Fz, Cz, and Pz for the difference between high self-efficacy group and low self-efficacy group.

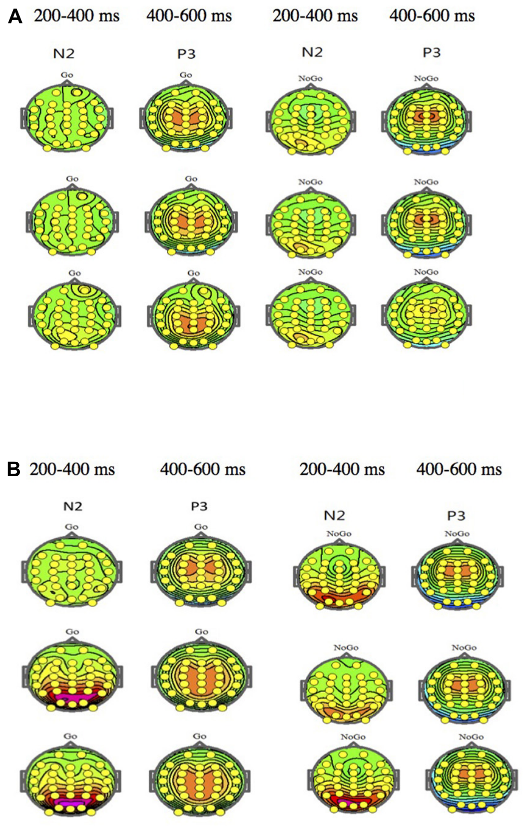


FIGURE 3
(A) The topographical map of the high self-efficacy group.
(B) The topographical map of the low self-efficacy group.

level when infrequent responses have to be made (Braver et al., 2001; Jones et al., 2002; Nieuwenhuis et al., 2003).

In contrast with previous studies, this study found there was no significant difference in No-Go P3 amplitudes between the high self-efficacy group and the low self-efficacy group. The P3 component has been used to ensure that stimulus analyses are appropriately linked with the correct behavioral actions in the monitoring processes; inhibition of the action operation is associated with the No-Go P3 (Verleger et al., 2005; Smith et al., 2008). The finding of this study indicates that conflict inhibition is not modulated by self-efficacy. It contradicts the findings of Wang (2018). Wang (2018) investigated the brain inhibitory effect of self-efficacy of college students to English vocabulary stimuli and found larger No-Go P3 amplitudes in the high self-efficacy group compared with the low self-efficacy group. A possible explanation is that Wang's study selected English biological and non-biological vocabulary as stimuli in the Go/No-Go tasks, and it is more difficult for students to identify these academic English words than the pictures used in the present study. The stimuli we used required less attention and the participants could easily withhold a dominant response to resist distraction. P3 is consciousness-dependent that is sensitive to cognitive demands during task processing such as task difficulty and the subjective probability of task stimuli or conditions (Kok, 2001; Hillman, 2004; Polich and Criado, 2006; Polich, 2007). P3 amplitude shows changes in the neural representation of the stimulus environment and reflects the allocation of task-relevant attentional control, with larger P3 amplitudes associated with the greater attentional allocation (Polich and Heine, 1996; Themanson and Rosen, 2014). Studies on attentional control implicate a network of brain regions, including the dorsolateral prefrontal cortex, parietal, and cingulate cortices (Bunge et al., 2002; Peterson et al., 2002; Durston et al., 2003; Weiss et al., 2003; Kerns et al., 2004; Langenecker et al., 2004). It is likely that due to low task difficulties for Go trials and generally reduced attention for No-Go trials, participants in the current study did not allocate much attentional control. Also, studies using different cues to induce varying levels of response preparation consistently found larger P3 amplitudes when inhibition is made more demanding (e.g., Bruin et al., 2001, Smith et al., 2007). It is possible that the effects of self-efficacy may be less powerful due to a possible increase in efficacy or confidence of participants (especially students with low self-efficacy) in their capabilities for tasks because of task repetition. The tendency of "conflict" tasks was not so large that the temptation to stop (as it is in the Go/No-Go task) called for less attentional resources or required less interference control and fewer additional cognitive loads. It also provides support for the need of exploring the mediation effects of task difficulties in future studies. Rosen (2010) and Themanson and Rosen (2014) also found students with greater self-efficacy showed larger P3 amplitudes. But they adopted a different task—a flanker task to examine the relationships between self-efficacy and neural indices of stimulus processing, task performance, and task-relevant attentional control. Their explorations only reflected part of the inhibition-related cognitive process. The elicitation and generation of the P3 component is a constant and ongoing process, and when we compare findings of different studies, a variety of associated individual difference factors have to be considered such as age, sex, intelligence, and personality (e.g., Stelmack and Houlihan, 1994; Polich, 1996). Furthermore, our findings contribute to the understanding of the social cognitive theory that explains in detail self-efficacy as a positive influence on cognitive processes (Bandura, 1993; Bandura et al., 1996; Kim, 2009; Diseth, 2011; Yusuf, 2011). Higher-order cognitive processes include not only inhibitory control but also cognitive flexibility and the ability to plan, monitor, and carry out goal-directed actions (Schacht et al., 2009). When we discuss the modulation of self-efficacy on inhibition, various associated factors could be considered.

Conclusion

This study analyzed neural correlates of inhibition modulated by self-efficacy. It was found that students with a higher self-efficacy level tended to have a stronger ability to detect a conflict. Conflict detection was modulated by self-efficacy, whereas conflict inhibition was not modulated by self-efficacy. The high self-efficacious students were more likely to be capable of detecting conflicts but not necessarily followed by inhibition of the action operation.

In classrooms, teachers could design meaningful or student-relevant activities to increase their self-efficacy to regulate their behaviors in response to conflicts or challenges, and thus enhance motivational aspects. Teachers also could emphasize the significance of positive reinforcement and a supportive environment, and teach students problem-solving and information processing skills to encourage students to persist longer in learning tasks and engage students actively in class work.

This study explored neural correlates of inhibition modulated by self-efficacy based on Go/No-Go task monitoring and processing, but it did not reflect the whole process of inhibition and also focused on a small sample of students. Future research could implement multi-task or more complex measures to assess inhibition modulated by self-efficacy with a larger sample size. We also need to know more about how other predictors contribute further to understanding inhibition behavior change.

To sum up, the findings of this study support previous reports that inhibition in response to a stimulus is associated with self-efficacy from the perspective of cognitive neurology, and provide evidence that inhibition, indexed by N2 amplitude, may be one mechanism through which self-efficacy improves task performance. The analysis of ERP components serves to yield a more complete picture of the cognitive mechanism underlying inhibition modulated by self-efficacy. This study provides a new perspective for studies on inhibition and self-efficacy, and it will contribute to our understanding of cognitive ability development.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by China University of Petroleum-Beijing. The patients/participants provided their written informed consent to participate in this study.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This work was supported by the Science Foundation of China University of Petroleum, Beijing (No. ZX20170005).

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Albert, J., López-Martín, S., Hinojosa, J., and Carretié, L. (2013). Spatiotemporal characterization of response inhibition. *NeuroImage* 76, 272–281. doi: 10.1016/j.neuroimage.2013.03.011

Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., and Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. *Developmental Psychology* 50, 2368–2379. doi: 10.1037/a0037493

Bandura, A. (1977). Self efficacy: Toward a unifying theory of behavioral change. *Psychological Review* 84, 191–215. doi: 10.1037/0033-295X.84.2.191

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology* 4, 359–373. doi: 10.1521/jscp.1986.4.3.359

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist* 28, 117–148. doi: 10.1207/s15326985ep2802_3

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Worth Publishers.

Bandura, A., Barbaranelli, C., and Pastorelli, C. C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development* 67, 1206–1222. doi: 10.2307/1131888

Bembenutty, H. (2011). Introduction: self-regulation of learning in postsecondary education. *New Directions for Teaching & Learning* 126, 3–8. doi: 10.1002/tl.439

Bokura, H., Yamaguchi, S., and Kobayashi, S. (2001). Electrophysiological correlates for response inhibition in a Go/NoGo task. *Clinical Neurophysiology* 112, 2224–2232. doi: 10.1016/S1388-2457(01)00691-5

Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L., and Avraham, S. (2001). Anterior cingulate cortex and response conflict: Effects of frequency, inhibition and errors. *Cerebral Cortex* 11, 825–836. doi: 10.1093/cercor/11.9.825

Bruin, K. J., Wijers, A. A., and Staveren, A. (2001). Response priming in a go/nogo task: Do we have to explain the go/nogo N2 effect in terms of response activation instead of inhibition? *Clinical Neurophysiology* 112, 1660–1671. doi: 10.1016/S1388-2457(01)00601-0

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., and Gabrieli, J. D. (2002). Immature frontal lobe contributions to cognitive control in children: Evidence from fMRI. *Neuron* 33, 301–311. doi: 10.1016/S0896-6273(01)00583-9

Caraway, K., Tucker, C. M., Reinke, W. M., and Hall, C. (2003). Self-efficacy, goal orientation, and fear of failure as predictors of school engagement in high school students. *Psychology in the Schools* 40, 417–427. doi: 10.1002/pits.10092

Ciesielski, K. T., Harris, R. J., and Cofer, L. F. (2004). Posterior brain ERP patterns related to the go/No-Go task in children. *Psychophysiology* 41, 882–892. doi: 10.1111/j.1469-8986.2004.00250.x

Cragg, L., and Nation, K. (2008). Go or No-Go? Developmental improvements in the efficiency of response inhibition in mid-childhood. *Developmental Science* 11, 819–827. doi: 10.1111/j.1467-7687.2008.00730.x

Dekker, M. C., Ziermans, T. B., Spruijt, A. M., and Hanna, S. (2017). Cognitive, parent and teacher rating measures of executive functioning: shared and unique influences on school achievement. *Frontiers in Psychology* 8:48. doi: 10.3389/fpsyg.2017.00048

Dimoska, A., Johnstone, S. J., and Barry, R. J. (2006). The auditory-evoked N2 and P3 components in the stop-signal task: Indices of inhibition, response-conflict or error-detection? *Brain & Cognition* 62, 98–112. doi: 10.1016/j.bandc.2006.03.011

Diseth, A. (2011). Self-efficacy, goal orientation and learning strategies as mediators between preceding and subsequent academic achievement. *Learning and Individual Differences* 21, 191–195. doi: 10.1016/j.lindif.2011.01.003

Durston, S., Davidson, M. C., Thomas, K. M., Worden, M. S., Tottenham, N., Martinez, A., et al. (2003). Parametric manipulation of conflict and response competition using rapid mixed-trial event-related fMRI. *Neuroimage* 20, 2135–2141. doi: 10.1016/j.neuroimage.2003.08.004

Enriquez-Geppert, S., Huster, R., Scharfenort, R., Mokom, Z., Zimmermann, J., and Herrmann, C. (2014). Modulation of frontal-midline theta by neurofeedback. *Biological Psychology* 95, 59–69. doi: 10.1016/j.biopsycho.2013.02.019

Enriquez-Geppert, S., Konrad, C., Pantev, C., and Huster, R. J. (2010). Conflict and inhibition differentially affect the N200/P300 complex in a combined go/nogo and stop-signal task. *Neuroimage* 51, 877–887. doi: 10.1016/j.neuroimage.2010.02.043

Farbiash, T., and Berger, A. (2015). Brain and behavioral inhibitory control of kindergartners facing negative emotions. *Developmental Science* 1–16. **v, doi: 10.1111/desc.12330

Gagne, J. R. (2017). Self-control in childhood: A synthesis of perspectives and focus on early development. *Child Development Perspectives* 11, 127–132. doi: 10.1111/cdep.12223

Gärtner, K. A., Vetter, V. C., Schäferling, Michaela, Reuner, G., and Hertel, S. (2018). Inhibitory control in toddlerhood – the role of parental co-regulation and self-efficacy beliefs. *Metacognition and Learning* 13, 241–264. doi: 10.1007/s11409-018-9184-7

Hillman, C. H. (2004). P3-ERP indices of physical activity effects on cognitive function. *Psychophysiology* 41, S18–S19. doi: 10.1111/j.1469-8986.2004.symp_15.x

Hoyniak, C. (2017). Changes in the NoGo N2 event-related potential component across childhood: A systematic review and meta-analysis. *Developmental Neuropsychology* 42, 1–24. doi: 10.1080/87565641.2016.1247162

Hum, K. M., Manassis, K., and Lewis, M. D. (2013). Neural mechanisms of emotion regulation in childhood anxiety. *Journal of Child Psychology and Psychiatry* 54, 552–564. doi: 10.1111/j.1469-7610.2012.02609.x

Huster, R. J., Enriquez-Geppert, S., Lavallee, C. F., Falkenstein, M., and Herrmann, C. S. (2013). Electroencephalography of response inhibition tasks: Functional networks and cognitive contributions. *International Journal of Psychophysiology* 87, 217–233. doi: 10.1016/j.ijpsycho.2012.08.001

Jodo, E., and Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a Go/No-Go task. *Electroencephalography and Clinical Neurophysiology* 82, 477–482. doi: 10.1016/0013-4694(92)90054-L

Jones, A. D., Cho, R. Y., Nystrom, L. E., Cohen, J. D., and Braver, T. S. (2002). A computational model of anterior cingulate function in speeded response tasks: Effects of frequency, sequence, and conflict. *Cognitive Affective & Behavioral Neuroscience* 2, 300–317. doi: 10.3758/CABN.2.4.300

Kerns, J. G., Cohen, J. D., MacDonald, A. W. I. I., Cho, R. Y., Stenger, V. A., and Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science* 303, 1023–1026. doi: 10.1126/science.1089910

Kim, E. (2009). Beyond language barriers: Teaching self-efficacy among East Asian international teaching assistants. *International Journal of Teaching and Learning in Higher Education* 21, 171–180. doi: 10.1037/e629232009-001

Kloo, D., and Sodian, B. (2017). The developmental stability of inhibition from 2 to 5 years. *British Journal of Developmental Psychology* 35, 582–595. doi: 10.1111/bjdp.12197

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology* 38, 557–577. doi: 10.1017/S0048577201990559

Lahat, A., Todd, R., Mahy, C. E., Lau, K., and Zelazo, P. D. (2010). Neurophysiological correlates of executive function: A comparison of European-Canadian and Chinese-Canadian 5-year-olds. *Frontiers in Human Neuroscience* 3:72. doi: 10.3389/neuro.09.072.2009

Langenecker, S. A., Nielson, K. A., and Rao, S. M. (2004). fMRI of healthy older adults during Stroop interference. *Neuroimage* 21, 192–200. doi: 10.1016/j.neuroimage.2003.08.027

Leue, A., Lange, S., and Beauducel, A. (2012). Modulation of the conflict monitoring intensity: The role of aversive reinforcement, cognitive demand, and trait-BIS. *Cognitive, Affective, & Behavioral Neuroscience* 12, 287–307. doi: 10.3758/s13415-012-0086-x

Lewis, M. D., Lamm, C., Segalowitz, S. J., Stieben, J., and Zelazo, P. D. (2006). Neurophysiological correlates of emotion regulation in children and adolescents. *Journal of Cognitive Neuroscience* 18, 430–443. doi: 10.1162/jocn.2006.18.3.430

Lewis, M. D., and Stieben, J. (2004). Emotion regulation in the brain: Conceptual issues and directions for developmental research. *Child Development* 75, 371–376. doi: 10.1111/j.1467-8624.2004.00680.x

Liew, J. (2012). Effortful control, executive functions, and education: Bringing self-regulatory and social-emotional competencies to the table. *Child Development Perspectives* 6, 105–111. doi: 10.1111/j.1750-8606.2011.00196.x

Litkowski, E. C., Finders, J. K., Borriello, G. A., Schmitt, S. A., and Purpura, D. J. (2020). Patterns of heterogeneity in kindergarten children's executive function: profile associations with third grade achievement. *Learning and Individual Differences* 80, 101–146. doi: 10.1016/j.lindif.2020.101846

Mcauley, E., Mullen, S. P., Szabo, A. N., White, S. M., Thomas, R. W., Mailey, E. L., et al. (2011). Self-regulatory processes and exercise adherence in older adults: executive function and self-efficacy effects. *American Journal of Preventive Medicine* 41, 284–290. doi: 10.1016/j.amepre.2011.04.014

Morgan, P., Farkas, G., Hillemeier, M. M., Pun, W. H., and Maczuga, S. (2019). Kindergarten children's executive functions predict their second-grade academic achievement and behavior. *Child Development* 90, 1802–1816. doi: 10.1111/cdev.13095

Nieuwenhuis, S., Yeung, N., Wildenberg, W., and Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive Affective & Behavioral Neuroscience* 3, 17–26. doi: 10.3758/CABN.3.1.17

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin* 126, 220–246. doi: 10.1037/0033-2909.126.2.220

Peterson, B. S., Kane, M. J., Alexander, G. M., Lacadie, C., Skudlarski, P., Leung, H. C., et al. (2002). An event-related functional MRI study comparing interference effects in the Simon and Stroop tasks. *Cognitive Brain Research* 13, 427–440. doi: 10.1016/S0926-6410(02)00054-X

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology* 95, 667–686. doi: 10.1037/0022-0663.95.4.667

Pintrich, P. R., Smith, D. A., Garcia, T., and Mckeachie, W. J. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. Ann Arbor, MI: University of Michigan. Retrieved from ERIC database. (ED338122).

Pintrich, P. R., Smith, D. A., Garcia, T., and Mckeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement* 53, 801–813. doi: 10.1177/0013164493053003024

Polich, J. (1996). Meta-analysis of P300 normative aging studies. *Psychophysiology* 33, 334–353. doi: 10.1111/j.1469-8986.1996.tb01058.x

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Polich, J., and Criado, J. R. (2006). Neuropsychology and neuropharmacology of P3a and P3b. *International Journal of Psychophysiology* 60, 172–185. doi: 10.1016/j.ijpsycho.2005.12.012

Polich, J., and Heine, M. R. D. (1996). P300 topography and modality effects from a single stimulus paradigm. *Psychophysiology* 33, 747–752. doi: 10.1111/j.1469-8986.1996.tb02371.x

Ramautar, J. R., Kok, A., and Ridderinkhof, K. R. (2004). Effects of stop-signal probability in the stop-signal paradigm: The N2/P3 complex further validated. *Brain and Cognition* 56, 234–252. doi: 10.1016/j.bandc.2004.07.002

Randall, W. M., and Smith, J. L. (2011). Conflict and inhibition in the cued-Go/NoGo task. *Clinical Neurophysiology* 122, 2400–2407. doi: 10.1016/j.clinph.2011.05.012

Rosen, P. J. (2010). Examining the relationship between self-efficacy and stimulus processing. *Honors Projects* 136. http://digitalcommons.iwu.edu/psych_honproj/136

Rueda, M. R., Posner, M. I., and Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology* 28, 573–594. doi: 10.1207/s15326942dn2802_2

Schacht, A., Nigbur, R., and Sommer, W. (2009). Emotions in go/nogo conflicts. *Psychological Research Prpf* 73, 843–856. doi: 10.1007/s00426-008-0192-0

Smith, J. L., Jamadar, S., Provost, A. L., and Michie, P. T. (2013). Motor and non-motor inhibition in the Go/NoGo task: An ERP and fMRI study. *International Journal of Psychophysiology* 87, 244–253. doi: 10.1016/j.ijpsycho.2012.07.185

Smith, J. L., Johnstone, S. J., and Barry, R. J. (2006). Effects of pre-stimulus processing on subsequent events in a warned Go/NoGo paradigm: Response preparation, execution and inhibition. *International Journal of Psychophysiology* 61, 121–133. doi: 10.1016/j.ijpsycho.2005.07.013

Smith, J. L., Johnstone, S. J., and Barry, R. J. (2007). Response priming in the Go/NoGo task: The N2 reflects neither inhibition nor conflict. *Clinical Neurophysiology* 118, 343–355. doi: 10.1016/j.clinph.2006.09.027

Smith, J. L., Johnstone, S. J., and Barry, R. J. (2008). Movement-related potentials in the Go/NoGo task: The P3 reflects both cognitive and motor inhibition. *Clinical Neurophysiology* 119, 704–714. doi: 10.1016/j.clinph.2007.11.042

Stelmack, R. M., and Houlihan, M. (1994). "Event related potentials, personality, and intelligence: Concepts, issues, and evidence," in *International handbook of personality and intelligence*, eds D. H. Saklofske and M. Zaidner (New York: Plenum Press), 349–365. doi: 10.1007/978-1-4757-5571-8_17

Themanson, J. R., Hillman, C. H., McAuley, E., Buck, S., Doerksen, S. E., Morris, K. S., et al. (2008). Self-efficacy effects on neuroelectric and behavioral indices of action monitoring in older adults. *Neurobiology of Aging* 29, 1111–1122. doi: 10.1016/j.neurobiolaging.2007.01.004

Themanson, J. R., and Rosen, P. J. (2014). Examining the relationships between self-efficacy, task-relevant attentional control, and task performance: Evidence from event-related brain potentials. *British Journal of Psychology* 106, 1–39. doi: 10.1111/bjop.12091

Verleger, R., Jaskowski, P., and Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology* 19, 165–181. doi: 10.1027/0269-8803.19.3.165

Waldvogel, D., van Gelderen, P., Muellbacher, W., Ziemann, U., Immisch, I., and Hallett, M. (2000). The relative metabolic demand of inhibition and excitation. *Nature* 406, 995–998. doi: 10.1038/35023171

Wang, Q. (2018). Analysis and Brain Mechanism of English Learning Characteristics of College Students Considering Self-Efficacy Regulation. *NeuroQuantology* 16, 461–467. doi: 10.14704/nq.2018.16.5.1349

Weiss, E. M., Golaszewski, S., and Mottaghy, F. M. (2003). Brain activation patterns during a selective attention test-a functional MRI study in healthy volunteers and patients with schizophrenia. *Psychiatry Res* 123, 1–15. doi: 10.1016/S0925-4927(03)00019-2

Wiebe, S. A., Sheffield, T. D., and Espy, K. A. (2012). Separating the fish from the sharks: A longitudinal study of preschool response inhibition. *Child Development* 83, 1245–1261. doi: 10.1111/j.1467-8624.2012.01765.x

Yusuf, M. (2011). The impact of self-efficacy, achievement motivation, and self-regulated learning strategies on students' academic achievement. *Procedia Social and Behavioral Sciences* 15, 2623–2626. doi: 10.1016/j.sbspro.2011.04.158

Zhao, D., Lin, H., Xie, S., and Liu, Z. (2019). Emotional arousal elicited by irrelevant stimuli affects event-related potentials (ERPs) during response inhibition. *Physiol. Behav.* 206, 134–142. doi: 10.1016/j.physbeh.2019.04.005

Appendix A: Self-efficacy survey

Please first answer the following questions about yourself. Your answers will be treated in a confidential manner and only identified to the researcher for this study.

1. Sex: ____________
2. Age: ____________
3. Education level: _____

The following questions ask about your self-efficacy (your self-perceptions or beliefs of capability to learn or perform tasks in classes). Answer in terms of how well the statement describes you. This usually takes about 5 min to complete. If you have any questions, let the researcher know immediately.

Please read each statement and check the box that best describes how you feel:

1 = Not at all true of me to 9 = Very true of me.

	Not at all true of me 1	2	3	4	5	6	7	8	Very true of me 9
I believe I will receive excellent grades in classes.									
I'm certain I can understand the most difficult material presented for classes.									
I'm confident I can understand the basic concepts taught in classes.									
I'm confident I can understand the most complex material presented by instructors in classes.									
I'm confident I can do excellent jobs on the assignments and tests in classes.									
I expect to do well in classes.									
I'm certain I can master the skills being taught in classes.									
Considering the difficulties of classes, teachers, and my skills, I think I will do well in classes.									

Appendix B

APPENDIX TABLE 1 Demographic characteristics of participants.

Self-efficacy	High ($N = 24$)	Low ($N = 26$)	Total ($N = 50$)
Sex			
Female	12	14	26
Male	12	12	24
Age			
19-21	11	11	22
22-24	14	14	28
Education Level			
Undergraduate	11	12	23
Graduate	13	14	27

APPENDIX TABLE 2 Mean accuracy for all the experimental conditions.

	Trial type	High self-efficacy	Low self-efficacy
Accuracy	No-Go	0.82	0.80
	Go	0.96	0.96
ResponseTimes(ms)	Go	410.23	412.79

APPENDIX TABLE 3 Analysis of variance (ANOVA) effects for N2 amplitude.

	df	F	*p*	partial η^2
site	2,96	6.21	0.003*	0.11
site*self-efficacy	2,96	2.26	0.11	0.05
go/nogo	1,48	4.48	0.04*	0.09
go/nogo*self-efficacy	1,48	4.76	0.03*	0.09
site*go/nogo	2,96	0.97	0.38	0.02
site*go/nogo*self-efficacy	2,96	1.61	0.21	0.03
self-efficacy	1,48	7.12	0.01*	0.13

*$p < 0.05$.

APPENDIX TABLE 4 Analysis of variance (ANOVA) effects for P3 amplitude.

	df	F	*p*	partial η^2
site	2,96	6.74	0.002*	0.12
site*self-efficacy	2,96	0.30	0.75	0.01
go/nogo	1,48	2.19	0.15	0.04
go/nogo*self-efficacy	1,48	3.21	0.08	0.06
site*go/nogo	2,96	18.77	0.00*	0.28
site*go/nogo*self-efficacy	2,96	1.42	0.25	0.03
self-efficacy	1,48	0.03	0.86	0.00

*$p < 0.05$.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 05 October 2022
DOI 10.3389/fpsyg.2022.892118

EDITED BY
Edwin James Burns,
Edge Hill University,
United Kingdom

REVIEWED BY
David R. Labbe,
École de technologie supérieure (ÉTS),
Canada
Abdolhamid Daneshjoo,
Shahid Bahonar University of Kerman, Iran

*CORRESPONDENCE
Aming Lu
luamingsu@126.com

SPECIALTY SECTION
This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

RECEIVED 08 March 2022
ACCEPTED 15 August 2022
PUBLISHED 05 October 2022

CITATION
Ren Y, Wang C and Lu A (2022) Effects of perceptual-cognitive tasks on inter-joint coordination of soccer players and ordinary college students.
Front. Psychol. 13:892118.
doi: 10.3389/fpsyg.2022.892118

Effects of perceptual-cognitive tasks on inter-joint coordination of soccer players and ordinary college students

Yuanyuan Ren, Cenyi Wang and Aming Lu*

School of Physical Education and Sport Science, Soochow University, Suzhou, China

Perceptual-cognitive tasks play a pivotal role in performing voluntary movements, which is crucial for good performances among soccer players. This study explored the effect of perceptual-cognitive tasks on the inter-joint coordination of soccer players and college students during landing. The classic multiple objective tracking (MOT) task was used to simulate the perceptual-cognitive task under a sports environment. Fifteen soccer players (age: 20.1 ± 1.5 year, height: 181.4 ± 7.4 cm, weight: 75.4 ± 10.7 kg) and twenty ordinary college students (age: 20.0 ± 2.3 years, height: 177.9 ± 4.9 cm, weight: 71.6 ± 9.9 kg) were enrolled to the study. Participants in the two groups were subjected to a single task (landing task) and dual-task (MOT task and landing task). Coordination and variability indicators were recorded using a Vicon infrared motion capture system and a force measuring platform. The results showed that the mean absolute relative phase of hip and knee joint ($MARP_{hip\text{-}knee}$), deviation phase of hip and knee joint ($DP_{hip\text{-}knee}$), and deviation phase of knee and ankle joint ($DP_{knee\text{-}ankle}$) of the two groups under the dual-task were significantly different compared with the parameters when participants were subjected to the single task. The dual-task had higher effect size on $DP_{hip\text{-}Knee}$ and $MARP_{hip\text{-}knee}$, indicating that dual-task had a greater impact on coordination of the hip and knee joints. $DP_{hip\text{-}knee}$ and $DP_{knee\text{-}ankle}$ of ordinary students were more extensive relative to those of the soccer players, and hip joint stiffness (K_{hip}) for ordinary students was lower than that of the soccer players under the different tasks. These findings implied that the perceptual-cognitive task markedly affected the inter-joint coordination of soccer players and college students, mainly by impairing the hip and knee coordination. Although there is less variability in lower extremity coordination patterns of soccer players compared to college students, the MOT task still affects their coordination ability.

KEYWORDS

cognition, inter-joint coordination, performance, dual-tasks, landing

Introduction

Perceptual-cognitive abilities in interactive sports are correlated with the visual attention features and the dynamic tracking tasks conducted in sports. Soccer players are required to focus on the position of the ball, teammate and the opponent concurrently high perceptual-cognitive abilities are important. The ability to divide attention and dynamically track multiple moving objects is known as Multiple Object Tracking (MOT) task. It plays a crucial role in performing voluntary movements, and highly contributes to success in team sports. MOT tasks can be used for effective evaluation and simulation of perceptual-cognitive tasks in dynamic environments of athletics (Ehmann et al., 2021).

Several laboratory studies have reproduced such dynamic situations using the MOT task paradigm, which is the main approach used for studying interactions between cognitive processing and motor behavior. The biomechanical effects of 3D-MOT on lower extremity landings have been extensively explored in recent studies, and the findings indicated that combination of perceptive-cognitive task with muscle fatigue alters knee kinematics and increases the risk of anterior cruciate ligament rupture (Mejane et al., 2019). MOT tasks can reduce the postural stability of soccer players, as well as increase the risk of injury of soccer players (Ren et al., 2021). Findings from a previous research showed that soccer players had a 14.4% higher injury rate during matches compared with the rate of injury observed during regular training (Romeas et al., 2016), which was partially attributed to MOT tasks undertaken during matches. Differences in attention allocation and perceptual-cognitive ability between soccer players and ordinary students have been observed in the past, and various results indicate that soccer players are proficient in cognitive tasks (Alves et al., 2013; Qiu et al., 2018). A study was conducted to compare the cognitive abilities of elite athletes, ordinary athletes, and healthy adults. The results showed that elite athletes had better performance in two control tasks and a visuospatial attention task compared with ordinary athletes and healthy adults (Alves et al., 2013). In addition, previous findings indicated that high-performance athletes who had been trained for a long time had better performance after subjection to tracking multiple targets with different attention loads relative to the performance of non-athletes (Qiu et al., 2018). The effect of perceptual-cognitive training on improving cognitive function and specific motor performance in humans has been explored in recent studies. Long-term MOT training improves athletic performance and enhances the winning ability during soccer matches (Romeas et al., 2016). Participants in a video-based perception training group in a study conducted previously showed improved decision-making, fewer recall errors, improved shooting, and other specialized skills after 4 weeks of intervention compared with that of the non-intervention control group (Gabbett et al., 2008). This indicates that resource allocation processes involved in complex motor environments and visual attention overlap in brain regions, and long-term training causes plasticity changes in brain activities (Dahlin et al., 2008).

Soccer is a popular and complex sport that requires perceptual-cognitive activities such as visual attention (Alvarez and Cavanagh, 2005). High requirements for coordination between limbs, good limb coordination, and stability are highly associated with excellent performance of soccer players (Zhang et al., 2021). Inter-joint coordination mode of the lower extremity is a major indicator for evaluating lower extremity load and sports injury during dynamic movement (Schoner and Kelso, 1988; Kim et al., 2019; Sinsurin et al., 2020; Zhang et al., 2021). Joint coordination during landing is mainly reflected through joint stiffness, and joint stiffness can be used to evaluate soft tissue injury in the lower extremity (Li et al., 2021). These findings imply that inter-joint significantly affects the performance of players. Coordinated activities can reduce the load on joints during movement by improving dynamic stability of the participant, resulting in efficient and accurate functional movement. The landing movement, which is a common movement type in sports, is a frequent movement in soccer activities such as heading of a soccer ball. Approximately 60% of injuries are associated with landing movements (Rahnama et al., 2002), thus landing movement is used to explore effective control strategies for preventing injuries. Studies conducted in the past reported possible interaction between perceptual-cognitive tasks and motor control, however, the effects of perceptual-cognitive tasks on inter-joint coordination and injury risk have not been fully elucidated. In addition, studies have not fully explored whether dual tasks (especially MOT tasks) affect the entire lower extremity, and possible differences in motor control and regulation modes in soccer players and college students should be elucidated.

Therefore, the purpose of this study was to evaluate the effects of MOT tasks interference on inter-joint coordination during landing. In addition, the motor control of soccer players and ordinary college students under different tasks was explored by comparing the effects of MOT tasks on the lower limb coordination patterns of the participants. The hypothesis formulated for the study was that attention processes involved in MOT tasks are important factors resulting in an increase in individual impairment and causes changes in lower extremity coordination mode. Furthermore, it was hypothesized that well-trained soccer players would exhibit better inter-joint coordination under interference of different tasks compared with ordinary college students who had not undergone long-term training. This study provides an important reference for further studies to design strategies for preventing injuries in individuals involved in sports.

Materials and methods

Participants

A total of 35 male participants from Soochow University were recruited to the study through posters. The participants included 15 well-trained players from a high-level soccer team in the university and 20 college students who had no experience participating in major athletic competitions or training. The inclusion criteria for the study were as follows: (1) Participants who had no history of diseases or injuries affecting the lower extremity joints within approximately half a year and were able to complete the experiment; (2) Subjects who had not undertaken strenuous exercise within 24 h before the test. Personal basic information was obtained using a questionnaire and participants who did not meet the inclusion criteria were excluded from the study (Table 1).

Materials and study equipment

A 3D projector and a screen were placed directly in front of the participants who were requested to perform perceptual-cognitive tasks. MOT tasks were presented as self-written programs, which were projected to a 220 cm screen using the projector. The distance between the participant and the screen was 130 cm (for MOT and landing tasks). Landing tests were performed on a force measuring platform using a large pressure sensor (Kistler company, Switzerland, Model 9287B) at a sampling rate of 1,000 Hz, to allow online detection of ground reaction force and determination of balance characteristics. Vicon infrared motion capture system with 8 infrared cameras (Vicon, Inc., Oxford, United Kingdom) with a sampling rate of 100 HZ and a Vicon plug-in gait model were used to obtain kinematic data on the sagittal plane. The model markers were placed at the following positions: bilateral anterior superior iliac spines, bilateral posterior superior iliac spines, bilateral proximal thigh (1/3 of the distance from the hip to the knee), bilateral distal thigh (2/3 of the distance from the hip to the knee), mid-lateral thigh (at the middle of the lateral thigh from the hip to the knee), lateral condyles of both knees, bilateral proximal calf (1/3 of the distance from the knee to the ankle), bilateral distal calf (2/3 of the distance from the knee to the ankle), the lateral middle of the calf (at the middle of the lateral calf from the knee to the ankle), bilateral malleolus, bilateral calcaneus, and bilateral second phalanges (Figure 1). All the pasting work of model markers was only conducted by a trained professional to reduce experimental error. Data were collected from 2019 to 2021.

TABLE 1 Basic information of the study participants (mean±SD).

Information	Players	Students	*t*-value	Value of *p*
Age (yr)	20.1 ± 1.5	20.0 ± 2.3	0.09	0.93
Height (cm)	181.4 ± 7.4	177.9 ± 4.9	1.57	0.13
Weight (kg)	75.4 ± 10.7	71.6 ± 9.9	0.98	0.34

yr, year; cm, centimeter; kg, kilogram.

FIGURE 1
Marker sticking positions.

Study design

All participants signed an informed consent form and provided details on personal information by filling a questionnaire. Subsequently, participants were briefed on the experiment process and action, and positions were labeled on their lower limbs. The initial maximum tracking speed of MOT tasks was then determined for each participant and the standardized speed was recorded. Further, participants warmed up by walking at 4 km/h for 5–10 min on a treadmill. Participants were then randomly assigned to single task (ST) or dual-task (DT) trials. The experimental protocol comprised ST and DT dimensions. Participants completed the landing movement in ST and completed synchronously the MOT task and landing movement in DT. Each participant followed the experiment procedure (Figure 2).

Conducting of MOT and landing tasks

Eight identical spheres were adopted in conducting MOT tasks (Figure 3). These spheres interacted dynamically in a three-dimensional space (bounced off each other, broke away from the virtual three-dimensional volume boundary, and blocked each other). The task was divided into four steps as follows: (a) Eight randomly organized identical spheres were presented on the screen; (b) The color of three target spheres changed; (c): The color of all the spheres changed to the original color and the spheres moved randomly around the screen at the same speed; (d): The participants were requested to point out the position of the three target spheres on the screen independently at the end of the task. The moving speed of target spheres were standardized according to the participant's threshold to ensure that all participants received the same MOT task load (Mejane et al., 2019). The standardization process was conducted as follows: The task was first presented at a random speed and the speed of the next MOT task was increased when the participant correctly identified three target spheres. The speed of the subsequent MOT task was decreased if the three target spheres were not correctly identified. The speed of spheres was gradually adjusted until the participant's initial maximum tracking speed was achieved. A standardized speed corresponding to 30% of the initial maximum tracking speed was chosen to ensure that all participants were able to perform dual tasks (Ren et al., 2021) and to minimize differences in tracking speed among participants. Movement of the MOT task on the screen was used as a cue for landing movement and was applied to ensure that both tasks were completed simultaneously. A failure was recorded when participants focused only on one of the two tasks and finished the tasks sequentially thus the test was repeated. Participants signed an informed consent form prior to taking part in the experiment, and each participant was requested to familiarize themselves with the MOT task and fully understand it (a correct rate of 90% or more) in advance to minimize learning effects.

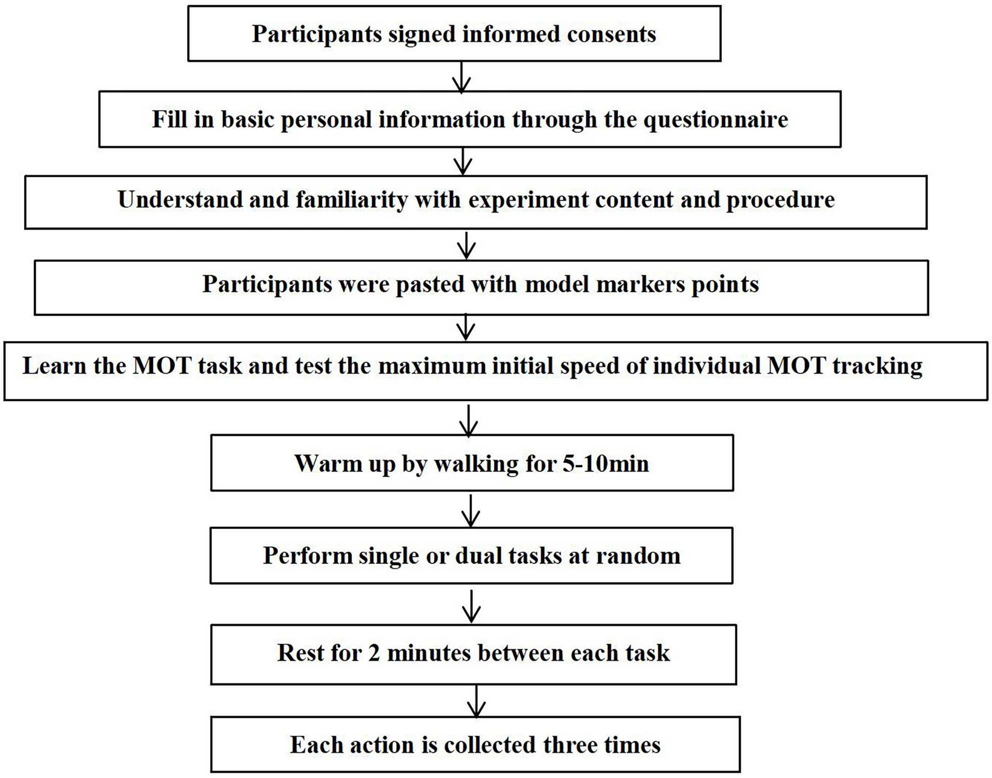


FIGURE 2
Schematic representation of the study design.

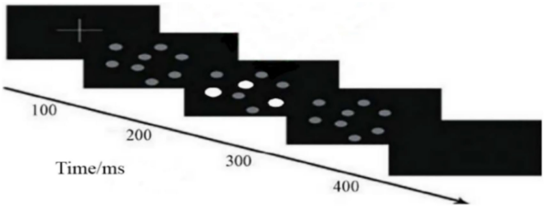


FIGURE 3
A representation of the multiple objective tracking (MOT) task experimental process.

In this study, each participant was requested to stand on a 40 cm platform with his hands on his hips and feet positioned upright. The participant used his dominant leg to step forward after receiving the "start" signal; subsequently, the participant leaned forward and fell vertically from the step without initial speed with the "toe-heel" landing mode. Each action was conducted in triplicates and the average value was calculated.

Data collection and analysis

The indicators of joint stiffness and coordination of the dominant lower extremity of the participants were analyzed in this study. Inter-joint coordination during landing mainly reflects joint stiffness, and reduction of joint stiffness is a potential risk factor for soft tissue injury (Li et al., 2021). Coordination-related parameters in this study included continuous relative phase (CRP), mean absolute relative phase (MARP), and deviation phase (DP), which were recorded using Vicon infrared motion capture system during ST and DT. Inter-joint coordination was evaluated to determine essential timing and sequencing of the neuromuscular system control over biomechanical degrees of freedom. Variability of coordination indicated adaptability of the neuromuscular system control. The landing process was defined as the period from touchdown (the moment vertical ground reaction $\leq$10 N was achieved) to the time maximum knee flexion was attained (Zhang et al., 2021). The duration of the landing process was normalized to 0–100%, with 0% representing the beginning of the landing process and 100% representing the end of the landing process. An inhouse MATLAB script (MatlabR2018b, MathWorks, Natick, MA, United States) was used for calculation of the indicators in this study.

CRP was used to describe the inter-joint coordination's mode and variability. MARP and DP were used to describe coordination (in-phase or anti-phase coupling) and variability between joints. The angle (θ) and angular velocity (ω) were normalized to minimize the effect of different movements in amplitudes and frequencies using formula (1) presented below. In the current study, the maximum angle of the sagittal plane of the lower extremity joints was normalized as "+2," the minimum angle was normalized as “0,” and the angular velocity (ω) was normalized to the maximum value (Sheikhi et al., 2021).

$$\omega_{(i)norm} = \omega / \max\left[\max(\omega_i), \max(-\omega_i)\right] \tag{1}$$

$$\theta_{(i)norm} = 2 * \left[\theta_j - \min(\theta_i)\right] / \max(\theta_i) - \min(\theta_i) \tag{2}$$

The relative phase angle (φ) of the different joints was calculated using formula (2) and standardized according to the phase diagram of the four quadrants (Figure 4). CRP values range from −180° to 180°, where a CRP close to 0° indicates that the two joints move “in-phase,” whereas a CRP close to 180° or −180° represents an “anti-phase” motion of the two joints (Sheikhi et al., 2021).

CRP values were calculated based on the difference between the distal joint and the proximal joint using formula (3) presented below. A positive value of CRP indicated that the proximal joint dominated the distal joint, whereas a negative value of CRP indicated that the distal joint dominated the proximal joint. MARP represents the mean value of each joint’s phase angles on the curve and is calculated using formula (4), and DP indicates the standard deviation of each point on the overall curve and is calculated using formula (5).

$$\text{CRP} = \varphi_{\text{distal}} - \varphi_{\text{proximal}} \tag{3}$$

$$\text{MARP} = \frac{\sum_{i=1}^{n} |\varphi|_i}{n} \tag{4}$$

$$\text{DP} = \frac{\sum_{i=1}^{n} SD_i}{n} \tag{5}$$

The joint stiffness (K_{joint}) was then calculated based on the net torque change in the sagittal plane (ΔM_{joint}) and displacement change of joint angle in the sagittal plane ($\Delta\theta_{\text{joint}}$) during landing (Farley et al., 1998). The formula for calculation of joint stiffness is presented below:

$$K_{\text{joint}} = \Delta M_{\text{joint}} / \Delta\theta_{\text{joint}} \tag{6}$$

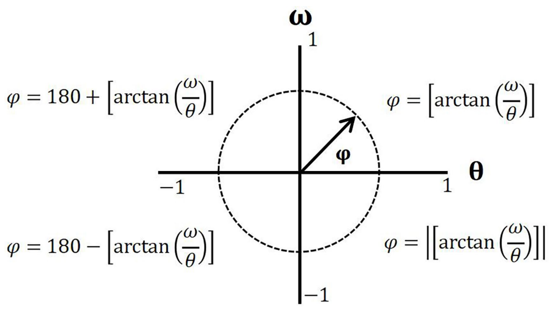


FIGURE 4
Standardization of the four-quadrant phase diagram.

Statistical analysis

SPSS statistics software (22.0, IBM Inc., Chicago, IL, United States) was used for statistical analysis. Shapiro–Wilk test was used to determine the normality of the data. Levene test was conducted to evaluate homogeneity of variance, and the studentized residuals method was used to determine presence of outliers. Mean (M) and standard deviation (SD) of the parameters were calculated and mixed repeated measure ANOVA was used to determine differences among variables with the group as the between-group factor (amateur soccer players and ordinary college students) and task as the within-group factor. Significance was set at $p < 0.05$.

Results

Changes in speed of spheres

The results showed that the average speed of the spheres was significantly different between the two groups of participants (Table 2). The initial maximum movement speed and standardized movement speed of the soccer players were higher than that of ordinary students.

Changes in inter-joint coordination

The CRP change curves of hip, knee, and ankle joints of soccer players and ordinary college students under two task conditions are presented in Figure 5. The $\text{CRP}_{\text{hip-knee}}$ and $\text{CRP}_{\text{knee-ankle}}$ values of the soccer players group in DT were significantly reduced than in ST; $\text{CRP}_{\text{hip-knee}}$ values of the ordinary student group in DT were significantly higher than in ST.

The MARP and DP changes under the two task conditions are presented in Figure 6. The $\text{MARP}_{\text{hip-knee}}$, $\text{DP}_{\text{hip-knee}}$ and $\text{DP}_{\text{knee-ankle}}$ changes in DT were significantly different relative to the

TABLE 2 Changes in speed of spheres in different groups (mean±SD).

Movement speed (°/s)	Players	Students	*t*-value	Value of *p*
Initial maximum speed	14.60 ± 2.44	12.77 ± 2.83	−2.092	0.043
Standardized speed	4.42 ± 0.81	3.83 ± 0.85	−2.179	0.035

K_{hip}, the hip joint stiffness; K_{knee}, the knee joint stiffness; K_{ankle}, the ankle joint stiffness; ST, single task; DT, dual task.

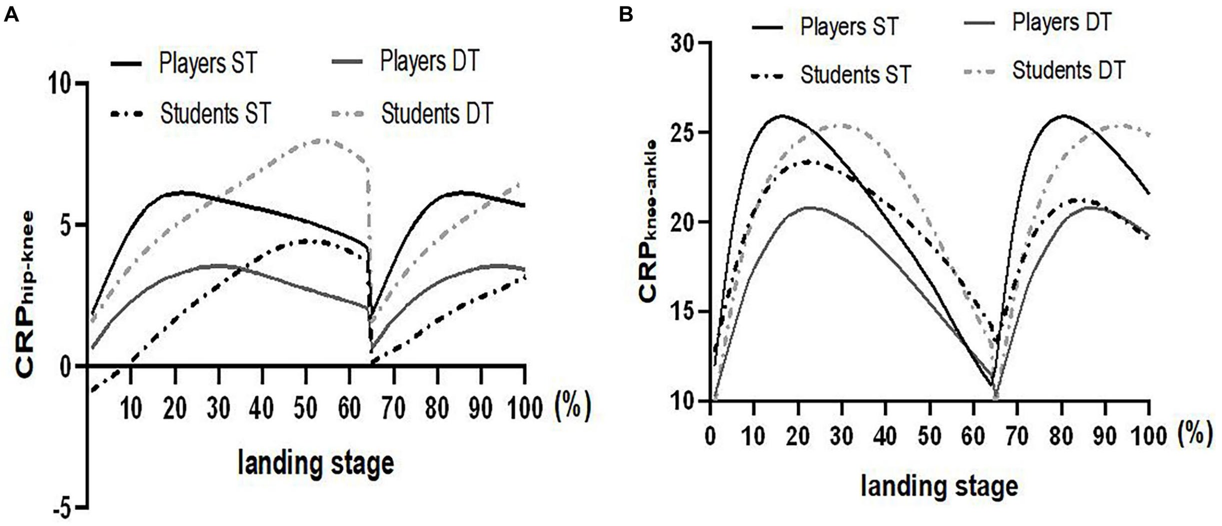


FIGURE 5
Changes in continuous relative phase (CRP) during landing.

values of the changes in ST ($F_{1,37}$ = 4.362, p = 0.041, partial η^2 = 0.070, Cohen's d = −0.523; $F_{1,37}$ = 4.831, p = 0.032, partial η^2 = 0.077, Cohen's d = −0.526; $F_{1,37}$ = 4.276, p = 0.043, partial η^2 = 0.069, Cohen's d = −0.498). In addition, dual-task had a greater effect size on $DP_{hip\text{-}Knee}$ and $MARP_{hip\text{-}knee}$, but a smaller effect size on $DP_{knee\text{-}ankle}$, indicating that dual-task had a higher effect on coordination of the hip and knee joints than single-task for both soccer players and college students. Moreover, the $DP_{Hip\text{-}Knee}$ and $DP_{Knee\text{-}ankle}$ of ordinary college students during landing were significantly larger relative to that of soccer players ($F_{1,27}$ = 9.888, p = 0.003, partial η^2 = 0.146, Cohen's d = 0.798; $F_{1,27}$ = 5.279, $p < 0.025$, partial η^2 = 0.083, Cohen's d = 0.586). It showed that the $DP_{Hip\text{-}Knee}$ and $DP_{Knee\text{-}ankle}$ values of college students were significantly higher than those of the soccer players in ST or DT.

Changes in joint stiffness

The K_{hip} of soccer players was significantly larger in DT than in ST, the K_{hip} of college students was also significantly larger in DT than in ST during landing ($F_{1,27}$ = 4.226, p = 0.044, partial η^2 = 0.068, Cohen's d = −0.549), and the effect size of K_{hip} was moderate. It shows that MOT tasks cause changes in lower extremity mode such as joint stiffness (Table 3). Moreover, No statistically significant difference was observed in K_{hip} of participants under the two tasks ($p > 0.05$).

Discussion

The purpose of this study was to explore whether MOT task interference affects inter-joint coordination during the landing task. The results showed that $MARP_{hip\text{-}knee}$, $DP_{hip\text{-}knee}$, and $DP_{knee\text{-}ankle}$ value under DT were higher compared with the values under ST, indicating that MOT tasks negatively affected inter-joint coordination of soccer players and ordinary college students. This finding is similar to findings from previous research whereby the effects of attentional control on gait and inter-joint coordination were evaluated using a dual-task paradigm. The previous findings showed that attentional control modulated the ability to maintain gait control and regulated inter-joint coordination patterns when adapting to dual-task gait perturbations (Wang et al., 2021). The kinematic chain theory indicates that increase in $MARP_{hip\text{-}knee}$ and $DP_{hip\text{-}knee}$ during landing causes changes in the coordination pattern and variability of the lower extremity hip-knee joint. Studies report that coordination of the hip-knee joint determines the dynamic stability of the lower extremity (Sinsurin et al., 2020), and hip-knee joint coordination is a dominant kinematic variable for reducing the risk of injury (Leonard et al., 2021). Therefore, the increase of $MARP_{hip\text{-}knee}$ observed in our study indicated that the "inverse" movement of the hip joint and knee joint increased over time. This increase indicated a significant contribution of the hip (proximal joint) to the movement of the participant under DT condition. Furthermore, the findings of the current study showed that dual tasks had a greater impact on hip and knee coordination, which agree with past findings that the hip joint is the primary stabilizer of lower extremity movement (Davis et al., 2019). Highly intense hip motion results in high motor instability. Therefore, the coordinated response of the hip—knee was explored in this study because it is a paramount kinematic variable for reducing the risk of injury. In addition, increase in $DP_{knee\text{-}ankle}$ indicated an increase in variability of the knee-ankle joint (calf movement). Active calf movement can decrease the dynamic stability of the hip

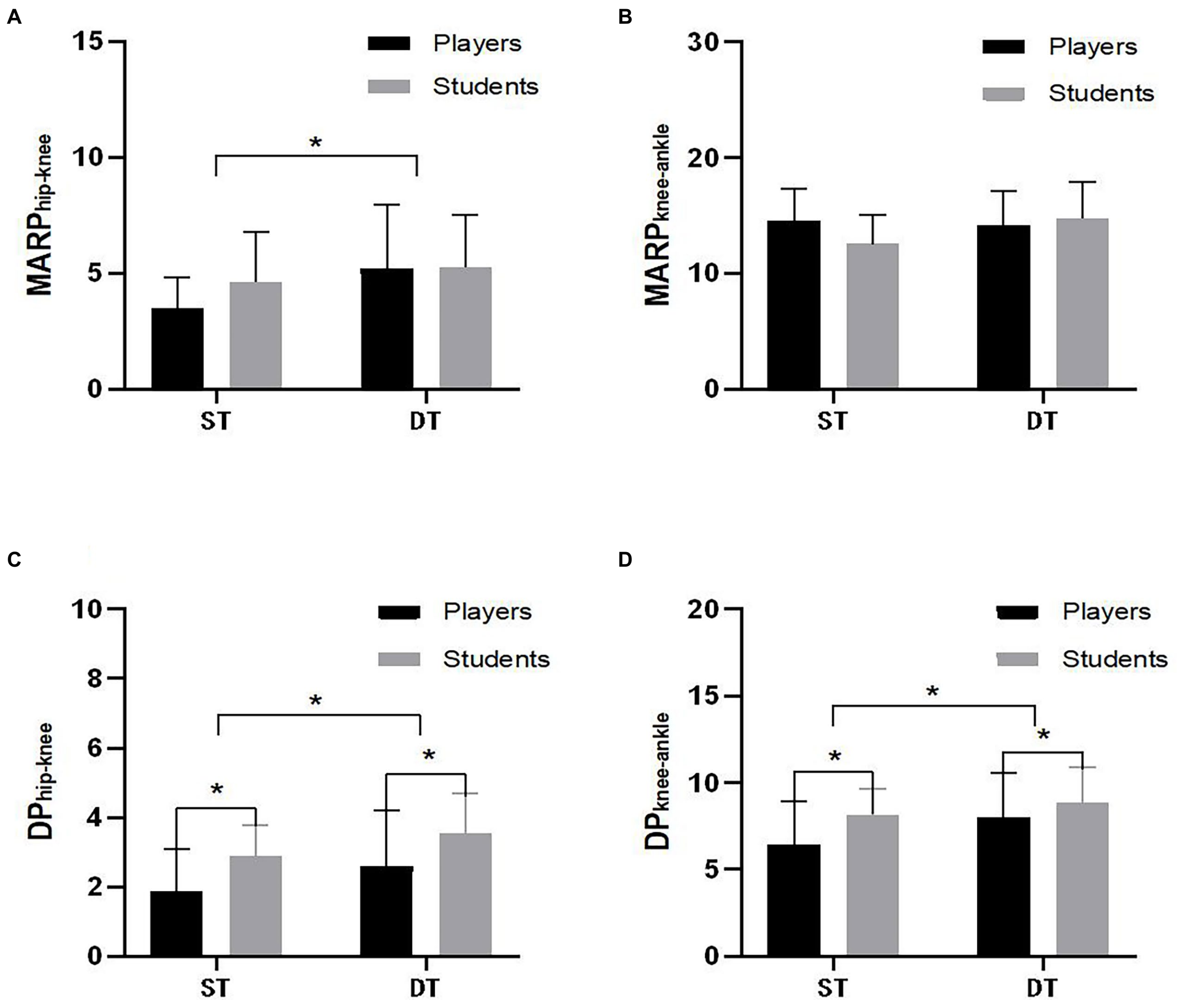


FIGURE 6
Changes in mean absolute relative phase (MARP) and deviation phase (DP) in the two groups under the two tasks.

TABLE 3 Changes in joint stiffness in the two groups under the two tasks (mean±SD).

Joint stiffness	Players		Students	
	ST	DT	ST	DT
K_{hip}	0.222 ± 0.060	0.279 ± 0.060	0.141 ± 0.114	0.168 ± 0.106
K_{knee}	0.015 ± 0.010	0.014 ± 0.010	0.014 ± 0.025	0.011 ± 0.098
K_{ankle}	0.041 ± 0.020	0.045 ± 0.020	0.117 ± 0.217	0.108 ± 0.208

K_{hip}, the hip joint stiffness; K_{knee}, the knee joint stiffness; K_{ankle}, the ankle joint stiffness; ST, single task; DT, dual task.

during coupling movement and change the load on the lower extremity joints during dynamic movement, resulting in increased occurrence of injuries. Other studies reported that distraction affects performance of female athletes when they passed the ball and completed a side-cut action simultaneously (Almonroeder et al., 2019), similar to the findings from our study. The results of the our study are inconsistent with previous findings (Qu and Hu, 2014), which indicated that younger adults choose whether to focus on cognitive tasks to adjust their motor coordination patterns based on the level of difficulty of the task. This difference can be attributed to differences in the types of tasks and loads used in the various studies. The sequential subtraction tasks used previously did not involve sensory perception. In addition, 3D-MOT tasks may not have the similar effects on individuals as other cognitive tasks. Therefore, further research should be conducted to explore the effects of different types of cognitive tasks on individuals.

A previous study reported that soccer players have to extract and process information after subjection to cognitive and sensory tasks especially when they take part in team sports (Romeas et al., 2016), which is consistent with the present findings. These specific cognitive functions (MOT tasks) decreased when participants were exposed to higher stress levels (during games or competitions), leading to a decrease in attention control ability (Romeas et al., 2016). A previous review article summarized various studies that reported the effects of distraction on landing activities with the findings indicating that distraction can impair performance in sports similar to the findings of the present study (Hughes and Dai, 2021). In addition, another past study indicates that complex cognitive task have a significant effect on jump-landing movement quality, which can be attributed to specific cognitive functions such as attention allocation (Schnittjer et al., 2021). Neurocognitive mechanism comprises a group of regions mainly located in the frontal and parietal regions of the cerebral cortex (precentral gyrus and paracentral gyrus). From the perspective of neurocognitive mechanism, the central motor command consists of a group of regions located primarily in the frontal and parietal regions of the cerebral cortex (precentral gyrus and paracentral gyrus), responsible for the control of muscle tissue by Beta cells from the superficial regions of the cerebral cortex into the corticospinal tract. These motor fibers cross first as they travel from the brain to the medulla and spinal cord before reaching the muscles. There is compelling evidence that activity in these motor areas may depend on attentional demands imposed by physical tasks (Bigliassi et al., 2016).

However, the MOT task, which is a high-level perceptual-cognitive interference, involves several cognitive processes, such as visual attention allocation (Trick et al., 2012). The effects observed can be attributed to the limited allocation of visual attention because the same neural network is implicated in controlling motor and cognitive functions. Notably, the MOT task requires allocation of a higher number of cortical resources, thus the dual task could have been affected resulting in changes in the motor system beyond conscious control. Results using theoretical capacity-sharing model explains this phenomenon (Pashler, 1994). In the present study, visual information played a key role in completion of the dual-task by the participants. Participants may have prioritized one task (MOT) by reducing self-control task (landing) because perceptual-cognitive tasks (MOT task) and motor-related demands (landing task) competed for frontal lobe resources. This affected postural control of the participants and impaired inter-joint coordination. Therefore, these results indicate that individual neuromuscular activity under the MOT task did not effectively control the challenging dual-task.

The results showed that the $DP_{\text{Hip-Knee}}$ and $DP_{\text{Knee-ankle}}$ values of college students were significantly higher than those of the soccer players in ST or DT, indicating that college students exhibited more inter-joint coordination variability compared with soccer players, while MOT tasks did not affect this finding. Previous studies have shown that excessive variation of lower extremity coordination mode could exacerbate impairment of neuromuscular control, resulting in abnormal biomechanical patterns and higher joint loads when subjects participate in sports (DiCesare et al., 2019). These results may reflect that soccer players exhibit less inter-joint coordination variability and have better coordination, and they perform better in lower extremity joint adaptability, sensory feedback, and load decentralization ability than ordinary college students. It might be attributed to the fact that ordinary college students do not have the same soccer experience or training. A recent study used a 3D-MOT task in high-intensity interval training (HIIT) to explore the effect on perceptual-cognitive performance and reported that the combination training resulted in task-specific benefits, which confirms our findings (Park et al., 2021). Romeas et al. (2016) reported that long-term MOT training improves athletic performance and decision-making ability of soccer players. Various studies have used MOT tasks for optokinetic simulation training, and the results showed that it improved posture stability for patients with motor coordination disorders such as stroke and hemiplegia (Chen et al., 2021; Komagata et al., 2021). Thus it is speculated that long-term MOT training may positively affect motor stability and improve inter-joint coordination. These findings can form a basis for designing further studies to explore application of perceptual-cognitive training in special populations with dual-task requirements. Notably, although soccer players had better coordination, the MOT task could affect their inter-joint coordination. Long-term training may improve dual-task processing ability and ensure stable coordination of soccer players but cannot eliminate the interference of MOT tasks.

The present study had some limitations. The sample size was not calculated for this study. Another limitation is that we have no idea how much these athletes train, and the ordinary students were not matched for any other characteristic. Further studies should be conducted to explore the effects of the level of training and both groups should be matched for all characteristics that may affect cognitive function.

Conclusion

The findings showed that neuromuscular activity of individuals under perceptual-cognitive task interference cannot effectively control the challenging dual-task movement and the perceptual-cognitive task can adversely affect inter-joint coordination of soccer players and college students, mainly the hip and knee coordination. Although soccer players had better coordination, the MOT task could affect their inter-joint coordination. The findings for the study provide a theoretical basis for improving individual inter-joint coordination and reducing the risk of injury.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the University of Soochow research ethics committee has approved the study (no. ECSU-2019000209). The patients/participants provided their written informed consent to participate in this study.

Author contributions

YR provided conceptualization, made formal analysis, and wrote the original draft. CW modified and corrected the text expression, reviewed, and edited. AL provided decisive advice on the key steps of each experiment implementation and article revision, in charge of supervision. All authors made substantial contributions and participated in the study design, data collection, and result analysis together. We revised different versions of the paper together, and the final submitted version was approved by all authors.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Almonroeder, T. G., Kernozek, T., Cobb, S., Slavens, B., Wang, J., and Huddleston, W. (2019). Divided attention during cutting influences lower extremity mechanics in female athletes. *Sports Biomech.* 18, 264–276. doi: 10.1080/14763141.2017.1391327

Alvarez, G. A., and Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychol. Sci.* 16, 637–643. doi: 10.1111/j.1467-9280.2005.01587.x

Alves, H., Voss, M. W., Boot, W. R., Deslandes, A., Cossich, V., Salles, J. I., et al. (2013). Perceptual-cognitive expertise in elite volleyball players. *Front. Psychol.* 4, 1–9. doi: 10.3389/fpsyg.2013.00036

Bigliassi, M., Silva, V. B., Karageorghis, C. I., Bird, J. M., Santos, P. C., and Altimari, L. R. (2016). Brain mechanisms that underlie the effects of motivational audiovisual stimuli on psychophysiological responses during exercise. *Physiol. Behav.* 158, 128–136. doi: 10.1016/j.physbeh.2016.03.001

Chen, Y., Chou, Y., and Hwang, I. (2021). Reliance on visual input for balance skill transfer in older adults: EEG connectome analysis using minimal spanning tree. *Front. Aging Neurosci.* 13, 11–20. doi: 10.3389/fnagi.2021.632553

Dahlin, E., Neely, A. S., Larsson, A., Backman, L., and Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science* 320, 1510–1512. doi: 10.1126/science.1155466

Davis, K., Williams, J. L., Sanford, B. A., and Zucker-Levin, A. (2019). Assessing lower extremity coordination and coordination variability in individuals with anterior cruciate ligament reconstruction during walking. *Gait Posture* 67, 154–159. doi: 10.1016/j.gaitpost.2018.10.010

DiCesare, C. A., Montalvo, A., Foss, K. D. B., Thomas, S. M., Hewett, T. E., Jayanthi, N. A., et al. (2019). Sport specialization and coordination differences in multisport adolescent female basketball, soccer, and volleyball athletes. *J. Athl. Train.* 54, 1105–1114. doi: 10.4085/1062-6050-407-18

Ehmann, P., Beavan, A., Spielmann, J., Ruf, L., Mayer, J., Rohrmann, S., et al. (2021). 360 degrees-multiple object tracking in team sport athletes: reliability and relationship to visuospatial cognitive functions. *Psychol. Sport Exerc.* 55, 101952–101958. doi: 10.1016/j.psychsport.2021.101952

Farley, C. T., Houdijk, H. H. P., Van Strien, C., and Louie, M. (1998). Mechanism of leg stiffness adjustment for hopping on surfaces of different stiffnesses. *J. Appl. Physiol.* 85, 1044–1055. doi: 10.1152/jappl.1998.85.3.1044

Gabbett, T. J., Carius, J., and Mulvey, M. (2008). Does improved decision-making ability reduce the physiological demands of game-based activities in field sport athletes. *J. Strength Cond. Res.* 22, 2027–2035. doi: 10.1519/JSC.0b013e3181887f34

Hughes, G., and Dai, B. (2021). The influence of decision making and divided attention on lower limb biomechanics associated with anterior cruciate ligament injury: a narrative review. *Sport Biomech.* 1–16. doi: 10.1080/14763141.2021.1898671

Kim, H., Son, S. J., Seeley, M. K., and Hopkins, J. T. (2019). Altered movement biomechanics in chronic ankle instability, coper, and control groups: energy absorption and distribution implications. *J. Athl. Train.* 54, 708–717. doi: 10.4085/1062-6050-483-17

Komagata, J., Sugiura, A., Takamura, H., Ohta, A., and Kitama, T. (2021). Effect of optokinetic stimulation on weight-bearing shift in standing and sitting positions in stroke patients. *Eur. J. Phys. Rehab. Med.* 57, 13–23. doi: 10.23736/S1973-9087.20.06184-5

Leonard, K. A., Simon, J. E., Yom, J., and Grooms, D. R. (2021). The immediate effects of expert and dyad external focus feedback on drop landing biomechanics in female athletes: an instrumented field study. *J. Orthop. Sport Phys.* 16, 96–105. doi: 10.26603/001c.18717

Li, Y., Ko, J., Walker, M. A., Brown, C. N., and Simpson, K. J. (2021). Joint coordination and stiffness during landing in individuals with chronic ankle instability. *J. Appl. Biomech.* 37, 156–162. doi: 10.1123/jab.2020-0272

Mejane, J., Faubert, J., Romeas, T., and Labbe, D. R. (2019). The combined impact of a perceptual–cognitive task and neuromuscular fatigue on knee biomechanics during landing. *Knee* 26, 52–60. doi: 10.1016/j.knee.2018.10.017

Park, S., Klotzbier, T. J., and Schott, N. (2021). The effects of the combination of high-intensity interval training with 3d-multiple object tracking task on perceptual-cognitive performance: A randomized controlled intervention trial. *Int. J. Environ. Res. Public Health* 18:4862. doi: 10.3390/ijerph18094862

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychol. Bull.* 116, 220–244. doi: 10.1037/0033-2909.116.2.220

Qiu, F., Pi, Y., Liu, K., Li, X., Zhang, J., and Wu, Y. (2018). Influence of sports expertise level on attention in multiple object tracking. *PeerJ* 6:e5732. doi: 10.7717/peerj.5732

Qu, X., and Hu, X. (2014). Lower-extremity kinematics and postural stability during stair negotiation: Effects of two cognitive tasks. *Clin. Biomech.* 29, 40–46. doi: 10.1016/j.clinbiomech.2013.11.006

Rahnama, N., Reilly, T., and Lees, A. (2002). Injury risk associated with playing actions during competitive soccer. *Br. J. Sports Med.* 36, 354–359. doi: 10.1136/bjsm.36.5.354

Ren, Y., Wang, C., Zhang, L., and Lu, A. (2021). The effects of visual cognitive tasks on landing stability and lower extremity injury risk in high-level soccer players. *Gait Posture* 92, 230–235. doi: 10.1016/j.gaitpost.2021.11.031

Romeas, T., Guldner, A., and Faubert, J. (2016). 3D-multiple object tracking training task improves passing decision-making accuracy in soccer players. *Psychol. Sport Exerc.* 22, 1–9. doi: 10.1016/j.psychsport.2015.06.002

Schnittjer, A., Simon, J. E., Yom, J., and Grooms, D. R. (2021). The effects of a cognitive dual task on jump-landing movement quality. *Int. J. Sports Med.* 42, 90–95. doi: 10.1055/a-1195-2700

Schoner, G., and Kelso, J. A. S. (1988). Dynamic pattern generation in behavioral and neural systems. *Science* 239, 1513–1520. doi: 10.1126/science.3281253

Sheikhi, B., Letafatkar, A., Thomas, A. C., and Ford, K. R. (2021). Altered trunk and lower extremity movement coordination after neuromuscular training with and without external focus instruction: a randomized controlled trial. *BMC Sports Sci. Med. Rehabil.* 13:92. doi: 10.1186/s13102-021-00326-9

Sinsurin, K., Vachalathiti, R., Srisangboriboon, S., and Richards, J. (2020). Knee joint coordination during single-leg landing in different directions. *Sports Biomech.* 19, 652–664. doi: 10.1080/14763141.2018.1510024

Trick, L. M., Mutreja, R., and Hunt, K. (2012). Spatial and visuospatial working memory tests predict performance in classic multiple-object tracking in young adults, but nonspatial measures of the executive do not. *Atten. Percept. Psychophys.* 74, 300–311. doi: 10.3758/s13414-011-0235-2

Wang, C., Wang, G., Lu, A., and Zhao, Y. (2021). Effects of attentional control on gait and inter-joint coordination during dual-task walking. *Front. Psychol.* 12, 1–12. doi: 10.3389/fpsyg.2021.665175

Zhang, K., Chen, Y., Li, Y., Yao, S., Su, Y., and Wang, I. (2021). Different drop heights in bilateral asymmetry and interjoint coordination during repetitive drop-jumps. *Symmetry* 13:1590. doi: 10.3390/sym13091590

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 07 October 2022
DOI 10.3389/fpsyg.2022.922291

OPEN ACCESS

EDITED BY
Naomi Sweller,
Macquarie University, Australia

REVIEWED BY
Giorgia Bussu,
Uppsala University, Sweden
Kathy Kar-man Shum,
Department of Psychology, University of Hong Kong, Hong Kong SAR, China

*CORRESPONDENCE
Oleg Zacharov
olegjang@hotmail.com

SPECIALTY SECTION
This article was submitted to Cognition,
a section of the journal Frontiers in Psychology

RECEIVED 17 April 2022
ACCEPTED 20 September 2022
PUBLISHED 07 October 2022

CITATION
Zacharov O, Huster RJ and Kaale A (2022) Working memory in pre-school children with autism spectrum disorder: An eye-tracking study. *Front. Psychol.* 13:922291.
doi: 10.3389/fpsyg.2022.922291

Working memory in pre-school children with autism spectrum disorder: An eye-tracking study

Oleg Zacharov[1]*, Rene Jürgen Huster[2] and Anett Kaale[1,3]

[1]Department of Special Needs Education, University of Oslo, Oslo, Norway, [2]Department of Psychology, University of Oslo, Oslo, Norway, [3]Norwegian Centre of Expertise for Neurodevelopmental Disorders and Hypersomnias, Oslo University Hospital, Oslo, Norway

Working memory (WM) was examined in pre-school children with Autism spectrum disorder (ASD) and children with typical development using eye-tracking technology. The children were presented with a digital A-not-B task (with a short and a long waiting condition) where they passively viewed animations of a moving train. Moreover, the current study investigated the relationship between non-verbal mental age (NVMA) and the performance on the task. No group differences were found in the average looking durations between the ASD and typically developing (TD) groups on either the short or long waiting conditions. Although the NVMA of the ASD group was lower than that of the TD group there were no correlations between NVMA and task performance in either group. The results suggest that WM in young children with ASD might not be different from that of TD children. However, the results might be due to ceiling effects of the task and thus needs to be further investigated.

KEYWORDS

working memory, pre-school children, autism spectrum disorder, typical development, nonverbal mental age, eye-tracking, A-not-B task

Introduction

Autism spectrum disorder (ASD) can be described by problems in social communication and interaction, and restricted and repetitive behaviors and interests (American Psychiatric Association, 2013). ASD can be reliably diagnosed at the age of 2, however, it is possible to detect it as early as 18 months of age (Hyman et al., 2020). The symptom presentation and severity vary greatly among individuals with ASD. In addition to the "classical" difficulties exhibited by individuals with ASD, emerging evidence demonstrates that executive functioning (EF) impairment may also be prevalent (Garon et al., 2018).

Executive functioning refers to a number of cognitive abilities that are essential for everyday functioning, including working memory (WM), cognitive flexibility, and inhibitory control (Anderson, 2014). These skills are especially crucial during the pre-school years. For example, Pellicano et al. (2017) reported that individual differences in WM and inhibitory control during the pre-school years were uniquely related to variation in school readiness for both typically developing (TD) children and children with ASD.

Working memory refers to the temporary storage and manipulation of information in the brain (Anderson, 2014). A meta-analysis of studies investigating WM revealed that individuals with ASD exhibited significant WM impairment (Cohen's $d = -0.61$) compared to the control group (Wang et al., 2017). No association between the WM impairment and age was found. Studies of WM in pre-school ASD has generated evidence demonstrating that pre-school autistic children exhibit impairments on various performance-based WM tasks and rating scales. Edmunds et al. (2021) measured WM in children from 2 to 4 year using three tasks; the Hide and Seek Task, the Boxes Task and the Self-Ordered Pointing Task (SOPT). They found that 2–4-year old children with ASD exhibited lower composite scores, combining all three WM tasks, compared to TD children. Similarly, the aforementioned study by Pellicano et al. (2017) demonstrated that 4 1/2–6-year-old children with ASD scored lower than TD children on the Corsi Blocks task measuring WM. In the Corsi Blocks task children are asked to repeat the tapping of blocks in the same order previously displayed by the experimenter.

However, in parallel to studies reporting group differences, there is also evidence showing that WM is intact in pre-school children with ASD. For example, Gardiner et al. (2017) reported no significant differences in WM performance between 3 1/2 and 7-year-old children with ASD and TD children on the Boxes task. In this task, children were asked to find Jack-in-the-box while keeping in mind boxes they already had searched. Also, no WM deficit was found on the Spinning Pots task between 4 and 6-year old TD children and children with ASD (Valeri et al., 2020). In the Spinning Pots task, children are asked to place an object in eight pots which are then covered and rotated. After each rotation the cover is lifted and the child is asked to find the placed object. This procedure is done until 8 objects are found or after 15 trials.

One of the factors that may contribute to the inconsistent findings could be associated with different parameters of the WM tasks. In pre-school ASD, WM has been predominantly investigated with performance-based tasks which vary in administration procedures and levels of difficulty. During administration of most WM tasks, pre-school children are required to understand and comply with verbal instructions, and produce some kind of motor responses, such as pointing. This may, in addition to the WM demands, further challenge children's mental capacity and thus may negatively influence their performance. Moreover, pre-school children with ASD often have language difficulties which may, despite their potentially intact WM, prevent them from understanding the instructions. Additionally, WM tasks developed for pre-schoolers vary in difficulty levels and may thus elicit differential performance across the pre-school age range. For example, for the Spinning Pots task, the performance was proposed to improve during the 18–42-month period among TD children (Garon et al., 2018). At 42 months and higher, children were expected to exhibit performance at or close to ceiling. This may explain the absence of group difference in the Valeri et al. (2020) study in which the minimum age of participants was 4 years. In contrast, in the Hide and Seek Task children are asked to place paper animal cutouts on the appropriate boxes. Here, the age range at which children's performance improves has been shown to be broader, namely 18–60 months (Garon et al., 2014). In the Edmunds et al. (2021) study the Hide and Seek Task was administered to participants between 2- and 4-years. Thus, although there are many WM tasks designed for pre-schoolers, some may capture the children's impairment in a given study while others may not, because they differ in WM demand, or the overall difficulty level relative to the age group in the study. This has important consequences for the interpretation of findings.

Different administration procedures and levels of difficulty of WM tasks would require children to possess sufficient levels of cognitive functioning. The inconsistent findings could be associated with the large heterogeneity in cognitive functioning among children with ASD. To control for this, researchers usually employ standardized tests to match children on language or some general cognitive ability, most commonly non-verbal mental age (NVMA). There is a general agreement that NVMA or language ability could be mediating factors of EF (Stephens et al., 2018). Doing so, however, restricts the representativeness of the group. Besides, while the NVMA is a preferred matching criterion in most studies with autistic pre-schoolers, there is scarcity of research in the ASD field as to whether NVMA is associated with WM performance. One study by Mungkhetklang et al. (2016) demonstrated that WM contributed to non-verbal problem solving for school age TD children and children with Intellectual Disability. Moreover, a recent study by Stephens et al. (2018) measured NVMA and verbal mental age with Mullen Scales of Early Learning (MSEL) in 6-year-old TD children and found that NVMA was a better predictor of EF that was measured by performance-based tasks. The verbal mental age, on the other hand, was a better predictor of parent-reported EF. Having a sample that has a larger spread in NVMA, not only would increase the sample representativeness, but could also shed some light on whether there is a relationship between NVMA and performance on WM tasks in pre-school children. A relationship would suggest that WM impairment could be related more to the general cognitive profiles of children, specifically the NVMA, rather than the diagnosis.

In recent years there has been a growing interest in the use of eye-tracking technology as tool to provide insight into psychological processes. By measuring where a person is looking, one can uncover essential information about how stimuli are being experienced, prioritized, and processed. There are two advantages for using eye-tracking on pre-school children with ASD. First, due to its millisecond-level precision, one can obtain a more precise measure of their attentional patterns. Second, eye-tracking is non-invasive and allows for data to be collected when children passively view the screen. Hence, this makes it an appropriate tool for investigating areas of interest also in minimally verbal and developmentally delayed pre-school children with ASD who often experience difficulties in dealing with complex cognitive demands. Research has shown that development of language in toddlerhood is important for the emergence and development of EF during early childhood (Stephens et al., 2018). In a study by Hill et al. (2015) 5–8-year-old ASD children with and without developmental language disorder (DLD) and a group of children with only DLD were assessed on verbal and non-verbal (visual) WM tasks. The autistic children with concurrent DLD performed worse on the verbal WM tasks than those without DLD. The children with ASD and DLD performed similar to the children with only DLD on most verbal WM tasks. However, there were no group differences in performance on non-verbal WM tasks, which suggests that increased demands associated with verbal WM task could lead to differential performance between ASD children with and without DLD.

The majority of WM tasks require verbal comprehension (i.e., understanding task instructions) for their completion. Developmental and language delays are common in young autistic children (Friedman and Sterling, 2019). Selecting a narrow sample capable of performing on certain tasks, which is usually the case in most studies, ignores those individuals who would more accurately represent the population at this age period. Using a task with removed verbal demands while still capable of assessing WM could be essential in such cases. Moreover, since eye-tracking could be suited for studying children with various cognitive functioning, which assumes varying levels of NVMA, it makes it possible to investigate the relationship between the measured WM performance and NVMA.

Most of the eye-tracking studies that are conducted on children with ASD have focused on social impairment, such as allocation of attention to social stimuli and predictive gaze (Hamner and Vivanti, 2019). There is a scarcity of eye-tracking research investigating EF in individuals with ASD, and to our knowledge only Fanning et al. (2018) has applied this technology to look at WM in pre-school children with ASD. In this study where they used an eye-tracking version of the A-not-B task, children were presented with a character in the middle of the screen for 3 s that then disappeared either into the left or right target location as a sound was played (croaking frog). After a 6 s waiting period another sound (croaking frog) was played for 3 s, implicitly asking children "from which location, A or B, did the sound originate?" It was reasoned that if children memorized which of the two locations the stimulus had disappeared to, on hearing the sound they would fixate at that location. Looking durations to the target and non-target locations were measured during the 3 s after the 6 s waiting period. WM performance was indicated by the duration of the preferential location at the target location. Researchers found no group differences on the eye-tracking version of the A-not-B task between 2- and 5-year-old TD children and children with ASD.

Similarly to the difficulty levels associated with various performance-based WM tasks, caution should also be applied to tasks using eye-tracking. Given that the A-not-B task originally was developed for infants, it is important to adjust the task to the range of the study participants' developmental level. A more elaborate version of the task might make group differences visible. For example, by introducing shorter and longer waiting periods it may be possible to manipulate the WM load, which will consequently influence the performance.

Overall, given the scarcity of eye-tracking research on WM in pre-school children with ASD and the aforementioned issues that may be associated with task parameters in previous studies, further investigation is necessary. Moreover, since eye-tracking allows for inclusion of developmentally delayed and/or minimally verbal children, it may shed some light on the involvement of NVMA in WM task performance, that would otherwise be difficult to do due to complex cognitive demands associated with manual and verbally loaded tasks.

Rationale

The aim of the current study is twofold. First the study will investigate and compare the WM performance of pre-school age children with ASD and TD on a novel computerized A-not-B task using the benefits of eye-tracking technology. Since the original A-not-B task is designed for infants and required children to reach for a hidden object, a number of modifications have been made. We kept the task passive without verbal instructions, but incorporated distractors and two WM load conditions (short vs. long waiting period). It is hypothesized that the long waiting period, which poses higher WM load, would affect the performance of children with ASD the most.

Being the most commonly used matching criterion in research on ASD, the NVMA is rarely studied along the EF performance in pre-school children with ASD. While matching provides control for variables that may affect the outcome, little is known whether NVMA could be related to WM performance. Hence, the current study will also investigate the relationship between children's NVMA and their performance on the A-not-B task. Considering the aforementioned research by Mungkhetklang et al. (2016) and Stephens et al. (2018), it is hypothesized that NVMA of both children with and without ASD would be associated with their performance on the eye-tracking version of the A-not-B task.

Materials and methods

Participants

Thirty-seven pre-school children participated in the study. 13 children, aged 31–68 months (*M* = 53.54, SD = 11.22), were diagnosed with ASD, and 24 children, aged 37–59 months (*M* = 49.50, SD = 6), were TD. There were 11 boys (84.6%) and 2 girls (15.4%) in the ASD group, and 11 boys (45.8%) and 13 girls (54.2%) in the TD group (**Table 1**). The recruitment of children with ASD was done through specialist health services, educational-psychological services, and pre-schools in Oslo and surrounding counties. The recruitment of TD children was done via pre-schools in Oslo and surrounding counties. The children in the ASD group were diagnosed by the specialist health services using the International Classification of Diseases, 10th version (World Health Organization [WHO], 1993) based on a detailed clinical evaluation, including tests, interview with caretakers and observations. As a part of the current study the parents scored their children on the Social Communication Questionnaire (SCQ) (Rutter et al., 2003). All but one child scored above the cut-off for ASD. However, SCQ data from 4 children (30.8%) are missing. Children with severe motor, visual or hearing problems did not participate in the current study. Also, due limitations in eye-tracking technology, children

TABLE 1 Descriptive statistics for the typically developing (TD) and autism spectrum disorder (ASD) groups.

	ASD (*n* = 13)	**TD (*n* = 24)**	***t***	***p***	**Hedges' *g***
CA (months)					
M (SD)	53.54 (11.22)	49.50 (6)	1.43	0.160	0.49
Range	31–68	37–59			
Social Communication Questionnaire–Parents					
M (SD)	16.33 (6.06)				
Range	8–27				
Missing data	4 (30.8%)				
NVMA (Months)					
M (SD)	31.58 (10.26)	48.96 (8.57)	−5.37	$p < 0.001$	1.90
Range	14–50	30–67			
Missing data	1 (7.7%)	–			
Receptive Language–Age (months)					

			Mann–Whitney U		
			U	***p***	**η 2**
M (SD)	23.67 (11.92)	41.42 (4.8)	27.5	$p < 0.001$	0.44
Range	8–46	30–48			
Missing data	1 (7.7%)	–			
Expressive Language—Age (Months)					
M (SD)	25.17 (18.9)	54.75 (13.38)	35.5	$p < 0.001$	0.38
Range	4–67	23–70			
Missing data	1 (7.7%)	-			
Child's Spoken Language					
Norwegian	7 (53.8%)	13 (54.2%)			
Norwegian + Other	1 (7.7%)	5 (20.8%)			
Missing data	5 (38.5%)	6 (25%)			
Gender					
Male	11 (84.6%)	11 (45.8%)			
Female	2 (15.4%)	13 (54.2%)			
Maternal Education					
Primary School	1 (7.7%)	–			
High School	2 (15.4%)	1 (4.2%)			
University	8 (61.5%)	17 (70.8%)			
Missing data	2 (15.4%)	–			

wearing glasses were not included in the study. The Regional Committees for Medical and Health Research Ethics approved the study and written informed consent was given by all parents.

Procedure

The present research was a part of a longitudinal project which aim was to study early development and learning in TD children and children with ASD. All children were tested during 1 day with multiple measures of language and cognitive ability. WM was measured with the eye-tracking version of the A-not-B task. The task was built in Experiment Builder (SR Research) and presented on the Lenovo ThinkPad W550s laptop. All children were seated approximately 70 cm from the laptop screen. Eye-tracking data was recorded with EyeLink 1000 Plus. No chin rest was used during the recording. Instead, a target sticker was placed on the children's forehead or cheek allowing the participants to move during the recording while the eye-tracking camera followed their eyes. The eye-tracking recording was performed on the left eye (monocular) for all participants. The sampling rate was set to 500 frames per second. A 16 mm camera lens was used. Children performed the task without familiarization phase. The experiment could be paused at any moment during the testing. During the pause, a flickering image could be played on the presentation laptop in order to attract the child's attention back to the screen. After the pause, the previous trial was restarted. The sound stimuli were played through external speakers. All the testing, including the eye-tracking, was carried out by the first author and a research assistant in a quiet room in the children's pre-school or in the laboratory at the University of Oslo. Test duration for each child ranged from 2 to 4 h including breaks. Social (e.g., praise, play brakes) and edible motivators (e.g., raisins, apple bits) were provided when necessary to increase children's compliance. As the eye-tracking WM task require the child to be attentive to the screen it was usually administered on the first hour of testing. All parents were asked to fill out questionnaires in order to obtain demographic information.

Measures

Cognitive and language level

Children's NVMA and expressive and receptive language were measured with the (MSEL; Mullen, 1995). Infants and children up to 68 months of age are eligible for MSEL. MSEL consists of five subscales, namely Gross Motor, Fine Motor, Expressive Language, Receptive Language, and Visual Reception. Composite of the Visual Reception and Fine Motor subscales were used to calculate the NVMA, while the Receptive and Expressive subscales were used to calculate language level.

Working memory measure

A-not-B task

The current task is a modification of the manual version of the A-not-B task. Although the task was initially proposed to measure WM in infants and toddlers (Diamond et al., 1997), original and modified manual A-not-B tasks have also been used with pre-school autistic children aged between 40 and 80 months of age (McEvoy et al., 1993; Griffith et al., 1999). The modifications were made to the original task in order to increase WM load. In the manual version of this task, the child is presented with two containers. During "A" trials, a toy is hidden in the container "A" and the researcher leaves the child to find the toy, usually by asking the question "where is the toy?" After a number of consecutive "A" trials, the toy is then hidden in the container "B." After hiding the toy during "B" trials, the experimenter claps his/her hands in order to disengage the child's attention from the hiding area. Afterward, the child is allowed to search for the toy. Infants and toddlers usually make a perseverative error during this trial by continuing to search for the toy in the container "A" (Diamond et al., 1997).

In the current task, children viewed a series of movie clips (width = 1,920, height = 1,080, frame rate = 25) of a train moving from the center into a tunnel either to the left or the right. The train moving into the left tunnel constituted "A" trials, and the train moving into the right tunnel constituted "B" trials. In total, there were 10 trials, six "A" trials and four "B" trials. The trials were presented in the following order: "A," "A," "A," "B," "B," "A," "A," "A," "B," "B." Each trial ("A" and "B") was separated into three parts: (1) a train going and disappearing into the tunnel, (2) a waiting period with ("B" trials) or without ("A" trials) distractor, and (3) the train returning back to center. A melody was played throughout the task. It was possible to pause the task at any moment and repeat an interrupted trial. As illustrated in **Figure 1**, at the beginning of each "A" trial, the train that was positioned at the center of the screen (x = 960, y = 540) would start moving horizontally to the left side of the screen and disappearing into the tunnel (x = 340, y = 540). The time it took for the train to disappear completely into the tunnel was 5,100 ms. A chugging train sound (duration = 5,100 ms) was played as the train moved toward and disappeared into the tunnel. After the train had disappeared, a waiting period was initiated. For "A" trails, the waiting period was 10,200 ms. 5,100 ms into the waiting period, the train whistle sound was played for 1,870 ms followed by a chugging train sound that was played for the rest of the waiting period and until the end of the "A" trial. The train whistle sound was implemented to signal the children that the train was about to come back from the tunnel. After the waiting period the train reappeared and moved back from the same tunnel to the center of the screen. Each "A" trial lasted in total for 20,400 ms. For "A" trials, the left tunnel was the correct target location. A 5-point calibration and validation procedure preceded the task.

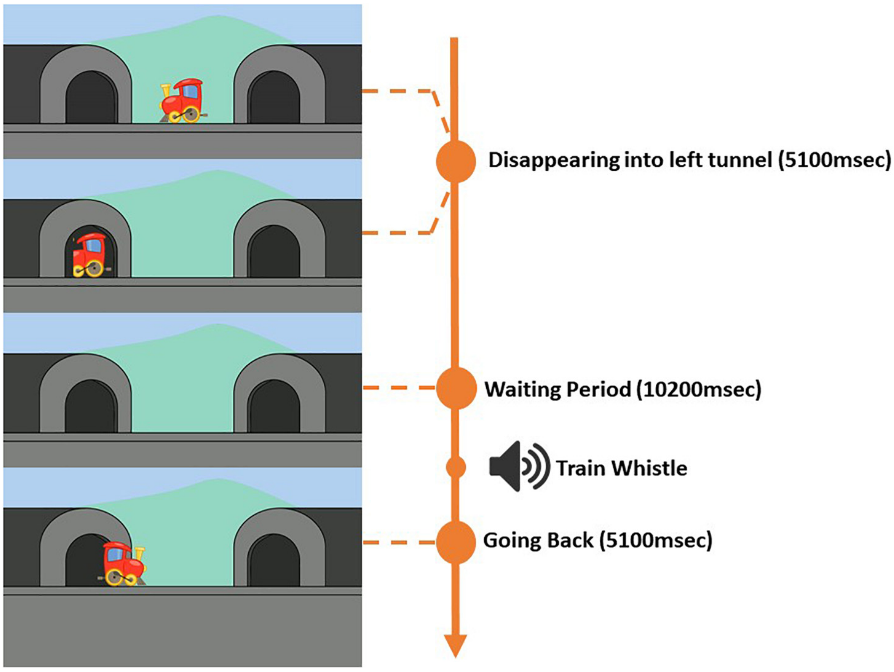


FIGURE 1
Progression of the A Trial.

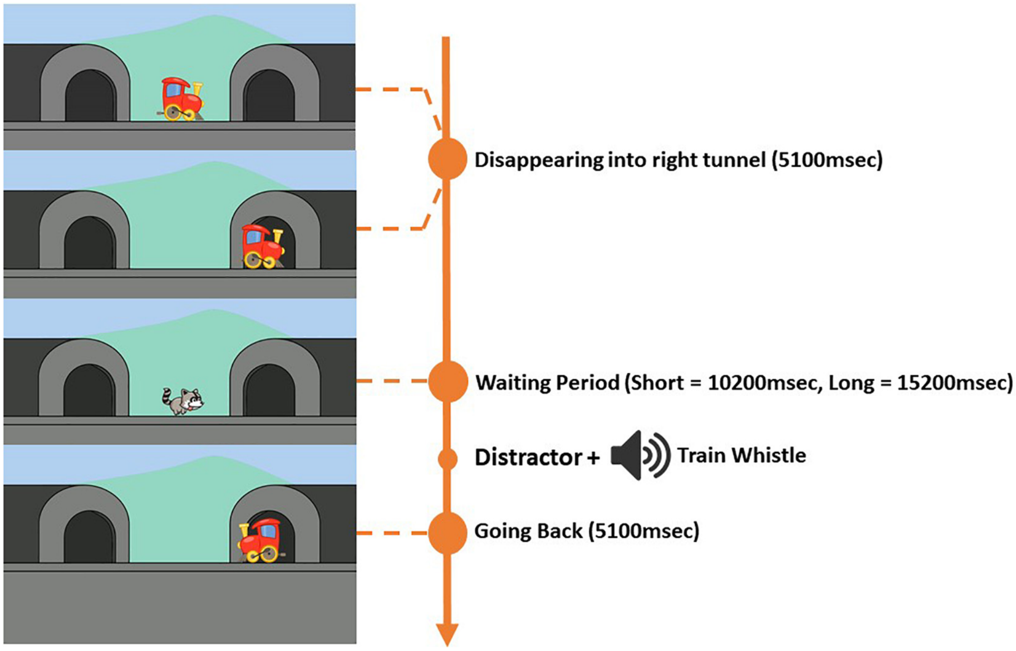


FIGURE 2
Progression of the B Trials with short (10,200 ms) and long (15,200 ms) waiting periods.

As for the "B" trials (Figure 2), there were two trial types, one with a shorter waiting period (10,200 ms) and one with a longer waiting period (15,200 ms). In total, there were two "B" trials with a shorter period and two "B" trials with a longer waiting period. The same trial types were never presented in succession. At the beginning of each "B" trial, a centrally positioned train (x = 960, y = 540), would start moving horizontally to the right side of the screen and disappearing into the tunnel (x = 1,580, y = 540) within 5,100 ms. As for the "A" trials, an accompanying chugging train sound would play for 5,100 ms until the train had completely disappeared. Depending on the trial type, a short or long waiting period was then initiated. Both "B" trials types had distractors presented during the waiting period. The distractors were in the form of moving animations that were located in the center of the screen. Each "B" trial had a different distractor. The distractors for the "B" trials were always presented at 1,000 ms into the waiting period, and were present for 4,100 ms. The distractors were implemented to divert attention and prevent children from fixating at the tunnel into which the train had disappeared for the rest of the waiting period. A train whistle sound was then played for 1,870 ms at 5,100 ms for "B" trials with a shorter waiting period and at 10,100 ms into the waiting period for "B" trials with a longer waiting period. A chugging train sound was then played for the rest of the waiting period and until the end of the "B" trial. After the waiting period the train would reappear from the same tunnel and move back to center of the screen within 5,100 ms. The total duration of the "B" trial with a shorter waiting period was 20,400 ms and the total duration of the "B" trial with a longer waiting period was 25,400 ms. For "B" trials, the right tunnel was the correct target location.

The eye-tracking version of the A-not-B task relies primarily on the voluntary sustained anticipatory looking until the appearance of the stimulus at the location where the participant expects it to appear. The addition of distractors was necessary to increase the difficulty of the A-not-B task (Watanabe et al., 2012). Moreover, "B" trials with a longer waiting period were incorporated to increase the demand for WM thus increasing the task's difficulty. These features were implemented in an attempt to make the task more appropriate for pre-school age children.

Data and statistical analysis

Descriptive data [e.g., chronological age (CA), NVMA, language level] were analyzed for both groups and reported as means, standard deviations, ranges or frequency and percentages. The eye-tracking data was preprocessed in Data Viewer (SR Research). Two areas of interest were created, equally encompassing both the left and right tunnels (**Figure 3**). The data from "A" trials was used primarily from TD children to test whether children would spend more time looking at the target location which would insinuate that the task was understood. Similarly to Fanning et al. (2018), it was speculated that if the location of the disappeared train was remembered, children would look longer at the correct tunnel upon hearing the train whistle sound stimulus during the waiting period. Fanning et al. (2018) tested this prediction in a pilot study where a group of TD children exhibited greater preferential fixation to the target versus non-target area. This prediction was also tested in the current study using data from the TD pre-school children. A t-test revealed that the TD children (n = 23) spent significantly more time looking at the left tunnel (M = 2,122 ms, SD = 666.82) than the right tunnel (M = 1,150 ms, SD = 470.95) ($t(24) = 5.195$, $p < 0.001$, 95% CI[584.79–1,358.67]).

For both "B" trials types (short and long waiting period), the dependent variables were the average looking durations at the left and the right tunnels during a 5-s period from the start of the train whistle sound to the reappearance of the train from the tunnel into which it has disappeared. The averages were made separately for the "B" trials with shorter and longer waiting

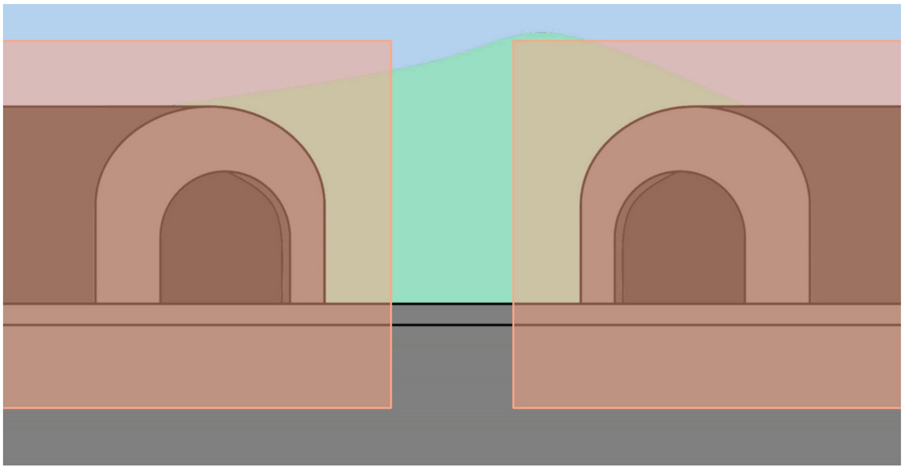


FIGURE 3
Two identical Interest Areas for the left and right tunnels.

periods. This way, it was possible to measure children's "correct" and "incorrect" anticipatory looking after the change of the hiding location. Similar to previous research, only fixations of 100 ms and longer were included in the analyses (Fanning et al., 2018). The average looking durations are reported in milliseconds.

First, for both "B" trial types, paired samples *t*-tests were used to investigate within group differences in looking durations between the left (incorrect) and right (correct) tunnels. Second, independent *t*-tests were conducted to compare looking durations at the right tunnel in the two types of "B" trials between the children with ASD and the TD children. Next, for each group, paired samples *t*-tests were conducted to investigate whether there was a statistical difference in looking durations at the right tunnel between "B" trials with shorter and longer waiting periods. To account for multiple tests, a Bonferroni correction was applied. An adjusted alpha level of 0.01 for statistical significance was used. Last, to investigate the relationship between NVMA and the average looking durations in the two groups during the two "B" trial types, a series of Spearman's correlation analyses were performed. One child from the ASD group was not included in the correlation analysis due to missing NVMA data. The data was analyzed in Statistical Package for the Social Sciences (SPSS) 27.

Results

Group difference on "B" trials with shorter waiting period

The children with ASD spent more time looking at the right tunnel (*M* = 1,961, SD = 1,096) than at the left tunnel (*M* = 1,032, SD = 469) during the B trials with shorter waiting period (**Figure 4**). This difference was statistically significant ($t(12) = -2.92$, $p = 0.013$, 95% CI[236–1,620]). Similarly, the TD children spent more time looking at the right tunnel (*M* = 2,020, SD = 868) than at the left tunnel (*M* = 975, SD = 620) (**Figure 4**). This difference was also significant ($t(23) = -3.98$, $p = 0.001$, 95% CI[502–1,588]). There were no significant differences in looking durations at the right tunnel between the ASD and TD groups ($t(35) = -0.181$, $p = 0.857$, 95% CI[−606–725], Hedges' $g = 0.061$).

Group difference on "B" trials with longer waiting period

The ASD group spent slightly more time looking at the right tunnel (*M* = 1,635, SD = 1,105) than at the left tunnel (*M* = 933, SD = 1,003), but the difference was not statistically significant ($t(12) = -1.95$, $p = 0.074$, 95% CI[−78.65–1,483]). The TD

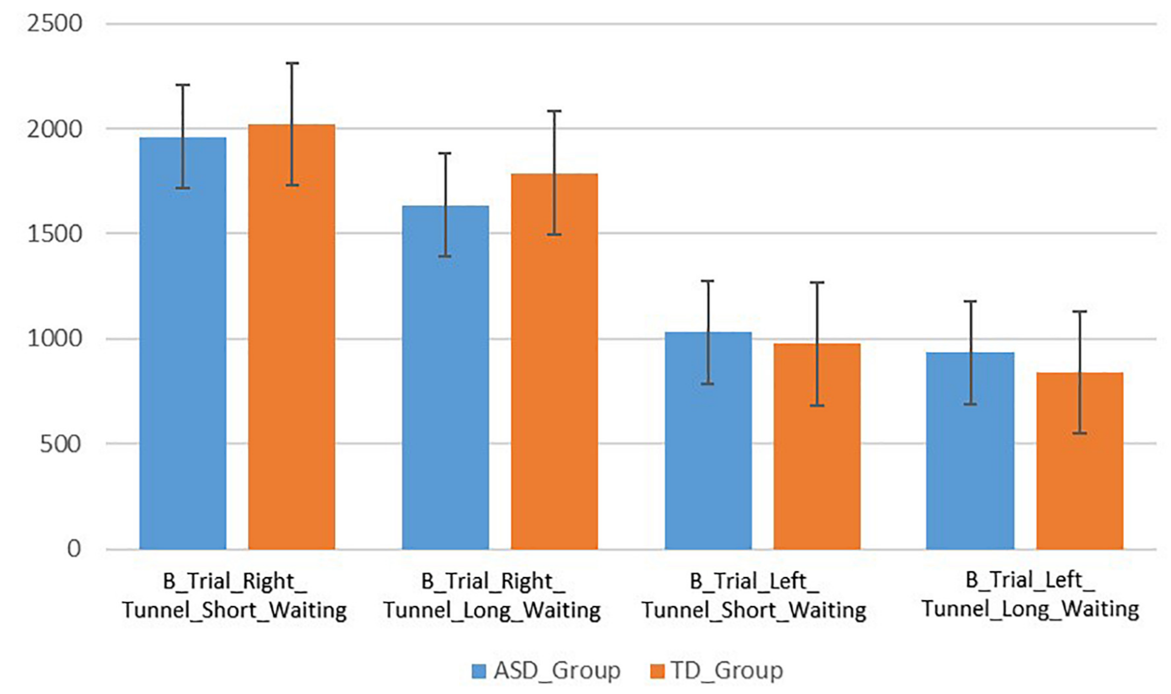


FIGURE 4
Bar graph showing the looking durations at the right and left tunnels in milliseconds on B trials with short and long waiting periods for both groups.

group spent more time looking at the right tunnel (M = 1,788, SD = 992) than at the left tunnel (M = 839, SD = 487). This difference was significant ($t(23) = -3.72$, p = 0.001, 95% CI[421–1,475]). There were no significant differences in looking durations at the right tunnel between the ASD and TD groups ($t(35) = -0.429$, p = 0.671, 95% CI[−569–874], Hedges' g = 0.148).

Within group difference on "B" trials with shorter versus longer waiting period

For the ASD group, there was no statistical significance difference in looking durations at the right tunnel between "B" trials with shorter and longer waiting periods ($t(12) = 1.296$, p = 0.219, 95% CI[−221.99–873.30]). For the TD group, there was no statistical significance difference in looking durations at the right tunnel between "B" trials with shorter and longer waiting periods ($t(23) = 1.466$, p = 0.156, 95% CI[−95.68–560.85]).

Relationship between non-verbal mental age and "B" trials with shorter and longer waiting period

No correlations were found between NVMA and looking durations at the right tunnel for either the children with ASD or TD on the trials with shorter and longer waiting periods (**Table 2**).

Discussion

The current study compared WM performance on an eye-tracking version of the A-not-B task between pre-school children with ASD and TD children. The group comparisons were performed on "B" trials with short and long waiting periods. Average looking durations were used as dependent variable. On average, both the ASD and TD groups spent significantly more time looking at the target location (right vs. left tunnel) during "B" trials with short waiting periods. Only the TD groups spent significantly more time looking at the target location during long waiting period. There were no significant differences in average looking durations to the right tunnel between the two groups during the tasks with either the short or the long waiting period. Moreover, neither group exhibited significant difference between looking durations at the target location during shorter vs. longer waiting period. This suggest that the introduction of a longer waiting period had no effect on looking durations at the target location. The results are contrary to the initial expectation that children with ASD would exhibit shorter looking durations at the target location than the TD group on both "B" trial types, and that the group differences would be larger in the condition with longer waiting period due to its higher WM demand. Complementary to the study by Fanning et al. (2018) that also found no significant differences between 2- and 5-year-old TD children and children with ASD on the eye-tracking A-not-B task, neither the addition of distractors nor a condition with longer waiting period in the current study lead to the identification of differences between the groups.

TABLE 2 Correlations between non-verbal mental age (NVMA) and looking durations for "B" trial with short and long waiting period for both groups.

	ASD ($n = 12^1$)	**TD (n = 24)**
Variable	**NVMA**	**NVMA**
«B» Short Waiting Period–Target Looking	−0.371	−0.017
«B» Long Waiting Period–Target Looking	−0.046	−0.052

[1] Data from one of the 13 children missing.

The current study also investigated the relationship between NVMA and average looking durations during the "B" trials with short and long waiting periods between the ASD and TD groups. As expected, the children in the ASD group had substantially lower NVMA as compared to the TD group. Still, correlations between NVMA and performance on "B" trials were not identified in either group.

The lack of group differences and correlation to NVMA could be related to children's developmental level, which possibly have exceeded the task's age range for which it is suited. The A-not-B task was previously used with children younger (15–30 months) than the current sample (Diamond et al., 1997). Studies by Griffith et al. (1999) and McEvoy et al. (1993) had samples with ages ranging from 40 to 80 months old performing the modified manual A-no-B tasks. However, no group differences were found in those studies. In the current sample, the mean NVMA of the ASD group was 31.5 months and the mean NVMA of the TD group was 48.9. Although an attempt was made in increasing tasks difficulty by implementing distractors and different waiting periods, the NVMA of children in the current study might be above the task's targeted age range.

Although eye-tracking provides the benefit of acquiring data from children who are minimally verbal and/or developmental delayed, it is important to ensure that the task targets the participants' developmental level. The aforementioned study by Valeri et al. (2020) in which 4–6-year old children with ASD displayed no impairment on Spinning Pots task which was proposed to be suitable for children between 18 and 42 months could act as an example of the importance of task selection. Additionally, in a study by Zacharov et al. (2021) pre-school children with ASD were administered two cognitive flexibility tasks designed for pre-schoolers. These tasks had different difficulty levels. On the more difficult task the children exhibited impairment, while on the other their performance was intact.

Small sample size is one of the main weaknesses of the current study. Thus, the interpretation of the results should be done with caution. Additionally, the gender ratio of the ASD sample is skewed, hence no potential gender differences could be investigated. However, the current study demonstrated that it is possible to measure WM with a cartoon-like task consisting of different WM loads using eye-tracking technology. Moreover, the current study further illuminated the important issue of task appropriateness when studying EF in pre-school children with ASD. The current task may be suitable for even younger children with ASD and TD. Hence, in future studies employing a similar eye-tracking version of the A-not B task, it is recommended to recruit younger participants than in the current sample.

Data availability statement

The datasets presented in this article are not readily available due to restrictions related to ethical regulations. Requests to access the datasets should be directed to AK, anett.kaale@isp.uio.no.

Ethics statement

The studies involving human participants were reviewed and approved by the Regional Committees for Medical and Health Research Ethics (REC), South East Norway. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

OZ conducted the data collection. All authors contributed in writing the manuscript, took part in the design and analyses of this study, and approved the submitted version.

Funding

This research was supported by grants from the Department of Special Needs Education and Faculty of Educational Science at University of Oslo, Norwegian Centre of Expertise for Neurodevelopmental Disorders and Hypersomnias at Oslo University Hospital, and the research group Communicative Processes (ComPros) at Department of Special Needs Education at University of Oslo.

Acknowledgments

We gratefully acknowledge the research assistants who contributed in the data collection, the children and the families who participated in this research and the pre-schools staff that welcomed us into the schools for testing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders*, 5th Edn. Arlington, VA: American Psychiatric Association. doi: 10.1176/appi.books.9780890425596

Anderson, J. R. (2014). *Cognitive psychology and its implications*, 8 Edn. New York, NY: Worth Publishers.

Diamond, A., Prevor, M. B., Callender, G., and Druin, D. P. (1997). Prefrontal cortex cognitive deficits in children treated early and continuously for PKU. *Monogr. Soc. Res. Child Dev.* 62, 1–206. doi: 10.2307/1166208

Edmunds, S. R., Colman, C., Vidal, P., and Faja, S. (2021). Brief report: Examining the links between language processes and working memory impairments in toddlers and preschoolers with ASD. *J. Autism Dev. Disord.* 52, 1872–1880. doi: 10.1007/s10803-021-05049-x

Fanning, P., Hocking, D. R., Dissanayake, C., and Vivanti, G. (2018). Delineation of a spatial working memory profile using a non-verbal eye-tracking paradigm in young children with autism and Williams syndrome. *Child Neuropsychol.* 24, 469–489. doi: 10.1080/09297049.2017.1284776

Friedman, L., and Sterling, A. (2019). A review of language, executive function, and intervention in autism spectrum disorder. *Semin. Speech Lang.* 40, 291–304. doi: 10.1080/13854046.2017.1290139

Gardiner, E., Hutchison, S. M., Müller, U., Kerns, K. A., and Iarocci, G. (2017). Assessment of executive function in young children with and without ASD using parent ratings and computerized tasks of executive function. *Clin. Neuropsychol.* 31, 1283–1305. doi: 10.1080/09297049.2013.857650

Garon, N., Smith, I. M., and Bryson, S. E. (2014). A novel executive function battery for preschoolers: Sensitivity to age differences. *Child Neuropsychol.* 20, 713–736. doi: 10.1002/aur.1893

Garon, N., Smith, I. M., and Bryson, S. E. (2018). Early executive dysfunction in ASD: Simple versus complex skills. *Autism Res.* 11, 318–330. doi: 10.1111/1467-8624.00059

Griffith, E. M., Pennington, B. F., Wehner, E. A., and Rogers, S. J. (1999). Executive functions in young children with autism. *Child Dev.* 70, 817–832. doi: 10.1007/s40474-019-00158-w

Hamner, T., and Vivanti, G. (2019). Eye-tracking research in autism spectrum disorder: What are we measuring and for what purposes? *Curr. Dev. Disord. Rep.* 6, 37–44. doi: 10.1186/s11689-015-9111-z

Hill, A. P., van Santen, J., Gorman, K., Langhorst, B. H., and Fombonne, E. (2015). Memory in language-impaired children with and without autism. *J. Neurodev. Disord.* 7:19. doi: 10.1542/9781610024716-part01-ch002

Hyman, S. L., Levy, S. E., Myers, S. M., and Council On Children With Disabilities, Section On Developmental and Behavioral Pediatrics (2020). Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics* 145:e20193447. doi: 10.1111/j.1469-7610.1993.tb01036.x

McEvoy, R. E., Rogers, S. J., and Pennington, B. F. (1993). Executive function and social communication deficits in young autistic children. *J. Child Psychol. Psychiatry* 34, 563–578.

Mullen, E. M. (1995). *Mullen scales of early learning*, AGS Edn. Circle Pines, MN: American Guidance Service Inc. doi: 10.3389/fpsyt.2016.00204

Mungkhetklang, C., Bavin, E. L., Crewther, S. G., Goharpey, N., and Parsons, C. (2016). The contributions of memory and vocabulary to non-verbal ability scores in adolescents with intellectual disability. *Front. Psychiatry* 7:204. doi: 10.1016/j.cogdev.2017.02.003

Pellicano, E., Kenny, L., Brede, J., Klaric, E., Lichwa, H., and McMillin, R. (2017). Executive function predicts school readiness in autistic and typical preschool children. *Cogn. Dev.* 43, 1–13.

Rutter, M., Bailey, A., and Lord, C. (2003). *The social communication questionnaire manual.* Los Angeles, CA: Western Psychological Services. doi: 10.1080/15248372.2018.1439493

Stephens, R. L., Langworthy, B., Short, S. J., Goldman, B. D., Girault, J. B., Fine, J. P., et al. (2018). Verbal and nonverbal predictors of executive function in early childhood. *J. Cogn. Dev.* 19, 182–200. doi: 10.1007/s10803-019-04102-0

Valeri, G., Casula, L., Napoli, E., Stievano, P., Trimarco, B., Vicari, S., et al. (2020). Executive functions and symptom severity in an Italian sample of intellectually able preschoolers with autism spectrum disorder. *J. Autism Dev. Disord.* 50, 3207–3215. doi: 10.1007/s11065-016-9336-y

Wang, Y., Zhang, Y. B., Liu, L. L., Cui, J. F., Wang, J., Shum, D. H., et al. (2017). A meta-analysis of working memory impairments in autism spectrum disorders. *Neuropsychol. Rev.* 27, 46–61.

Watanabe, H., Forssman, L., Green, D., Bohlin, G., and von Hofsten, C. (2012). Attention demands influence 10- and 12-month-old infants' perseverative behavior. *Dev. Psychol.* 48, 46–55. doi: 10.1037/a0025412

World Health Organization [WHO] (1993). *The ICD-10 classification of mental and behavioural disorders.* Geneva: World Health Organization.

Zacharov, O., Huster, R. J., and Kaale, A. (2021). Investigating cognitive flexibility in preschool children with autism spectrum disorder. *Front. Psychol.* 12:737631. doi: 10.3389/fpsyg.2021.737631

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 11 October 2022
DOI 10.3389/fnins.2022.898526

OPEN ACCESS

EDITED BY
Hong Xu,
Nanyang Technological University,
Singapore

REVIEWED BY
Diankun Yu,
University of California, San Francisco,
United States
Chen Yan,
The Second Affiliated Hospital
of Guangzhou Medical University,
China

*CORRESPONDENCE
Sheng-yong Bao
szsrmkf@163.com
Wen-wen Cheng
525587688@qq.com
Zhuo-ming Chen
sangshunan@163.com

†These authors have contributed
equally to this work

SPECIALTY SECTION
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

RECEIVED 17 March 2022
ACCEPTED 30 June 2022
PUBLISHED 11 October 2022

CITATION
Cui L-y, Cheng W-w, Shan S-r, Lv W, Sun C-m, Li R, Zhou S, Chen Z-m and Bao S-y (2022) Spontaneous quantitative processing in Chinese singular and plural picture naming: An event-related potentials analysis. *Front. Neurosci.* 16:898526. doi: 10.3389/fnins.2022.898526

Spontaneous quantitative processing in Chinese singular and plural picture naming: An event-related potentials analysis

Li-yan Cui[1,2]†, Wen-wen Cheng[3]*†, Sha-rui Shan[4], Wen Lv[1], Chen-ming Sun[1], Run Li[1], Shu Zhou[5], Zhuo-ming Chen[2]* and Sheng-yong Bao[1]*

[1]Department of Rehabilitation Medicine, Shenzhen People's Hospital, The Second Clinical Medical College, Jinan University, The First Affiliated Hospital, Southern University of Science and Technology, Shenzhen, Guangdong, China, [2]Department of Rehabilitation Medicine, The First Affiliated Hospital of Jinan University, Guangzhou, China, [3]Department of Neurology, Maoming People's Hospital, Maoming, China, [4]Department of Rehabilitation Medicine, The First Affiliated Hospital of Guangdong Pharmaceutical University, Guangzhou, China, [5]Department of Neurology, Nanfang Hospital, Southern Medical University, Guangzhou, China

Chinese nouns lack inflection and cannot reflect the quantitative relationship between singular and plural numbers. However, neural processes of picture naming are different from those of words. We assume that Chinese single and plural picture naming is different, and they may involve quantitative processing. Therefore, Experiment 1 was designed by picking picture naming as the task and Chinese as the target language and compared the accuracy, reaction time, and event-related potentials (ERPs) between single and plural picture naming, where two types of pictures were mixed. Although the *T*-test showed no significant differences in behavioral data, there were differences in ERPs. ERP differences involved two effects: P1 of 160–180 ms and P2 of 220–260 ms in the parietal-occipital lobe. These differences are suggested to reflect the neural differences in quantitative processing. Therefore, Chinese singular and plural picture naming consists of word production and implicit quantitative processing simultaneously. To explore the relationship between the two processings, we added a semantic factor (inanimate vs. animate items) to the quantity factor of Experiment 1 and carried out Experiment 2, with the observation indexes unchanged. There were no significant differences in behavioral data among the four conditions. After variance analysis, ERPs results indicated an interaction between semantic and quantitative factors in the central area at 180–280 ms. In summary, we suggest that Chinese singular and plural picture naming includes two simultaneous neural processing tasks: word production and quantitative processing, which interact in the central area at 180–280 ms.

KEYWORDS

picture naming, singular, plural, word production, quantitative processing, event-related potentials (ERP), Chinese

Introduction

Pictures become symbols of their objects by physical similarity. Several studies propose that recognizing images involves consistent cognitive processes as are applied when perceiving real objects (Potter, 1979; Glaser, 1992). And picture naming comprises the same process of conceptually driven word production (Glaser, 1992; Bock and Levelt, 1994). Numerous studies have investigated the neuropsychological process of picture naming and word production using various methods, including functional magnetic resonance imaging (fMRI) (van Turennout et al., 2000), magnetoencephalography (MEG) (Salmelin et al., 1994), positron emission tomography (PET) (Papathanassiou et al., 2000), electroencephalogram (EEG) (Hassan et al., 2015), etc. It is evidenced that picture naming and word production share the same temporal and spatial signatures (Indefrey and Levelt, 2004). Most studies support the "Lemma model" of lexical access proposed by Levelt et al. (1999), which divides word production into six steps: visual extraction, lexical concept formation, lemma selection, phonological coding, phonological words coding, and pronunciation. Therefore, picture naming is an important experimental paradigm in cognitive psychology.

A recent study, which tracked spatiotemporal dynamics networks of picture naming cognitive activity, applied EEG source connectivity analysis to further elucidate the "Lemma model" (Hassan et al., 2015). From the sight of pictures to the completion of articulation, picture naming comprises six brain network states (BNSs). BNS1 (0–119 ms), mainly involving the inferior occipital, is related to visual feature extraction. BNS2 (120–150 ms), primely comprising occipital regions, is responsible for visual information process and object recognition. BNS3 (151–190 ms) indicates lexical retrieval, lemma retrieval, and lemma selection occurrence at the occipital and bilateral inferior temporal sulcus. BNS4 (191–320 ms) spreads to the left inferior temporal gyrus for integrating access to phonological forms. BNS5 (321–480 ms), mainly involving precentral, is responsible for phonetics and articulation. In addition to articulation, BNS6 (481–535 ms) may be related to introspection over the left insular gyrus.

Based on most studies on the neuropsychological process of picture naming were carried on a single object/image, it is generally believed that picture naming is equivalent to the word production process of the "Lemma model." However, in daily life, objects are more often presented in a plural form than in a singular form (Schiller and Caramazza, 2003). At present, from the perspective of neuro mechanism, there is no research focusing on the naming of singular and plural pictures. Only a few articles have studied the differences between singular and plural pictures in behavioral experiment, but the argument lies in the activation of lexico-syntactic (Khwaileh et al., 2015; Beyersmann et al., 2018). Several articles have studied the quantitative concept of singular and plural pictures with a congruent or incongruent quantifier, founding that the number feature of pictures is different from the word production process and they were not a competitive relationship (Schiller and Caramazza, 2002; Arcara et al., 2019). Most studies focused on the word production process of singular and plural nouns. The research materials are often phrases or sentences, and the focus is mainly on inflectional grammar (Sahin et al., 2009; Gimenes and Brysbaert, 2016). These studies have shown that the word production processes of singular and plural nouns are different due to inflection. Given these, we wondered whether the procedure of singular picture naming would be different from that of plural picture naming.

Numerical magnitude is an abstract quality of a set. It can be represented in a symbolic or non-symbolic form ("10," "ten," and "●●●●●●●●●●") (Holloway and Ansari, 2010). The singular (the set "1") opposed to the plural, which encompasses all other numbers treated as a whole (the set "2, 3, 4,…"), has a different quantitative meaning (Hodent et al., 2005). Accumulating studies suggest that the ability of non-verbal representation of numerical magnitudes is native (Starkey and Cooper, 1980; Wynn, 1992; Butterworth, 2005), as it is shown that very young infants and newborns can distinguish syllables, moving objects, collections of objects, and simple dots (Antell and Keating, 1983; Wynn, 1992; Wynn et al., 2002; McCrink and Wynn, 2004). A similar phenomenon is also found in animals (Wynn, 1992; Romo et al., 1999). Further research has indicated that this not only involves a perceptual pattern but also arithmetical operations (Wynn et al., 2002; McCrink and Wynn, 2004). Even if continuous variables (such as area and contour length of items) are controlled, in a looking-time procedure on numbers of items, a 5-month baby did not simply expect "more" or "less" than the initial number of items seen but rather expected exactly the correct number of items (for example, perform 1 + 1 or 2 – 1 by Mickey toys, the baby looking more time on the wrong numbers of toys). This supports that the nature of human infants' numerical knowledge is based on the accumulator mechanism (magnitude-based estimation system) proposed by Meck and Church (1983), Gallistel and Gelman (1992), Wynn et al. (2002), and McCrink and Wynn (2004) but not an automatic object-tracking mechanism (Kahneman et al., 1992; Trick and Pylyshyn, 1994).

In 1993, Dehaene and Changeux proposed a classic neural processing model of quantitative processing based on non-symbolic numbers (Dehaene and Changeux, 1993). First, objects of different sizes and positions are input and characterized by the retina. Then, the sizes and positions are standardized through a topological map formed by a fixed group of neurons. Finally, the quantitative detector summarizes all outputs and forms a neural map that is highly correlated with quantity. Verguts and Fias further developed this model and applied it to the processing of symbol numbers (Verguts and Fias, 2004). Both computational models focus on sequentially occurring summation coding and spatial coding, which characterize quantitative information of objects to the quantitative processing neural network. While the summation coding assumes that neuronal activity increases linearly with an increasing number (Nieder and Merten, 2007; Roitman et al., 2007; Santens et al., 2010), the spatial coding assumes that certain neurons are associated with specific numbers, suggesting that neuron activation is logarithmic to number, and they generate maximum activation for the number of preferences (Nieder et al., 2002, 2006; Piazza et al., 2004). It is believed that the selection of the coding methods is task-dependent (Parker and Newsome, 1998; Salinas, 2006). The summation coding may be preferred in the number comparison task (Romo et al., 1999) and the spatial coding may be preferred in the discrepancy comparison task (Nieder and Merten, 2007). Based on these two computational models, computer simulations can account for several phenomena in the numerical domain, including the distance effect and Fechner's law for numbers (Verguts and Fias, 2004). The computer simulations may also demonstrate that human infants and several animal species possess some elementary abilities for numerical processing or calculation, despite the lack of language or task acknowledgment (Verguts and Fias, 2004). Therefore, basic numerical abilities are natural and native. Quantity processing may be widely present in daily life in a subliminal manner when one subject is not aware of having seen objects or a number symbol (Dehaene et al., 1998; Naccache and Dehaene, 2001). Some studies have also found quantitative processing under non-computing tasks (Roitman et al., 2007).

If picture naming is the same as the process of word production. In a language with inflection (such as English), the neural processing of singular and plural picture naming are markedly different because the morphologies and pronunciations of naming have altered, according to the "lemma model" (e.g., basketball vs. basketballs/mouse vs. mice) (Sahin et al., 2009). Interestingly, we found that the word productions of singular and plural pictures in a language without inflection (such as Chinese) were consistent (e.g., Lan Qiu vs. Lan Qiu) (Yu et al., 2013). In other words, Chinese singular and plural picture namings share the same neural processings. However, we know that the essence of inflection in nouns is to indicate the number of objects. And when pictures were presented, we could visually clearly see the number difference of singular and plural objects, even if Chinese lack inflection. Boldly, we doubted that we could even perceive the difference in quantity. Based on these, we proposed a hypothesis: singular and plural picture naming may include both word production and quantitative processing. In other words, singular and plural pictures are still quantitatively processed under the task of naming. According to this theory, the neural processing of singular and plural picture namings are different, at least in quantitative processing. It will bring about a great challenge to the traditional treatment method of picture naming for Chinese aphasia, and even the picture learning method for Chinese children (using singular pictures). And the concept of quantitative processing may also provide a new theoretical direction for further research on the mechanism of picture naming.

The inflection of English would interfere with our judgment on the existence and characteristics of quantitative processing. However, in case neural processes of Chinese singular and plural picture namings are different, it will probably be due to the quantitative processing. Therefore, to prove our hypothesis, Experiment 1 was designed, which used event-related potential (ERP) technology and Chinese picture naming task to compare neural processes of singular and plural pictures. At the same time, we assume that the two processes in one task could not be unrelated, so Experiment 2 was conducted to explore the relationship between them.

Experiment 1

Study design

Participants

Twenty young healthy postgraduate students (8 male; aged 22–30 years old, and mean age = 25.2 years old, SD: 2.48 years) from Jinan University took part in this experiment as paid participants. These subjects were right-handed, native Chinese speakers. All reported having no previous history of neurological, reading, or learning disorders. All had normal or corrected-to-normal vision. All participants signed written informed consent after all the experimental procedures were fully explained. The study was approved by the Medical Ethics Committee of the First Affiliated Hospital of Jinan University.

Stimuli

The linguistic stimuli were 66 concrete, countable, inanimate, and different objects, which were adopted from the Ni'picture database (Ni et al., 2019; **Supplementary Appendix 1**). They consisted of two types of pictures: 33 pictures representing one object (singular pictures) and 33 pictures representing three objects (plural pictures). The properties of the two types of items were matched in object familiarity, visual complexity, name agreement, image agreement, image variability, age of acquisition, and word frequency (see **Table 1**). The 66 pictures from Ni'picture database were edited by Adobe Photoshop CC 2018. They were all set 1,000 × 1,000 pixels, not changing the color and distinguishability of objects. The size of an object in a singular picture was 30,000 pixels, and each size of the object in a plural picture was 10,000 pixels. Objects were placed randomly in both types of pictures (see **Figure 1**).

The reason why we chose three objects was to exclude confounding factors. As many studies have shown that there are significant neural differences between small (<3 or 4) and large numbers (>3 or 4) (Jevons, 1871; Piazza et al., 2003), which may be related to attention (Sophiana and Crosby, 2008) and

TABLE 1 Mean (SD) scores for the list of 132 pictures; 33 singular-inanimate (Q–S–), 33 plural–inanimate (Q + S–), 33 singular-animate (Q–S +), and 33 plural-animate (Q + S +).

Properties Type	Q–S–	Q + S–	Q–S +	Q + S +	*F*	*P*
N-A	0.78(0.18)	0.76(0.18)	0.68(0.20)	0.75(0.18)	1.826	0.146
Im-A	3.97(0.51)	4.03(0.67)	3.80(0.53)	3.89 (0.53)	0.991	0.399
Fam	3.11(0.65)	3.09(0.80)	3.09(0.93)	3.15(0.81)	0.044	0.988
Vi-C	2.44(0.94)	2.47(0.93)	2.70(0.68)	2.64(0.61)	0.828	0.481
Im-V	2.13(0.65)	2.20(0.74)	1.97(0.70)	2.01(0.55)	0.845	0.472
A-o-A	3.99(0.69)	3.81(0.86)	3.74(0.80)	3.62(0.70)	1.308	0.275
Wo-F	80.61(7.89)	76.52(9.17)	81.22(7.92)	79.83(9.14)	1.979	0.120

N-A, Naming Agreement; Im-A, Image agreement; Fam, Familiarity; Vi-C, Visual Complexity; Im-V, Image variability; A-o-A, Age of Acquisition; Wo-F, Word Frequency; SD, Standard deviation.

visuospatial working memory (Luck and Vogel, 1997). If the plural number selection is greater than 3 or 4, the final result may be affected by differences in neural processing of small and large numbers. Therefore, in this study, 3 which can be identified at a glance was selected as the plural number (such as l vs. 3).

Procedure

In a sound-attenuated dimly lit chamber, the participants were put on an electrode cap of Ag-AgCl and sat about 120 cm away from a 23-in computer monitor. Eyes were on the same horizontal line as the center of the screen, avoiding excessive eye movements. Stimuli were presented against a dark gray background by the MindXP software developed by our lab, and participants were asked to name the objects rapidly. Meanwhile, the voice by a microphone and EEG were recorded. Before the experiment, participants were pre-tested to ensure that they knew the exact name of the 66 objects. Additional 5 pictures were arranged as a pre-experiment to familiarize participants with the experimental process.

The experiment consisted of two blocks and continued for 6.46 min. One block required approximately 2.7 min to display all the 66 different pictures. One minute was set for a rest between blocks. That is to say, this experiment contains a total of 132 trials, single and plural pictures were 66 trials, respectively. As **Figure 1** presented a block began with a fixation cross displayed in the center of the screen for 500 ms. Then pictures for naming were displayed for 1,200 ms, followed by blank screens for a random duration between 1,000 and 1,500 ms to avoid psychological expectations. The 66 pictures were represented in pseudo-random orders: three pictures with similar pronunciation or the same type (singular or plural) did not appear consecutively.

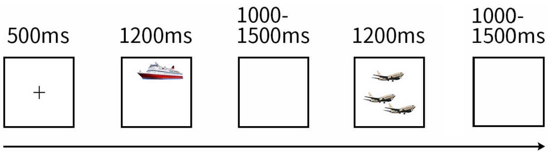


FIGURE 1
Task Design of Experiment 1. The picture shows the design of the experimental task. All trials followed the depicted sequence. A block began with a fixation cross a picture. Then a singular/plural picture was displayed and participants were asked to name the objects rapidly after the stimuli were presented, followed by a blank screen (three pictures with similar pronunciation or the same type (singular/plural) did not appear consecutively).

Electrophysiological recordings

The EEG recording system was provided by Nanfang Hospital, Southern Medical University, with a 19-channel EEG amplifier (Symptom Instrument®). It used an international 10–20 system with linked earlobes as the reference (FP1, FP2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, and Pz). EEG was continuously recorded at a sampling rate of 1,000 Hz. Recording bandwidth was set at 0.5 to 100 Hz. Electrode impedances were kept below 10 kΩ.

Data analyses

Behavioral analyses

Accuracy and reaction times (RTs) were recorded for each participant by a vocal response using Cool Edit Pro 2.1. The error picture naming included no response (including unnamed and RT over 1,200 ms), word error, and fluency error. And the mean RTs were calculated based on the correct trials. Data were compared between the singular and plural groups using two-tailed paired *t*-tests. Data analysis was performed by SPSS 22.0 software.

Event-related potentials analyses

MindWave-sorting software and statistical parametric mapping (SPM) software developed by our lab were used for ERP analyses (application in literature, Zhou et al., 2004, 2019; Cheng et al., 2021). MindWave-sorting software was used for the pre-processing of the EEG data, including automatic correction and ERP extraction. After detection of the ocular, muscular, and any other artifacts at the threshold of ± 70 μV, MindWave-sorting software automatically corrected the EEG signal using principal component analysis (Lins et al., 1993a,b). Then, the epochs were segmented, ranging from −100 ms to 600 ms after stimulus onset, with a baseline correction (using the mean amplitude of 100 ms pre-stimulus interval). Here two ERPs were obtained (singular and plural ERPs) in 19 channels. SPM software was used to obtain the average waveform for each ERP. A pairwise comparison for the two ERPs was performed using two-tailed paired *t*-tests, where correction for multiple testing on the 19 channels was based on the false discovery rate procedure (FDR, Benjamini and Yekutieli, 2001; Lage-Castellanos et al., 2010). And the differences were presented as a topographical map using an interpolation method relevant to a generalized cortical imaging technique (Zhou et al., 1998). A fixed sliding step of 20 ms without overlapping data was set for the topographical map. And 0.05 was set as the significance threshold.

Results

Behavioral data

The overall accuracy was near ceiling (94.98 ± 2.02%) for a total of 132 stimuli. The mean RTs from the onset of pictures to the pronunciation was 612.23 ± 80.02 ms. Specific descriptive statistics of the accuracy and RTs were presented in **Table 2**, and the paired *t*-test showed no significant differences between the two types of picture naming in both the accuracy and RTs. The error rates (in%) were reported in **Supplementary Table 1**.

Waveform and component analysis

The grand-average ERPs time-locked to the content word (from −100 to 600 ms) for all the 19 electrodes (FP1, FP2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, and Pz) were shown in **Figure 2**. Two phases (P1 effects of 160–180 ms and P2 effects of 220–260 ms in parietal-occipital lobe) showed significant differences in the waveform, but they were consistent after 300 ms. In the first phase (P1 effects), the plural type exhibits greater average amplitudes than the singular type at O1, O2, P4, C4, and T4 electrodes in the range of 160–180 ms. In the second phase (P2 effects), the waveform showed a higher positive average amplitude in the singular type than the plural type at O1, O2, and P3 electrodes within the range of 220–260 ms. All the specific statistics at typical electrodes within a 20-ms time window were shown in **Table 3**.

TABLE 2 Behavioral performance summary in Experiment 1 (mean ± SD) (*N* = 20).

Behavioral performance	Singular picture naming	Plural picture naming	*t*	*p*
Accuracy (%)	95.05 ± 1.99	94.90 ± 2.10	0.23	0.818
Reaction times (ms)	597.10 ± 79.10	627.36 ± 80.01	−1.20	0.236

Spatiotemporal pattern: SPM (t)

Figure 3 showed topographical maps of SPM (t) (0–600 ms) derived from two-tailed paired *t*-tests. The red/bright blue bin of the color scale corresponded to the 0.05 significance threshold: $t_{(1,19)} = \pm 2.09$; the white dots on the maps represented the electrode sites with significant effects. The two types of pictures initially showed differences in neural processing at the parietal-occipital lobe of 160–180 ms. As the neural processing progressed, they differed at 220–260 ms in the parietal-occipital lobe again.

Discussion

Stimulus items were matched for familiarity and naming accuracy of image materials of the two sets of pictures. As anticipated, naming accuracy was found to be very high for singular and plural pictures, with no significant difference. Besides, no significant difference was found in the RTs between the two types of pictures. However, based on ERP results, the two types of pictures showed differences in amplitude in two effects (parietal-occipital P1 and P2) within the first 300 ms of picture naming. Therefore, using behavioral data to determine the difference between singular and plural picture naming may be insufficient.

To the best of our knowledge, this is the first study to compare neural processes of Chinese singular and plural picture naming. Given that word production of Chinese singular and plural nouns is consistent, we suggest that ERP differences between the two types of pictures (P1 and P2 effects) may represent the differences in quantitative processing (singular vs. plural). First, study proved that the EEG signal was not contaminated by speech artifacts up to 100 ms before articulation (Fargier et al., 2017). In this study, the ERP differences were in 0–300 ms, while pronunciation was around 650 ms, so P1 and P2 effects would not be interfered by pronunciation. Second, the P1 effect was not related to visual perception. Visual perception is an exogenous component and generally differs within 100 ms, while the P1 effect was after 100 ms. Third, ERP differences cannot be explained by the differences in singular and plural word production. In word production, 160–180, and 220–260 ms periods represent lexical retrieval, lemma selection, and phonological coding, respectively (Hassan et al., 2015). Accordingly, the activated brain areas gradually transition from the back to front, from the occipital lobe to the inferior temporal sulcus and the frontal area. However, in the present study, both effects of ERPs differed in the parieto-occipital lobe. Fourth, the ERP differences were not inflection. Inflection occurred in the left inferior frontal gyrus

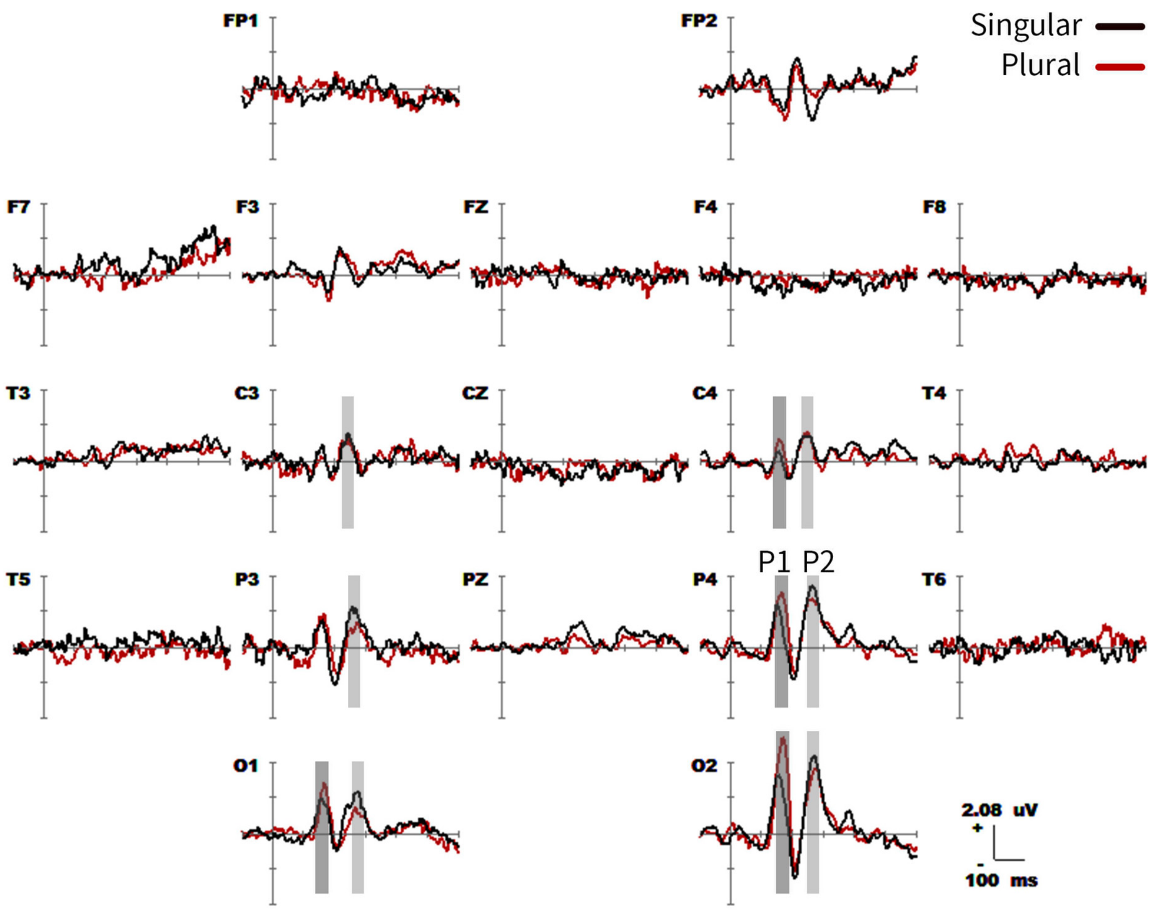


FIGURE 2
Grand average event-related potential (ERP) waveforms (from −100 to 600 ms) are shown for 19 electrodes across singular (black traces) vs. plural (red traces) picture naming, from the 20 subjects. The baseline ERP measurement is the mean amplitude of a 100-ms pre-stimulus interval.

TABLE 3 Significant waveform effects in 19 channels within a 20-ms time window in Experiment 1 (N = 20).

Effect	**P1 (O2)**		**P1 (P4)**		**P1 (C4)**		**P1 (T4)**		**P1 (O2)**	
	Stat.	***p*/WO**	**Stat.**	***p*/WO**	**Stat.**	***p*/WO**	**Stat.**	***p*/WO**	**Stat.**	***p*/WO**
t/p	−3.52	0.002	−2.79	0.012	−2.37	0.034	−2.66	0.017	−2.16	0.044
Cohen’s d/WO	−1.62	160	−1.28	160	−1.09	160	−1.22	160	−0.99	180
Effect	P2 (O1)		P2 (O1)		P2 (O2)		P2 (P3)		P2 (O1)	
	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO
t/p	2.75	0.013	2.15	0.049	2.53	0.020	2.44	0.026	2.14	0.046
Cohen’s d/WO	1.26	220	0.99	240	1.16	240	1.12	240	0.98	260

Significant waveform effects, electrodes and time windows (20 ms interval as a time window) with significant amplitude differences of singular vs. plural ERPs within 0–600 ms, and their corresponding maximum statistics (t, *p*, and Cohen’s d); P1 (O2), P1 effect (electrode); P2 (O2), P2 effect (electrode); Stat., statistics; p, FDR-corrected *p*-value of paired *T*-test; WO, window set (at 20 ms intervals). WO, window onset.

at 280–400 ms, with a peak at 320 ms (Sahin et al., 2009). In this study, the ERPs of the two types of pictures showed no difference after 260 ms. It also suggests that Chinese picture naming lacks inflection processing. Last but not least, the ERP differences in this study were basically consistent with the previous literature on the time course, amplitude features, and activated brain regions of quantitative processing ERPs (Libertus et al., 2007; Pinhas et al., 2015). The first ERP difference (P1 effect of 160–180 ms in parietal-occipital lobe) had a larger amplitude as a larger quantity, which conforms to the characteristics of the

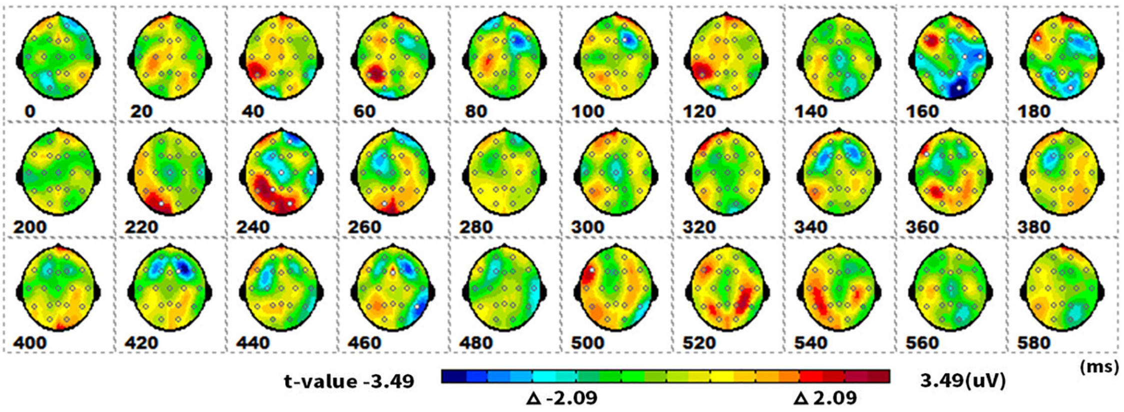


FIGURE 3
The spatiotemporal patterns of SPM (t) (0 to 600 ms) are derived from the two-tailed paired *t*-tests (singular vs. plural). Each map was interpolated from the average *t*-values within the fixed 20-ms time window, and the red/bright blue bin of the color scale corresponded to the 0.05 significance threshold: $t_{(1,19)} = \pm 2.09$. The white dots represented the electrode sites with significant effects.

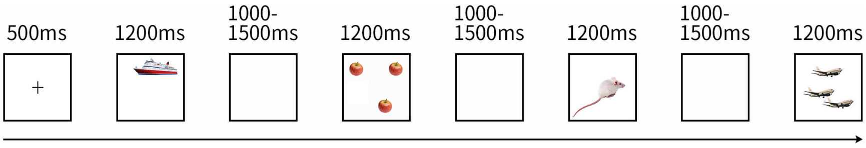


FIGURE 4
Task Design of Experiment 2. The picture shows the design of the experimental task. All trials followed the depicted sequence. A block began with a fixation cross a picture. Then a singular – inanimate/plural – animate/singular – inanimate/plural – animate picture was displayed and participants were asked to name the objects rapidly after the stimuli were presented, followed by a blank screen (three pictures with similar pronunciation or the same quantity (singular vs. plural)/semantics (animate vs. inanimate) did not appear consecutively).

summation coding. The second ERP difference (P2 effect) was in the parieto-occipital lobe at 220–260 ms and corresponded to the spatial coding. Singular images got larger amplitudes. The reason may be that in the daily life the picture naming is always based on a singular picture. And the number one was closer to the participants' psychological preference number (Libertus et al., 2007; Nieder and Merten, 2007; Pinhas et al., 2015). However, the summation coding was a positive effect in our study and a negative effect in the previous literature (Libertus et al., 2007; Pinhas et al., 2015). We considered that the difference in polarity of this effect is due to the different experimental tasks (picture naming vs. counting task). Picture naming affected the ERP waveform of summation coding.

TABLE 4 Behavioral performance summary in Experiment 2 (mean ± SD) (*N* = 24).

	Reaction times (ms)		Accuracy (%)	
	Singular	Plural	Singular	Plural
Animate pictures	644.56 ± 112.80	667.33 ± 105.86	94.71 ± 1.27	95.46 ± 1.77
Inanimate pictures	640.34 ± 127.18	644.46 ± 116.40	94.67 ± 1.55	94.63 ± 1.79

The neural processes of Chinese singular and plural picture naming are different. It supports our hypothesis that in the act of singular and plural picture naming, there are two simultaneous neuropsychological processes: word production and quantitative processing. Meanwhile, both neural processes showed electrical activity in the parieto-occipital lobe at 140–220 ms. Whether this is a mere coincidence or there is a certain connection between these two neural processes warrants further investigation. Therefore, Experiment 2 was designed to explore this issue.

Experiment 2

According to the "Lemma model," picture naming goes through two psychological processes from visual feature extraction to lexical concept formation/semantic formation: concept gathering (color, shape, movement, motion features, hearing, smell, taste, etc.) and viewpoint selection (relative relation within objects). The relative relationship can be orientational, quantitative, etc. That is, semantic formation contains quantitative information. Conversely, quantity differences of picture naming can also be expressed in semantics and vocabulary (for example, basketball vs. basketballs). Therefore, it is reasonable to hypothesize that semantic and quantitative information in singular and plural picture naming may have a certain connection. Based on this, they all showed electrical activity in the parieto-occipital lobe at 140–220 ms, corresponding to the semantic formation and summation encoding, respectively. Hence, Experiment 2 was designed by taking Chinese as the target language and controlling semantics (S) and quantity (Q) as two factors, with two levels: inanimate (S−) vs. animate (S +) and singular (Q−) vs. plural (Q +), to explore the correlation between semantic and quantity factors using a 2 × 2 variance analysis.

Study design

Participants

According to the standard of Experiment 1, another 25 postgraduate students (12 male; aged 18–29 years old, and mean age = 24.1 years old, SD: 3.09 years) completed the experiment.

Stimuli

Following the method in Experiment 1, 132 pictures with different objects were prepared. Four conditions (S−Q−, S + Q−, S−Q +, S + Q +) each had 33 pictures. And the inanimate singular and plural pictures (S−Q−, S−Q +) were exactly the same as in Experiment 1.

Procedure

The experiment was performed as previously described in Experiment 1. The task was still to name the pictures as quickly as possible. The experiment consisted of 264 trials, with 66 trials for each of the four conditions, and divided into 4 blocks (2.7 min each). The stimulus was shown in **Figure 4**, using the same method as Experiment 1, but with the added restriction that no more than three inanimate or animate pictures could be seen sequentially.

Electrophysiological recordings

Electrophysiological recordings were the same as in Experiment 1.

TABLE 5 Two-factor ANOVA of repeated measures of behavioral data (*N* = 24).

	Reaction times		Accuracy	
	F	***P***	***F***	***p***
Semantic	0.328	0.568	1.777	0.186
Quantitative	0.323	0.571	1.164	0.283
Interaction	0.155	0.694	1.454	0.231

Data analyses

Behavioral analyses

The accuracy and RTs of picture naming were analyzed by a repeated-measures ANOVA.

Event-related potentials analyses

The software and methods of ERP processing and analysis were similar to those applied in Experiment 1. Since it was likely established in Experiment 1 that quantitative processing occurred before 300 ms, semantic processing of naming also ended before this period. The waveforms and topographic maps of ERP in Experiment 2 were intercepted from −100 to 400 ms. Eventually, Experiment 2 had 4 ERPs (S−Q−, S + Q−, S−Q + , S + Q +) in 19 channels, segmented within 0 to 400 ms. The within-subject factors were semantic (animate vs. inanimate) and quantity factors (singular vs. plural). Then a two-way repeated-measures ANOVA was performed on the four variables, with multiple testing on the 19 channels corrected using the FDR procedure (Benjamini and Yekutieli, 2001; Lage-Castellanos et al., 2010). Similarly, the differences were represented by a topographical map with a fixed sliding window of 20 ms, and the white dots on maps indicated significant effects.

Results

Behavioral data

The overall naming accuracy of all 264 stimuli was very high, approximately 94.87 ± 1.61%. The mean RTs were 649.17 ± 113.85 ms. **Table 4** described the specific values of naming accuracy and RTs of the four conditions. And the results of the repeated-measures ANOVA, which indicated that there was no significant difference in behavioral performance among the four conditions, were presented in **Table 5**. The error rates were reported in **Supplementary Table 2**.

Waveform and component analysis

The grand-average waveforms of the four ERPs (−100 to 600 ms) are shown in **Figure 5**. There were differences in the waveforms of singular and plural pictures (whether they are animate or inanimate) in the parieto-occipital lobe (O1, O2, P3, P4, and PZ) at 160–180 ms (P1 effect), and parieto-occipital lobe (O1, O2, P4, and PZ) at 220–260 ms (P2 effect). The difference of waveforms between animate and inanimate pictures (regardless of singular and plural factors) was in the parieto-occipital temporal lobe (O1, O2, P3, P4, T5, T6, T3, T4, and F7) at 100–140 ms (N1 effect). And the animate pictures got larger N1 than the inanimate pictures. **Table 6** detailed the average statistics of the waveforms at typical electrodes.

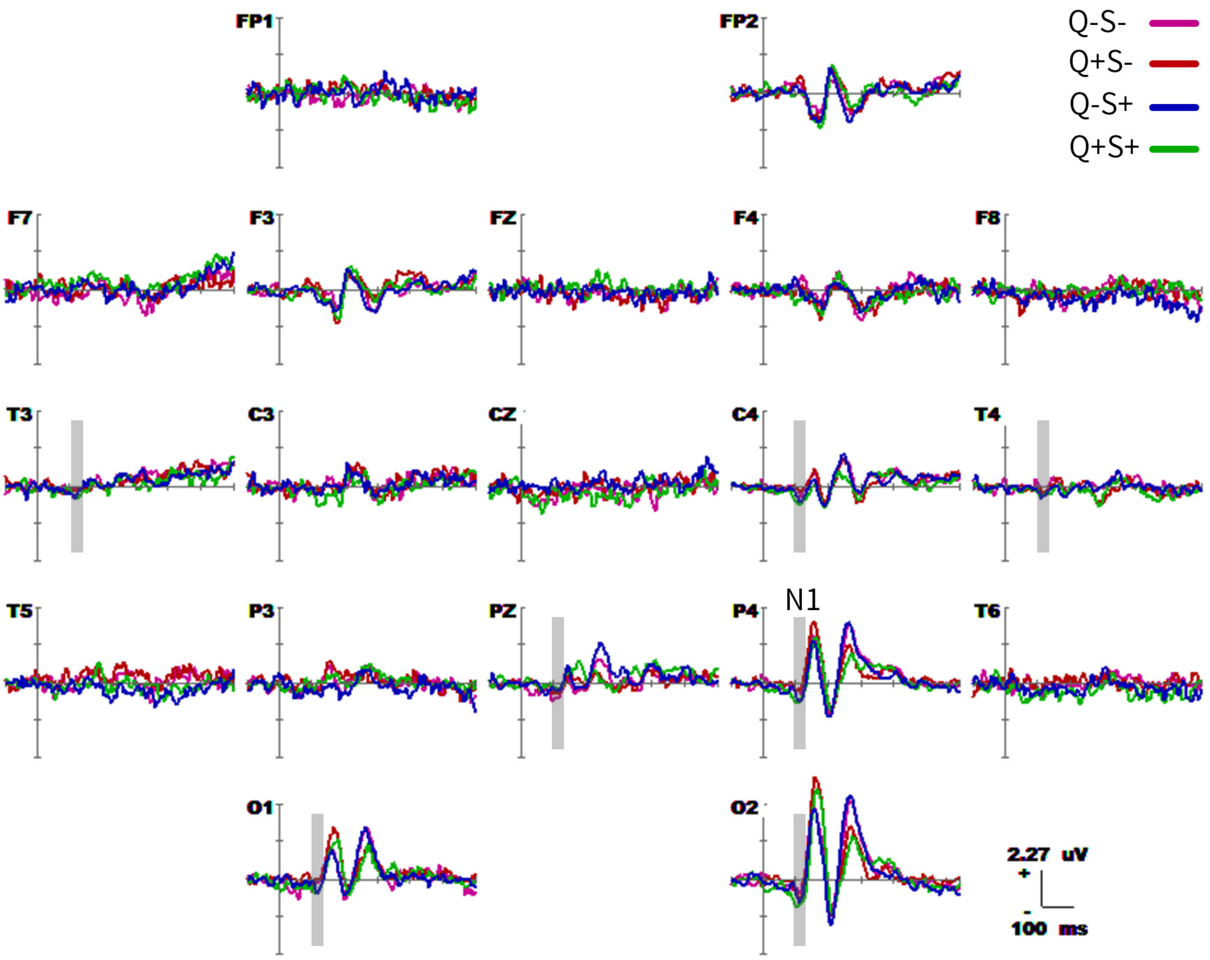


FIGURE 5
Grand average event-related potential (ERP) waveforms (from −100 to 600 ms) are shown for 19 electrodes across four trial types. The purple, red, blue, and green traces correspond to group average ERP of the singular − inanimate (Q−S−), plural- − inanimate (Q + S−), singular − animate (Q−S +), and plural-animate (Q + S +) conditions, respectively. The baseline ERP measurement is the mean amplitude of a 100-ms pre-stimulus interval.

Spatiotemporal pattern: SPM (f)

Figure 6 showed topographical maps of SPM (f) (0 to 400 ms), which were derived from the two-way repeated-measures ANOVA of waveforms. Figure 6A indicated the main effect of quantity. The quantity factor (singular vs. plural) led to significant differences in two ERPs: the first in the parieto-occipital lobe at 160–180 ms, and the second in the parieto-occipital area at 220–260 ms. Figure 6B showed that semantic processing (animate vs. inanimate) mainly induced an ERP difference in the parieto-occipital temporal area at 100–140 ms. Figure 6C revealed that semantic and quantity factors in picture naming have an interactive effect at 180–280 ms in the central area.

The results of the *post hoc* tests were presented as topographic maps in Figure 7. The ERP differences of singular and plural picture naming – the amplitude differences in the parieto-occipital area at 160–180 ms and 220–280 ms were larger in the animate items (S + : Q− – Q +) than that in the inanimate items (S−: Q− – Q +). The plural items (Q + : S− – S +) got greater ERP differences in animate and inanimate picture naming than the singular items (Q−: S− – S +). The ERPs differed not only in the parieto-occipital temporal area at 100–140 ms but also in the parieto-occipital area at 160–180 ms and 220–280 ms.

Discussion

The naming accuracies and RTs of singular vs. plural pictures in Experiment 2 were similar to those in Experiment 1. And semantic and quantitative differences did not have significant effects on behavioral results. Nonetheless, this study revealed that semantic and quantity factors of Chinese singular and plural picture naming interact in the central

TABLE 6 Significant waveform effects in 19 channels within a 20-ms time window in Experiment 2 (N = 24).

Effect	***F*(1,23)/*p***	**P1 (O1)**		**P1 (O2)**		**P1 (Pz)**		**P1 (P3)**		**P1 (P4)**	
	ES/WO	**Stat.**	***p*/WO**	**Stat.**	***p*/WO**	**Stat.**	***p*/WO**	**Stat.**	***p*/WO**	**Stat.**	***p*/WO**
Quantitative	F/*p*	9.67	0.006	11.34	0.003	5.96	0.036	11.32	0.003	10.16	0.004
	η2p/WO	0.30	160	0.33	160	0.21	160	0.33	160	0.31	160
	F(1,23)/*p*	P1 (O1)		P1 (O2)		P1 (P3)		P2 (O1)		P2 (O2)	
	ES/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO
Quantitative	F/*p*	6.64	0.02	4.54	0.044	10.85	0.003	13.36	0.001	24.92	0.000
	η2p/WO	0.22	180	0.16	180	0.32	180	0.37	220	0.52	220
	F(1,23)/*p*	P2 (Pz)		P2 (P4)		P2 (O1)		P2 (O2)		P2 (Pz)	
	ES/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO
Quantitative	F/*p*	20.94	0.010	11.8	0.003	12.3	0.002	23.3	0.000	21.07	0.000
	η2p/WO	0.48	220	0.34	220	0.35	240	0.50	240	0.48	240
	F(1,23)/*p*	P2 (P4)		P2 (O2)		P2 (Pz)		P2 (P4)		P2 (O2)	
	ES/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO
Quantitative	F/*p*	11.5	0.004	10.40	0.010	7.44	0.013	4.91	0.044	4.79	0.041
	η2p/WO	0.33	240	0.31	260	0.24	260	0.18	260	0.17	280
	F(1,23)/*p*	N1 (O1)		N1 (O2)		N1 (P4)		N1 (C4)		N1 (F7)	
	ES/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO
Semantic	F/*p*	10.59	0.003	5.90	0.034	11.28	0.003	8.79	0.008	6.08	0.03
	η2p/WO	0.32	100	0.20	100	0.33	100	0.28	100	0.21	100
	F(1,23)/*p*	N1 (F4)		N1 (O2)		N1 (T5)		N1 (P4)		N1 (T5)	
	ES/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO
Semantic	F/*p*	8.31	0.01	4.99	0.048	12.5	0.002	10.3	0.004	4.76	0.0475
	η2p/WO	0.27	100	0.18	120	0.35	120	0.31	120	0.17	140
	F(1,23)/*p*	N1 (P3)		N1 (T6)							
	ES/WO	Stat.	*p*/WO	Stat.	*p*/WO						
Semantic	F/*p*	6.02	0.025	10.13	0.004						
	η2p/WO	0.21	140	0.31	140						
	F(1,23)/*p*	Cz		Cz		Cz		Cz		Cz	
	ES/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO	Stat.	*p*/WO
Interaction	F/*p*	7.35	0.012	5.78	0.025	7.31	0.013	9.47	0.005	7.62	0.011
	η2p/WO	0.24	180	0.20	200	0.24	220	0.29	240	0.25	280

Significant waveform effects, electrodes and time windows (20 ms interval as a time window) with significant amplitude differences of singular vs. plural/inanimate vs. animate/interaction ERPs within 0–600 ms, and their corresponding maximum statistics (F, *p*, and η2p); Stat., statistics; p, FDR-corrected *F*-value of ANOVA; WO, window set (at 20 ms intervals).

area at 180–280 ms using ERP data. It is supported by some research in recent years that semantics interact with quantitative processing. It was found that the peripheral region of the lateral and medial parietal cortex in semantic networks is selective for numbers (Huth et al., 2016). Meanwhile, an fMRI study found that quantifiers were identical to the processing of animal names (Wei et al., 2014). Another fMRI study suggested that calculation and language processes shared a common neural substrate since both of them activated the temporal lobe (Zago et al., 2008). However, the two tasks were separate procedures, and the idea of common neural substrates is not convincing. In our study, the interaction of semantic and quantity factors is unified in time and space owing to the one task and 2 × 2 variance analysis adopted.

Event-related potentials results showed that the main effect of the quantitative factor in Experiment 2 is consistent with the differences of singular and plural picture naming in Experiment 1, confirming that there was spontaneous quantitative processing at picture naming. The main effect of the semantic factor is also consistent with previous literature, manifested in the parieto-occipital temporal area at 100–140 ms. Semantic and quantitative factors affect each other in the process of picture naming, with a significant interaction at 180–280 ms. According to the time nodes of semantic, quantitative, and interactive processing, we considered that the summation coding and semantic processing interact with each other. And they affect the lexical retrieval, lemma selection, and integrating access to phonological forms of picture naming (Hassan et al., 2015). This is consistent with

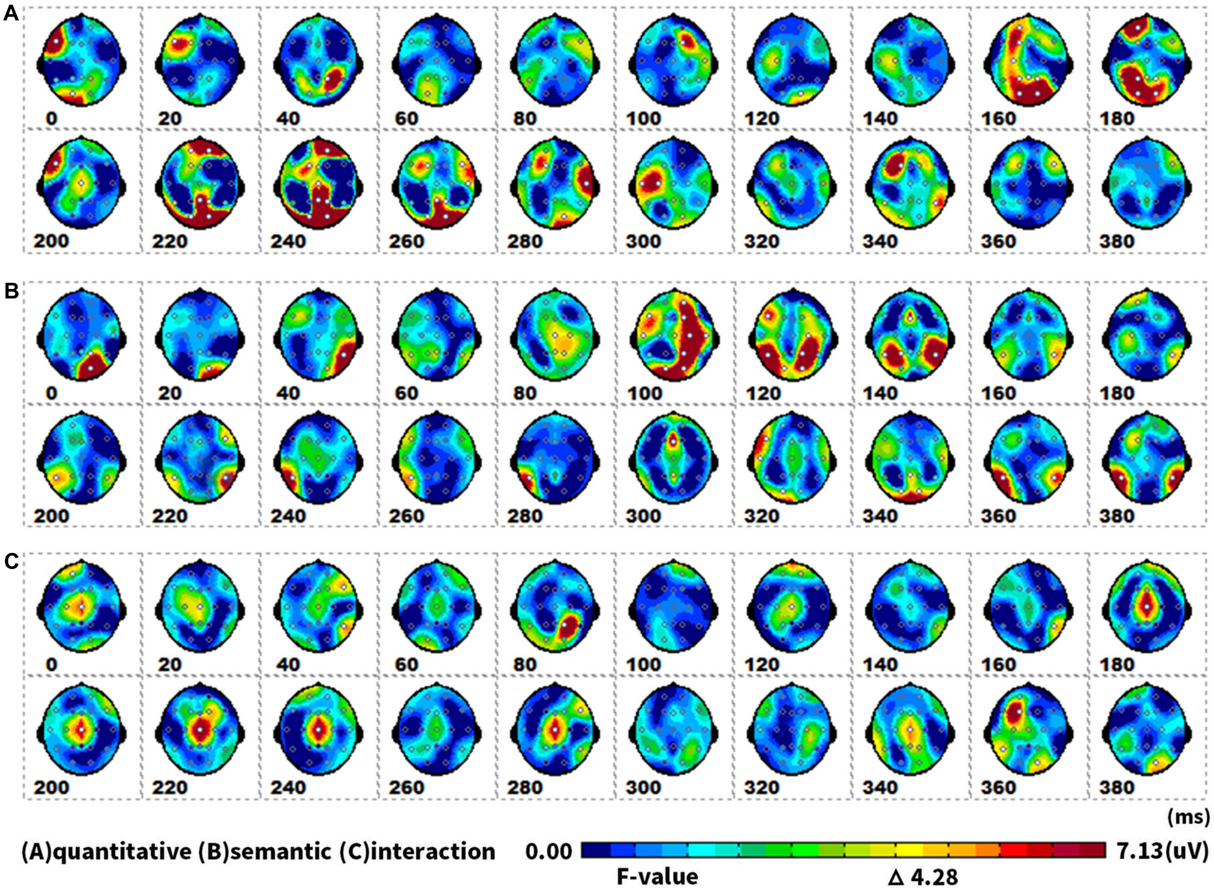


FIGURE 6

The spatiotemporal patterns of SPM (f) (0 to 400 ms) are derived from the two-way (semantic: animate vs. inanimate, and quantitative: singular vs. plural) repeated measures ANOVA: **(A)** the quantitative effect, **(B)** the semantic effect, and **(C)** the interaction effect. Each map was interpolated from the average *F*-values within the fixed 20-ms time window, and the bright yellow bin of the color scale corresponded to the 0.05 significance threshold: $F_{(1,23)} = 4.28$. The white dots represented the electrode sites with significant effects.

the differences in English singular and plural nouns (words and pronunciation). *Post hoc* tests showed that the difference in quantitative processing of animate pictures was larger than that of inanimate pictures. Plural pictures have greater differences in semantic and quantitative processing than singular pictures. This also indicates that semantics interact with quantitative processing. Although there was no statistical difference in behavior results among the four conditions, the RTs of the plural animate pictures were the largest. And this may be related to increased cognitive load.

General discussion

Because the vocabularies/word production processes of Chinese singular and plural picture naming were the same, it is often overlooked that the neural processing process of these two types of picture naming may be inconsistent. This paper investigated the neuropsychological processes of Chinese singular and plural picture naming. "Experiment 1" revealed that the neural electrical activities of these two types of picture naming were different. The ERP differences between the two conditions could not be explained by word production, but were close to the ERP differences in different quantities. It indicates that the Chinese singular and plural picture naming is not only a program of word production but also may involve quantitative processing. "Experiment 2" further explored that these two processes are not isolated, and there is an interaction between them.

Although we explored the neuropsychological processes associated with single and plural picture naming and highlighted the theories of picture naming for the first time, this study has several limitations. First, in this study, we only choose 1 and 3 which were both small numbers for the comparison of singular and plural numbers. In the future, different plural numbers should be tested to provide additional evidence for the neural mechanism of picture naming (such

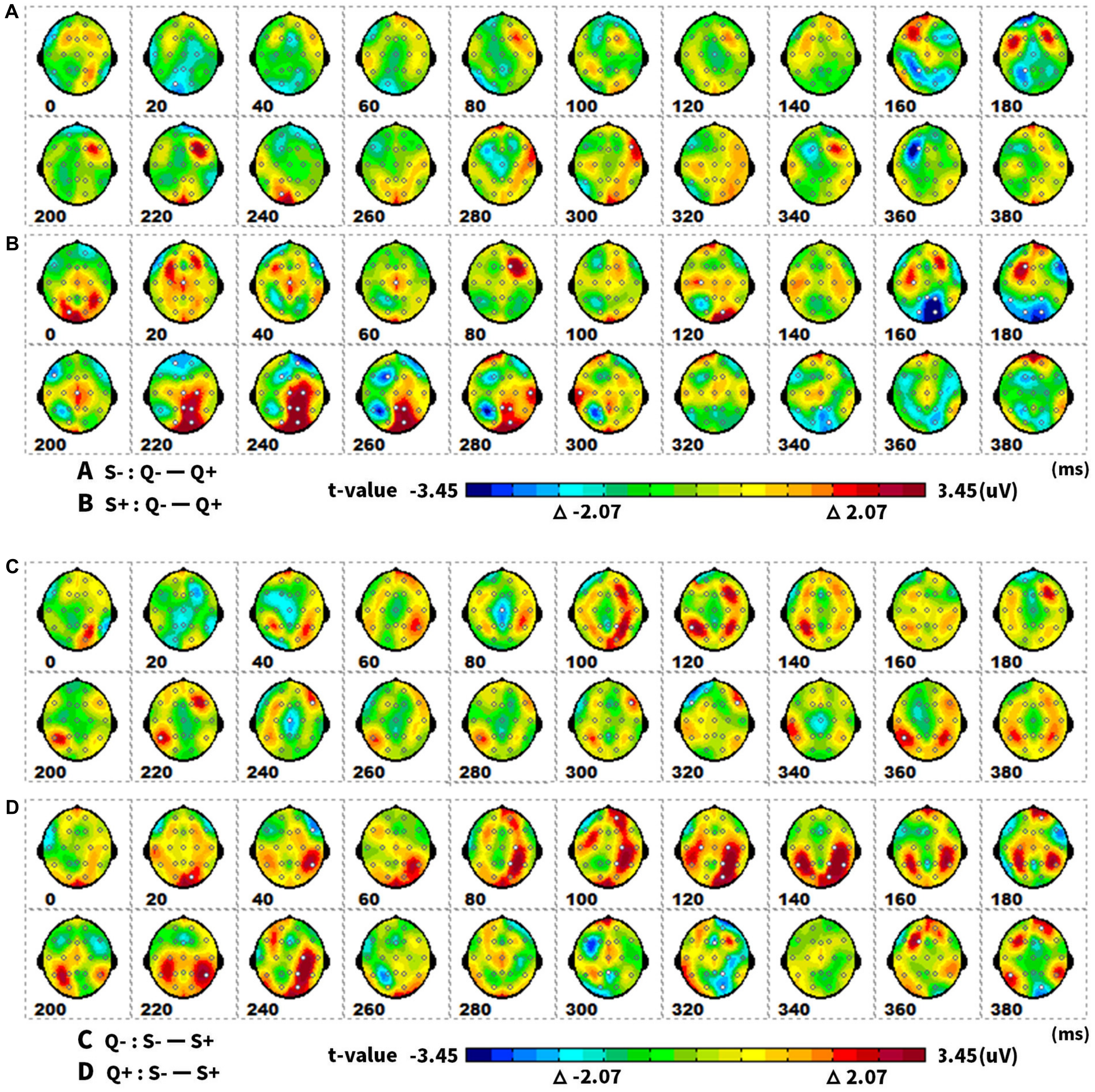


FIGURE 7
The spatiotemporal patterns of SPM (t) (0 to 400 ms) are derived from the pairwise comparisons between the conditions (Q−, Q + , S−, S +): **(A)** (S−. Q− − Q +), **(B)** (S + . Q− − Q +), **(C)** (Q−. S− − S +), **(D)** (Q + . S− − S +). Each map was interpolated from the average *t*-values within the fixed 20-ms time window, and the red/bright blue bin of the color scale corresponded to the 0.05 significance threshold: *t*(1,23) = ± 2.07. The white dots represented the electrode sites with significant effects. Q−, singular; Q + , plural; S−, inanimate; S + , animate.

as 5, 6, etc.). Second, in the present study, singular and plural pictures had the same visual complexity. If objects of single and plural images were set to the same size, more information about quantitative processing could be obtained. Third, the target language used for picture naming in the present study was Chinese, and given the universality of neural processes involved in picture naming with regard to language, the present findings should be validated in different languages.

Conclusion

This is the first study to investigate neuropsychological processes associated with singular and plural picture naming in the Chinese language. Results showed that singular and plural picture naming may involves two simultaneous neural processes: word production and quantitative processing. Moreover, these two processes share a common neural substrate – they interact at 180–280 ms in the central area.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of the First Affiliated Hospital of Jinan University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

L-YC and W-WC contributed to the conception and design of the study and performed the writing of this manuscript. L-YC, S-RS, C-MS, and RL acquired the recording data. SZ and Z-MC revised the main manuscript text. S-YB and WL modified the article.

Funding

This study received financial support from "National Key R&D Program of China" (2020YFC2005700), "High-Level Hospital Construction Research Project of Maoming People's Hospital," and "Guangdong Science and Technology Project" (2021A1414020006) during the conduct of the study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.898526/full#supplementary-material

References

Antell, S. E., and Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Dev.* 54, 695–701. doi: 10.2307/1130057

Arcara, G., Franzon, F., Gastaldon, S., Brotto, S., Semenza, C., Peressotti, F., et al. (2019). One can be some but some cannot be one: ERP correlates of numerosity incongruence are different for singular and plural. *Cortex* 116, 104–121. doi: 10.1016/j.cortex.2018.10.022

Benjamini, B. Y., and Yekutieli, D. (2001). The control of the false discovery rate in mutliple testing under dependency. *Ann. Stat.* 4, 1165–1188. doi: 10.1214/aos/1013699998

Beyersmann, E., Biedermann, B., Alario, F.-X., Schiller, N. O., Hameau, S., and Lorenz, A. (2018). Plural dominance and the production of determiner-noun phrases in French. *PLoS One* 13:e0200723. doi: 10.1371/journal.pone.0200723

Bock, K., and Levelt, W. J. M. (1994). "Language production: Grammatical encoding," in *Handbook of Psycholinguistics*, ed. M. A. Gernsbacher (London: Academic Press), 945–984.

Butterworth, B. (2005). The development of arithmetical abilities. *J. Child Psychol. Psychiatry.* 46, 3–18. doi: 10.1111/j.1469-7610.2004.00374.x

Cheng, W., Huang, Q., Chen, Y., Dai, W., Cui, L., Shan, S., et al. (2021). Parsing the Neural Mechanisms of Short-Term and Long-Term Associations in the Flanker Tasks: An ERP Analysis. *Front. Behav. Neurosci.* 15:626907. doi: 10.3389/fnbeh.2021.626907

Dehaene, S., and Changeux, J. P. (1993). DehaeneDevelopment of elementary numerical abilities: a neuronal model. *J. Cogn. Neurosci.* 5, 390–407. doi: 10.1162/jocn.1993.5.4.390

Dehaene, S., Naccache, L., Le Clec, H. G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., et al. (1998). Imaging unconscious semantic priming. *Nature* 395, 597–600. doi: 10.1038/26967

Fargier, R., Burki, A., Pinet, S., Alario, F.-X., and Laganaro, M. (2017). Word onset phonetic properties and motor artefacts in speech production EEG recordings. *Psychophysiology* 55, 1–10. doi: 10.1111/psyp.12982

Gallistel, C. R., and Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition* 44, 43–74. doi: 10.1016/0010-0277(92)90050-r

Gimenes, M., and Brysbaert, M. (2016). The processing of singular and plural nouns in English. French, and Dutch: New insights from megastudies. *Can. J. Exp. Psychol.* 70, 316–324. doi: 10.1037/cep0000074

Glaser, W. R. (1992). Picture naming. *Cognition* 42, 61–105. doi: 10.1016/0010-0277(92)90040-O

Hassan, M., Benquet, P., Biraben, A., Berrou, C., Dufor, O., and Wendling, F. (2015). Dynamic reorganization of functional brain networks during picture naming. *Cortex* 73, 276–288. doi: 10.1016/j.cortex.2015.08.019

Hodent, C., Bryant, P., and Houdé, O. (2005). Developmental science Language-specific effects on number computation in toddlers. *Dev. Sci.* 8, 420–423. doi: 10.1111/j.1467-7687.2005.00430.x

Holloway, I. D., and Ansari, D. (2010). Developmental specialization in the right intraparietal sulcus for the abstract representation of numerical magnitude. *J. Cogn. Neurosci.* 22, 2627–2637. doi: 10.1162/jocn.2009.21399

Huth, A. G., Heer, W. A., Griffiths, T. L., Theunissen, F. E., and Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532, 453–458. doi: 10.1038/nature17637

Indefrey, P., and Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/j.cognition.2002.06.001

Jevons, W. S. (1871). The power of numerical discrimination. *Nature* 3, 281–282. doi: 10.1038/003281a0

Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* 24, 175–219. doi: 10.1016/0010-0285(92)90007-o

Khwaileh, T., Body, R., and Herbert, R. (2015). Morpho syntactic processing of Arabic plurals after aphasia: dissecting lexical meaning from morpho-syntax within word boundaries. *Cogn. Neuropsychol.* 32, 340–367. doi: 10.1080/02643294.2015.1074893

Lage-Castellanos, A., Martínez-Montes, E., Hernández-Cabrera, J. A., and Galán, L. (2010). False discovery rate and permutation test: an evaluation in ERP data analysis. *Stat. Med.* 29, 63–74. doi: 10.1002/sim.3784

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/S0140525X99001776

Libertus, M. E., Woldorff, M. G., and Brannon, E. M. (2007). Electrophysiological evidence for notation independence in numerical processing. *Behav. Brain Funct.* 3:1. doi: 10.1186/1744-9081-3-1

Lins, O. G., Picton, T. W., Berg, P., and Scherg, M. (1993a). Ocular artifacts in EEG and event-related potentials. I: Scalp topography. *Brain Topogr.* 6, 51–63. doi: 10.1007/BF01234127

Lins, O. G., Picton, T. W., Berg, P., and Scherg, M. (1993b). Ocular artifacts in recording EEGs and event-related potentials. II: Source dipoles and source components. *Brain Topogr.* 6, 65–78. doi: 10.1007/BF01234128

Luck, S. J., and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281. doi: 10.1038/36846

McCrink, K., and Wynn, K. (2004). Large-number addition and subtraction by 9-month-old infants. *Psychol Sci.* 15, 776–781. doi: 10.1111/j.0956-7976.2004.00755.x

Meck, W. H., and Church, R. M. (1983). A mode control model of counting and timing processes. *J. Exp. Psychol. Anim. Behav. Process.* 9, 320–334. doi: 10.1037/0097-7403.9.3.320

Naccache, L., and Dehaene, S. (2001). The priming method: imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cereb. Cortex* 11, 966–974. doi: 10.1093/cercor/11.10.966

Ni, L., Liu, Y., Yu, W., and Fu, X. (2019). The China Image Set (CIS): A New Set of 551 Colored Photos With Chinese Norms for 12 Psycholinguistic Variables. *Front. Psychol.* 10:2631. doi: 10.3389/fpsyg.2019.02631

Nieder, A., Diester, I., and Tudusciuc, O. (2006). Temporal and spatial enumeration processes in the primate parietal cortex. *Science* 313, 1431–1435.

Nieder, A., Freedman, D. J., and Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science* 297, 1708–1711. doi: 10.1126/science.1072493

Nieder, A., and Merten, K. (2007). A labeled-line code for small and large numerosities in the monkey prefrontal cortex. *J. Neurosci.* 27, 5986–5993. doi: 10.1523/JNEUROSCI.1056-07.2007

Papathanassiou, D., Etard, O., Mellet, E., Zago, L., Mazoyer, B., and Tzourio-Mazoyer, N. (2000). A common language network for comprehension and production: a contribution to the definition of language epicenters with PET. *Neuroimage* 11, 347–357. doi: 10.1006/nimg.2000.0546

Parker, A. J., and Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* 21, 227–277.

Piazza, M., Giacomini, E., Le Bihan, D., and Dehaene, S. (2003). Single-trial classification of parallel pre-attentive and serial attentive processes using functional magnetic resonance imaging. *Proc. Biol. Sci.* 270, 1237–1245. doi: 10.1098/rspb.2003.2356

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron* 44, 547–555. doi: 10.1016/j.neuron.2004.10.014

Pinhas, M., Buchman, C., Lavro, D., Mesika, D., Tzelgov, J., and Berger, A. (2015). The Neural signatures of processing semantic end values in automatic number comparisons. *Front. Hum. Neurosci.* 9:645. doi: 10.3389/fnhum.2015.00645

Potter, M. C. (1979). "Mundane symbolism: The relations among objects, names, and ideas," in *Symbolic Functioning in Childhood*, eds N. R. Smith and M. B. Franklin (Hillsdale, NJ: Erlbaum), 41–65.

Roitman, J. D., Brannon, E. M., and Platt, M. L. (2007). Monotonic coding of numerosity in macaque lateral intraparietal area. *PLoS Biol.* 5:e208. doi: 10.1371/journal.pbio.0050208

Romo, R., Brody, C. D., Hernández, A., and Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399, 470–473. doi: 10.1038/20939

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., and Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* 326, 445–449. doi: 10.1126/science.1174481

Salinas, E. (2006). How behavioral constraints may determine optimal sensory representations. *PLoS Biol.* 4:e387. doi: 10.1371/journal.pbio.0040387

Salmelin, R., Hari, R., Lounasmaa, O. V., and Sams, M. (1994). Dynamics of brain activation during picture naming. *Nature* 368, 463–465. doi: 10.1038/368463a0

Santens, S., Roggeman, C., Fias, W., and Verguts, T. (2010). Number processing pathways in human parietal cortex. *Cereb. Cortex* 20, 77–88. doi: 10.1093/cercor/bhp080

Schiller, N. O., and Caramazza, A. (2002). The selection of grammatical features in word production: the case of plural nouns in German. *Brain Lang.* 81, 342–357. doi: 10.1006/brln.2001.2529

Schiller, N. O., and Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *J. Memory Lang.* 48, 169–194. doi: 10.1016/S0749-596X(02)00508-9

Sophiana, C., and Crosby, M. E. (2008). What eye fixation patterms tell us about subitizing. *Dev. Neuropsychol.* 33, 394–409. doi: 10.1080/87565640801982460

Starkey, P., and Cooper, R. G. Jr. (1980). Perception of numbers by human infants. *Science* 210, 1033–1035. doi:

Trick, L. M., and Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychol. Rev.* 101, 80–102. doi: 10.1037/0033-295x.101.1.80

van Turennout, M., Ellmore, T., and Martin, A. (2000). Long-lasting cortical plasticity in the object naming system. *Nat. Neurosci.* 3, 1329–1334. doi: 10.1038/81873

Verguts, T., and Fias, W. (2004). Representation of number in animals and humans: a neural model. *J. Cogn. Neurosci.* 16, 1493–1504. doi: 10.1162/0898929042568497

Wei, W., Chen, C. S., Yang, T., Zhang, H., and Zhou, X. L. (2014). Dissociated neural correlates of quantity processing of quantifiers, numbers, and numerosities. *Hum. Brain Mapp.* 35, 444–454. doi: 10.1002/hbm.22190

Wynn, K. (1992). Addition and subtraction by human infants. *Nature* 358, 749–750. doi: 10.1038/358749a0

Wynn, K., Bloom, P., and Chiang, W. C. (2002). Enumeration of collective entities by 5-month-old infants. *Cognition* 83, B55–B62. doi: 10.1016/s0010-0277(02)00008-2

Yu, X., Bi, Y., Han, Z., and Law, S. P. (2013). An fMRI study of grammatical morpheme processing associated with nouns and verbs in Chinese. *PLoS One* 8:e74952. doi: 10.1371/journal.pone.0074952

Zago, L., Petit, L., Turbelin, M. R., Andersson, F., Vigneau, M., and Tzourio-Mazoyer, N. (2008). How verbal and spatial manipulation networks contribute to calculation: an fMRI study. *Neuropsychologia* 46, 2403–2414. doi: 10.1016/j.neuropsychologia.2008.03.001

Zhou, S., Wei, J., Luo, Y., and Yang, W. (1998). New methods for three dimensional mapping of brain waves. *Brain Topogr.* 11, 103–110. doi: 10.1023/A:1022250404601

Zhou, S., Xiong, S., Cheng, W., and Wang, Y. (2019). Flanker paradigm contains conflict and distraction factors with distinct neural mechanisms: an ERP analysis in a 2-1 mapping task. *Cogn. Neurodyn.* 13, 341–356. doi: 10.1007/s11571-019-09529-w

Zhou, S., Zhou, W., and Chen, X. (2004). Spatiotemporal analysis of ERP during chinese idiom comprehension. *Brain Topogr.* 17, 27–37. doi: 10.1023/b:brat.0000047334.48256.9f

frontiers | Frontiers in Psychology

TYPE Methods
PUBLISHED 08 November 2022
DOI 10.3389/fpsyg.2022.724498

OPEN ACCESS

EDITED BY
Hong Xu,
Nanyang Technological University,
Singapore

REVIEWED BY
Antonio Prieto,
National University of Distance
Education (UNED), Spain
José Manuel Reales,
National University of Distance
Education (UNED), Spain

*CORRESPONDENCE
Oliver Y. Chén
olivery.chen@bristol.ac.uk

SPECIALTY SECTION
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

RECEIVED 13 June 2021
ACCEPTED 13 September 2022
PUBLISHED 08 November 2022

CITATION
Chén OY, Phan H, Cao H, Qian T,
Nagels G and de Vos M (2022)
Probing potential priming: Defining,
quantifying, and testing the causal
priming effect using the potential
outcomes framework.
Front. Psychol. 13:724498.
doi: 10.3389/fpsyg.2022.724498

Probing potential priming: Defining, quantifying, and testing the causal priming effect using the potential outcomes framework

Oliver Y. Chén[1]*, Huy Phan[2,3], Hengyi Cao[4,5,6],
Tianchen Qian[7], Guy Nagels[8,9] and Maarten de Vos[10,11,12]

[1]Faculty of Social Sciences and Law, University of Bristol, Bristol, United Kingdom, [2]School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom, [3]The Alan Turing Institute, London, United Kingdom, [4]Department of Psychology, Yale University, New Haven, CT, United States, [5]Center for Psychiatric Neuroscience, Feinstein Institutes for Medical Research, Manhasset, NY, United States, [6]Division of Psychiatry Research, Zucker Hillside Hospital, Glen Oaks, NY, United States, [7]Donald Bren School of Information and Computer Sciences, University of California, Irvine, Irvine, CA, United States, [8]Department of Neurology, Universitair Ziekenhuis Brussel, Jette, Belgium, [9]Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom, [10]Faculty of Engineering Science, KU Leuven, Leuven, Belgium, [11]Faculty of Medicine, KU Leuven, Leuven, Belgium, [12]KU Leuven Institute for Artificial Intelligence, Leuven, Belgium

Having previously seen an item helps uncover the item another time, given a perceptual or cognitive cue. Oftentimes, however, it may be difficult to quantify or test the existence and size of a perceptual or cognitive effect, in general, and a priming effect, in particular. This is because to examine the existence of and quantify the effect, one needs to compare two outcomes: the outcome had one previously seen the item vs. the outcome had one not seen the item. But only one of the two outcomes is observable. Here, we argue that the potential outcomes framework is useful to define, quantify, and test the causal priming effect. To demonstrate its efficacy, we apply the framework to study the priming effect using data from a between-subjects study involving English word identification. In addition, we show that what has been used intuitively by experimentalists to assess the priming effect in the past has a sound mathematical foundation. Finally, we examine the links between the proposed method in studying priming and the multinomial processing tree (MPT) model, and how to extend the method to study experimental paradigms involving exclusion and inclusion instructional conditions.

KEYWORDS

priming effect, causal inference, potential outcomes framework, word fragment completion test, significant test, between-subjects study

Introduction

Imagine you are asked to fill in a fraction of a word, say _aze_ _e. Suppose the target word is gazette. What would your performance be if you have seen a list of words including gazette before the game? Intuitively, seeing a list of words containing the target answer improves one's performance. But, how could we formally test whether the improvement exists, and how could we quantify the amount of improvement?

More specifically, we call such a phenomenon where having been exposed to an item (e.g., viewing a word or an object) facilitates the subsequent recovery of the item based on a partial or reduced perceptual cue (e.g., viewing a partial word or a fragment of an object) *repetition priming* (Hayman and Tulving, 1989; Tulving and Schacter, 1990, 1992).

Neurobiologically, this (priming) effect on word identification is facilitated and carried out through activations in the brain involving memory and learning. Although the exact neural bases of priming are as of yet little known, several lines of evidence have hinted that priming is mediated by neural systems outside of the medial temporal and diencephalic regions (Tulving and Schacter, 1990), and that priming is related to changes in cortical modules that are involved in processing specific attributes of stimulus information (Squire, 1987). Neuropsychologically, posterior cortical areas in the right hemisphere seem to be associated with object identification (Warrington and Taylor, 1978); passive reading of familiar words produces selective bilateral activation in the extrastriate cortex, suggesting that visual identification (not necessarily understanding) of words has an anterior-occipital locus (Schwartz et al., 1980; Funnell, 1983; Satori and Job, 1987; Petersen et al., 1988).

Yet, despite neurobiological and neuropsychological advances, little do we know about how to formally test the existence of a priming effect, whether the effect is causal and if so, how to quantify it. The difficulty, in part, lies in the need to compare two scenarios where only one is observable. Specifically, to claim that there exists a priming effect (e.g., the effect of a word study on word identification), one must first quantify the outcomes of two scenarios (e.g., word identification accuracy after viewing the target words vs. the accuracy without viewing the target words) and then compare these two outcomes to draw a (statistical) conclusion. But, only one of these[1] is observable on each individual. How, then, could we compare an observable outcome with an unobservable one?

Here, linking Neyman and Rubin's works on causal inference and Tulving and Schacter's earlier works on priming, we aim to define, test, and quantify the causal priming effect using the *potential outcomes* framework (Neyman, 1923; Rubin, 1974, 1977, 1978). We demonstrate how to use this framework to study the priming effect by analyzing data from a between-subjects study (Hayman and Tulving, 1989). We also show that what has been previously used intuitively to study the priming effect has a sound mathematical foundation. But before we proceed, it is perhaps useful to discuss the reasons for choosing this framework, the relationship between priming and memory, and the convenience of studying priming using a word fragment completion test.

A brief introduction of causal inference

Let us begin by briefly introducing and comparing three useful approaches to study causation: Campbell's ("validity testing") framework, Pearl's (causal diagram) framework, and the Neyman–Rubin's (potential outcomes) framework.

Campbell's framework focuses on evaluating the validity of standard designs for experimentation in the social sciences and finding extraneous variables that may confound causal interpretations (Campbell, 1957).

The Neyman–Rubin framework focuses on the *magnitude* of the causal effect; it emphasizes the mathematical argument that can yield an analytical estimate of the causal effect. As only one of the two outcomes in the Neyman–Rubin framework can be observed from each individual, they are usually referred to as *potential outcomes* (Neyman, 1923; Rubin, 1974, 1977, 1978).

Pearl's framework introduces *directed* graphs into causal analysis, with nodes indicating variables (e.g., exposure and outcome) and edges indicating causal links (Pearl, 1993, 1995, 2001, 2009a). In addition, the $do(\cdot)$ operator[2] and the back-door and front-door criteria make some otherwise difficult causal effects identifiable (see later).

A comparison between Campbell's, Neyman–Rubin's, and Pearl's causal models

Similarities

Most psychologists are familiar with Campbell's method; perhaps few have had exposure to the Neyman–Rubin model (West and Thoemmes, 2010). In our view, however, Design 6 [3] in Campbell (1957) shows spirit of both Neyman–Rubin's

---

[1] Either viewing the target words or not viewing them; one cannot un-view the words that one had viewed.

[2] Here, $do(X = x)$ means the model forces X to take the value of x; in other words, one sets X (*via* intervention) to be a constant value x.

[3] Consider two experiments: $A\ X\ O_1$ *vs.* $A\ [\ ]\ O_2$, where A, X (or lack thereof), and O are ordered from left to right in time, and A, X, and O indicate a random sampling assignment, a treatment, and the outcome, respectively. In Campbell's approach, the presence of X on the left of O_1 means O_1 is the outcome of a group after receiving a treatment X, and the absence of X (*i.e.*, blank space) on the left of O_2 means O_2 is the outcome of another group without receiving a treatment.

potential outcomes framework[4] and Pearl's causal diagram[5]. As for Pearl's and Neyman–Rubin models, oftentimes, they are mathematically equivalent[6] (see Section 7.4.4 of Pearl, 2009b).

Differences

Compared with Campbell's approach, the Neyman–Rubin framework offers an analytical language for identifying and quantifying the causal effect. Compared with Neyman–Rubin formulations, Pearl's method is oftentimes easier for social scientists to understand and visualize the causal problems using vivid graphic representations. In certain cases[7], controlling for covariates using the Neyman–Rubin method may fail to identify a causal effect – a major criticism from the Pearl school. Furthermore, under the potential outcomes framework, there is a subtle difference between Neyman's null (where the null hypothesis considers zero average causal effect) and Fisher's null (where the null hypothesis considers zero individual causal effect) for many realistic situations, which may cause confusions (Ding, 2017). The Rubin school argues[8] that causation, especially causation involving directed causation and dynamic causation, cannot be simply explained by graphs. Pearl's method assumes that the $do(\cdot)$ operator itself does not perturb the (causal) system, about which some may caste doubts; in addition, oftentimes this assumption cannot be tested. For experimentalists, it is sometimes impractical to apply the $do(\cdot)$ operator to intervene certain variables such as gender and age. Finally, in practice, it may be difficult to obtain a complete picture of the causal diagram (e.g., the directed causal map of the brain network).

Weighing pros and cons and in light of priming research, in this article, we derive the potential causal priming framework in Neyman–Rubin's language and accompany graphs in Pearl's style to visualize causal relationships (see the Discussion section for future directions).

Remark 1. We encourage interested readers to compare, in detail, the potential outcomes framework with Campbell's framework (e.g., West and Thoemmes, 2010) and the potential outcomes framework with Pearl's framework [e.g., Gelman's blog post (Gelman, 2009) and Pearl's response under the post].

Remark 2. There are other fine works on causal inference; we refer our readers to them for further reading (Peters, 1941; Cochran and Chambers, 1965; Hill, 1965; Goldberger, 1972; Ding et al., 2016).

A brief discussion of memory

Different memory systems

Whereas the focus of the article is on priming, it is perhaps beneficial to familiarize oneself with the memory systems, in general. This is because on the one hand, priming is related to memory, and on the other hand, it is arguably independent of explicit and semantic memory (Tulving and Schacter, 1990). By stating explicit and semantic memory, one has already implied there exists some categorization of memory systems. Although we do not intend to and cannot fully examine the hypothesis regarding the number of memory systems present, a summary of a few key classifications of memory systems may help the readers to deal with priming conceptually. Tulving (1985) argued that there exist three types of memory systems: episodic (associated with self-knowing consciousness), semantic (associated with knowing consciousness), and procedural (associated with non-knowing consciousness). Cohen and Squire (1980) and Mishkin et al. (1984) argued that there are two types of memory systems: the former coined the two systems according to the concepts of "knowing how" and "knowing that," and the latter distinguished the habit system from the "memory" system. Others have proposed more specific classifications, arranged either hierarchically (Pribram, 1984)[9] or interactively without a fixed relationship to each other (Johnson, 1983). More specifically to priming, it is hypothesized that there exists a pre-semantic perceptual system [called the *perceptual representation system* (PRS)] that manages priming; the PRS operates independently of the explicit and semantic memory (Tulving and Schacter, 1990). In brief, the hypothesis of the PRS suggests that there is a dissociation between priming and explicit memory and that there is a dissociation between (pre-semantic) priming and semantic memory (Warrington and Taylor, 1978; Parker et al., 1983; Graf et al., 1984; Hashtroudi et al., 1984; Cermak et al., 1985; Light et al., 1986; Shimamura, 1986; Nissen et al., 1987; Kopelman and Corn, 1988; Parkin and Streete, 1988; Tulving and Schacter, 1990).

Process dissociation model

Interposed between the classification of multiple memory systems and the study of priming is the need to separate the latter from other, for example, semantic and explicit

---

[4] The letter *A* in Campbell's approach is equivalent to the randomization and matching mechanism in Neyman-Rubin's approach, or in Campbell's words: "*A* is the point of selection, the point of allocation of individuals to groups . . . At time *A* the groups were equal, even if not measured. . ."

[5] Pearl's circles and arrows are equivalent, in spirit, to Campbell's letters and orders: in Campbell's notation, if *X* is placed on the left of *O*, it implies there is a directed arrow from *X* to *O*.

[6] Namely, $P\{Y|do(Z) = z\} = P\{Y(z)\}$, where $P\{Y|do(Z) = z\}$ (in Pearl's language) means forcing Z to take the value z by removing all father nodes of Z, and $P\{Y(z)\}$ (in Neyman-Rubin's language) means the potential outcome of Y under z.

[7] Suppose (1) U and W both cause X, (2) U and W cause T and Y, respectively, and (3) T causes Y. Using Pearl's method, the relationship from T to Y is causal. But using Neyman-Rubin's method, by controlling for X, the relationship from T to Y is not causal.

[8] We attribute a part of the summary between Neyman-Rubin's and Pearl's methods to works from Peng Ding.

[9] The discussion was on primates.

memory processes. This need is partly sprawled empirically from findings where patients with amnesia reported significantly worse explicit memory (intentional use of memory) than normal subjects but showed as large a priming effect (an arguably automatic, passive use of memory) as the normal subjects (Warrington and Weiskrantz, 1974; Graf et al., 1984; Cermak et al., 1985; see Shimamura, 1986 for a review). Practically, to separate and estimate the contribution of unconscious, automatic, controlled, and intentional processes, Jacoby (1991) proposed the *process dissociation framework* and argued its utility in studying perception, memory, and thought. The key point of the framework is to use regression models to separate the effect of (consciously controlled) recollection from that of (automatic) familiarity [see Experiment 3 in Jacoby (1991) for details].

The role of memory in encoding instructions

Participants in a priming study need to follow instructions. Working memory, the ability to maintain and process information (Baddeley and Hitch, 1974), plays an important role in encoding both spoken (Baddeley et al., 1984; Hanley and Broadbent, 1987) and written (Wright, 1978; Wright and Wilcox, 1978) instructions. Cognitive load (including intrinsic, extraneous, and germane loads) that connects instructional design to cognitive functions is related to working memory. The cognitive load consumes a part of the working memory, and particularly, with appropriate instructional design, the germane load positively affects learning (Cooper et al., 2001).

A brief introduction to the word fragment completion (WFC) test

The word fragment completion (WFC) test is widely used to assess priming. In general, the test consists of a study phase and a test phase. During the study phase, subjects are instructed to view a list of words, including target words (e.g., gazette) and non-target words (called buffers). The test phase starts after an interval (e.g., 2 h). During this phase, the subjects are randomly assigned into two groups; each group undertakes one of the following tasks: (1) uncovering studied words (e.g., gazette) given a cue and (2) uncovering non-studied words given a cue. Some fragment completion tests will include an additional test, which involves repeating the word identification of gazette either with the same cue or with a different cue (see examples in the Results section). In simple terms, priming is said to have occurred when the success rate of cue-based item identification after studying the item is higher than that of a non-studied item (see **Figure 1**).

Under the Neyman–Rubin's potential outcomes framework, the priming effect of receiving a word study, which consists of the target words (the exposure of interest), on word identification (the outcome) can be defined as follows:

> *It is the difference between the two potential outcomes: the first is the outcome had an individual received the word study which consists of the target words, and the second is the outcome had the same individual not received the word study (or received a word study which did not contain the target words).*

We restrict our focus on the priming effect during a non-semantic word completion test, although the framework can be extended to studying semantic tasks such as rating the pleasantness when viewing a word and giving its definition. This is, in part, because priming is not affected by semantic and non-semantic encoding (Tulving and Schacter, 1990). Similarly, as priming occurs in more complex studies such as visual object recognition (Schacter et al., 1990), the framework introduced in this article may also be useful to quantify these priming effects. Although we focus on modeling the causal effect in studies of implicit memory, it may shed some light on studies of explicit memory (see the Discussion section). Finally, we note that when the instructions were not implicit but explicit during a fragment completion test, the test should be, in spirit, considered more as a "cued recall test" than a "fragment completion test."

Method

Notations and definitions

We begin by defining the notations used throughout this article. We use Z to denote whether a word study concerning viewing a list of words (including target words, such as gazette, which we use as an example throughout this article, and non-target words, such as *vermouth*) is undertaken at time $t_0 = 0$ (see **Figure 1**). Specifically, $Z = 1$ means that a subject has undertaken a word study including the target words (and henceforth referred to as having undertaken a word study for simplicity), and $Z = 0$ means that the subject has not undertaken the word study (or have undertaken a word study with all non-target words, which, for simplicity, we will henceforth refer to as not having undertaken a word study). In this study, we consider that Z takes binary values (i.e., having vs. not having conducted a word study), although our approach can be extended to categorical Z that takes more than two values (e.g., word studies consisting of words with low, intermediate, and advanced level of complexity). The word complexity can be quantified by, for example, evaluating the combination of syllable shapes and word patterns. As such, a further extension of Z can take any value between 0 and 100 to indicate complexity of each word (see the Discussion section for continuous and time-dependent cases).

Let X_1 denote a cue (e.g., X_1 = _aze_ _e) given during a WFC test at a time t_1 ($t_1 = t_0$, typically t_1 is 2 h after t_0) (see **Figure 1**). We write $\mathrm{Y}(X = \mathrm{x}, Z = \mathrm{z})$ as the outcome of

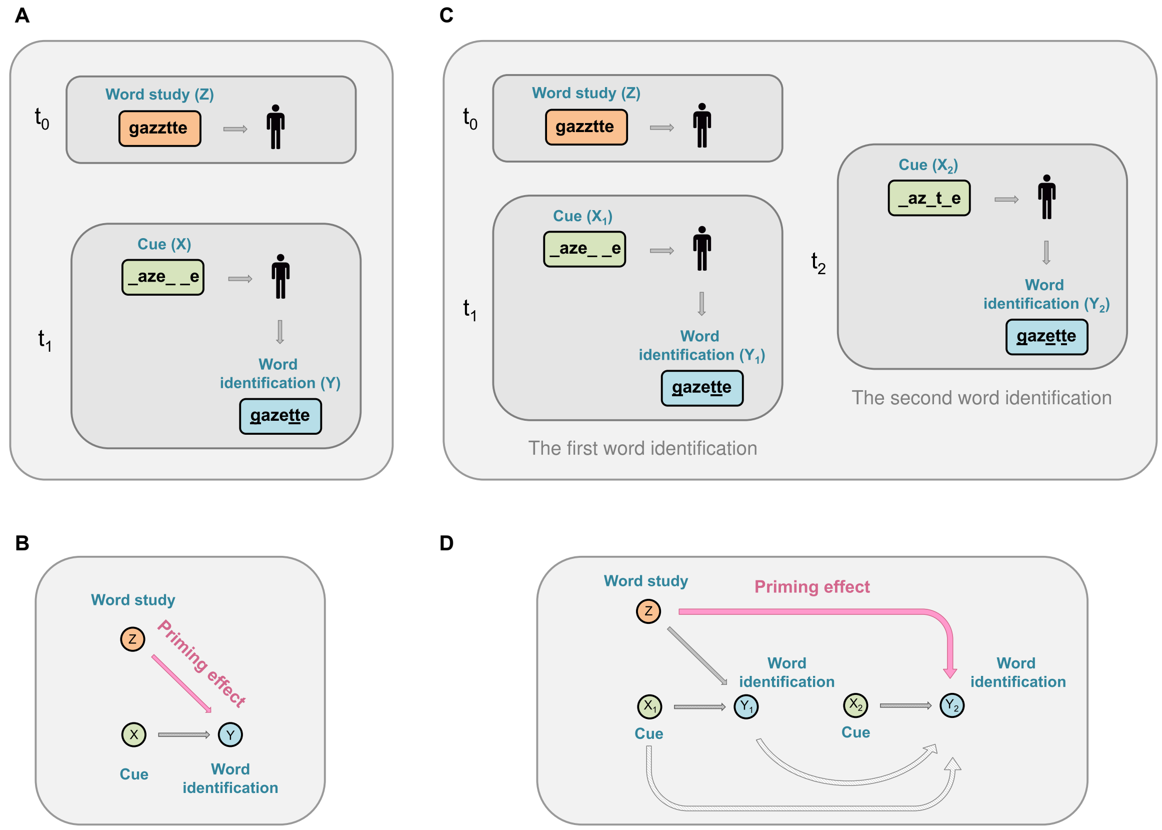


FIGURE 1
A schematic representation of the priming effect during a word completion experiment. **(A)** First, a subject views a list of words (including a target word gazette, or Z) during a word study. Next, a cue that consists of a fragment of the word (e.g., _aze_ _e, or X) is given to the subject. The subject is then asked to fill in the blanks. The subject may successfully uncover the target word or another word gazelle; or report a non-word gazeele, or an incomplete entry such as _azelle. **(B)** Priming effect refers to the phenomenon that being previously exposed to the word gazette (Z) primes (i.e., facilitates) the identification of the word (Y) given a partial or reduced cue (X). **(C)** A word study followed by two-word identifications. Left: The same word study and the subsequent cue-based identification, as in **(A)**. Right: An additional cue-based identification. In the figure, a different cue (e.g., _az_t_e, or X_2) is used during the second word identification for illustration purposes; two identical cues can also be used. **(D)** Priming effect where being previously exposed to the word gazette (Z) primes (i.e., facilitates) the identification of a word (Y_2) during a second word identification test given a partial or reduced cue _az_t_e (X_2).

word identification on the experiment unit (i.e., an individual participant), given that the unit received a word study $Z = z$ at t_0 and a cue $X = x$, where the upper case indicates a random variable and the lower case refers to its realized value. If there is an additional test, let X_2 denote the cue (e.g., $X_2 =$ _az_t_e) given during the second word completion test at time t_2, where t_2 can be, for example, 2 h after t_1. By design, we have $t_2 > t_1 > t_0$. Between t_1 and t_2, participants can undertake tasks irrelevant to the experiment, such as taking a cognitive psychology class. We define Y_1 and Y_2 as the corresponding word identification outcomes given cues X_1 and X_2, respectively.

In the following, we always assume that each cue X corresponds to a single answer, and we drop the notational dependence of Y on the target word where there is no confusion. For example, Y ($X =$ _aze_ _e, $Z = 1$) $= 1$ means that the word identification is correct (e.g., the identified word is gazette, the target word, or gazelle, another correct answer[10]) given the cue $X =$ _aze_ _e, after a word study Z consisting of a target word gazette. Y ($X =$ _aze_ _e, $Z = 1$) $= 0$ means that the word identification is incorrect given the cue $X =$ _aze_ _e, after a word study Z consisting of the target word gazette. Similarly, Y ($X =$ _aze_ _e, $Z = 0$) $= 1$ means that the word identification is correct (i.e., the identified word is gazette or gazelle) given the cue $X =$ _aze_ _e, had a word study Z not been conducted. Y ($X =$ _aze_ _e, $Z = 0$) $= 0$ means that the word identification is incorrect given the cue $X =$ _aze_ _e, had a word study Z not been conducted.

Definition 1.1 (Causal priming effect)

The causal priming effect on an experiment unit (i.e., a subject) given a cue X = x is defined as

---

[10] This indicates the study phase did not affect the WFC.

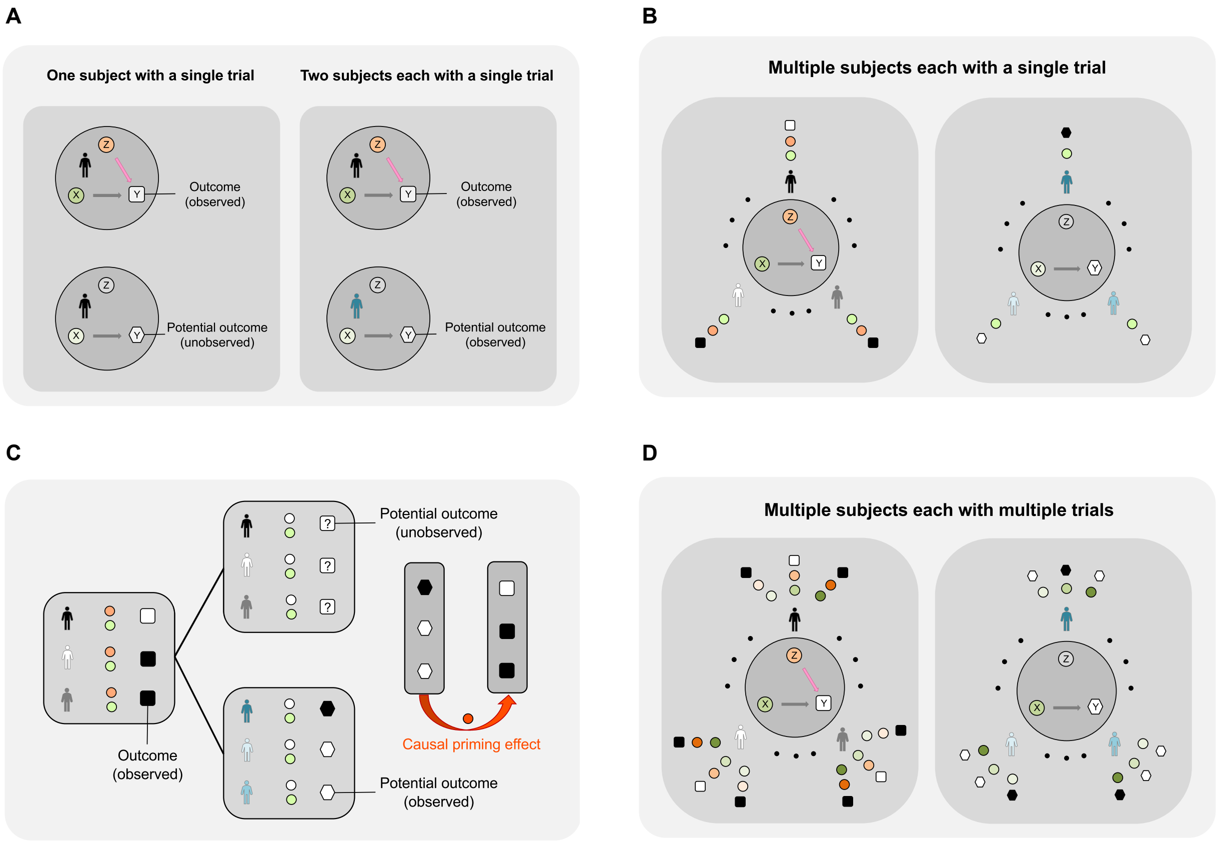


FIGURE 2

Using the potential outcomes framework to study the causal priming effect during a word completion experiment. **(A)** Identifying the causal priming effect with one trial. Top left: The figure describes a scenario where the subject first conducts a word study (*Z* = 1) and then aims to identify a target word based on a partial cue (X). The outcome (or Y) is observable. Bottom left: The figure describes a scenario where a subject does not conduct a word study (*Z* = 0) and aims to identify a target word based on a partial cue (X). If the subject has already participated in a word study, this (potential) outcome (or Y) under no word study is not observable. Top right: The same experiment as in the top left. *Bottom right*: Since the potential outcome of the subject (indicated by a human icon in black) is not observable, a different subject (indicated by a human icon in blue) is asked to identify the same target word based on the same partial cue (X), without a word study (*Z* = 0). If the two subjects are similar, then the causal priming effect is estimated by the difference between the outcomes (denoted by a letter Y in a square and a letter Y in a hexagon). **(B)** To reduce the possibility that a particular word may yield various priming effects on different subjects, multiple individuals are randomized to either conduct a word study (*Z* = 1) or not (*Z* = 0). The individuals with grayscale colors receive a word study; the individuals with bright colors do not receive a word study. For the left image, green circles represent various cues, orange circles represent word studies (*Z* = 1), and black or white squares represent the word study outcomes, where a black square indicates success (*Y* = 1) and a white square indicates failure (*Y* = 0). For the right image, green circles represent the cues corresponding to those on the left image, gray circles with the letter Z refer to having not conducted a word study (*Z* = 0), and black or white hexagons represent the word study outcomes, where a black hexagon indicates success (*Y* = 1) and a white hexagon indicates failure (*Y* = 0). **(C)** Average priming effect of a word study on word identification among multiple subjects using one trial. Left: Half of the subjects are randomized to perform the word identification experiments after a word study (indicated by orange circles). The black and white squares indicate the observed outcomes. Top middle: The potential outcomes of the same subjects (who have participated in a word study) had they not participated in a word study (indicated by blank circles). The squares with question marks indicate that these outcomes are not observable. Bottom middle: The remaining subjects perform the word identification experiments without a word study; the outcomes are observable and are indicated by black and white hexagons for successful and failed cases, respectively. Right: Due to randomization, the difference in the outcomes between the two groups (divided by the sample size) gives an estimate of the average priming effect of the word study on word identification in a sample. **(D)** The experiment can be further extended to multiple subjects with multiple words.

$Y(X = x, Z = 1) - Y(X = x, Z = 0)$; this quantifies the difference between the outcome Y from a study unit that has conducted a word study ($Z = 1$) versus the outcome Y from the same unit had no word study been conducted ($Z = 0$), given the same cue $X = x$.

Only one of the two potential outcomes can be observed from each subject. In other words, the individual-level causal priming effect is non-identifiable. Therefore, a natural inquiry into the causal priming effect is to uncover the average priming effect across multiple subjects (see Figures 2A,B).

Definition 1.2 (Average priming effect of a single studied word)

Consider one target word in a wordlist that is viewed by a total of N subjects. Then, the *average priming effect* (APE) of a word study Z on a word identification Y given a cue X is defined as follows:

$$APE_N^{\{1\}} = \frac{1}{N}\left\{\sum_{i=1}^{N} Y_i\,(X = x,\ Z = 1) - Y_i\,(X = x,\ Z = 0)\right\}$$

where $\mathrm{Y}_i\,(X = \mathrm{x}, Z = \mathrm{z})$ indicates the word identification result of the i^{th} subject after given a cue $X = x$ and a word study Z ($Z = 1$ means after viewing a wordlist consisting of target words and $Z = 0$ means without a word study). The superscript $\{1\}$ indicates that it is the average priming effect for one target word, and the subscript N indicates that the effect is defined on N individuals.

Unfortunately, we cannot observe both $\mathrm{Y}\,(X = \mathrm{x}, Z = 1)$ and $\mathrm{Y}(X = \mathrm{x}, Z = 0)$ on the same subject. This is because after having assigned (or not assigned) a word study Z (e.g., $Z = 1$) and the word identification test result Y has been reported, we cannot go back to the time t_0 to assign a different Z (i.e., $Z = 0$). Certainly, one could experiment on the same unit in two trials (one with the word study $Z = 1$ and the other with $Z = 0$), which consists of a repeated-measures study (e.g., Challis and Brodbeck, 1992). The first study, however, may have a carryover or learning effect on the second. Therefore, we cannot ascertain that the priming effect is due to the word study Z or the information learned (e.g., the cue X, the word identification Y, or the study mechanism) from the first test (see the Discussion section for details).

Estimating average priming effect involving one target word in a 2*K* trial study

Consider a sample of $2K$ subjects, where half of the subjects undertake a word study and half do not (see **Figure 2B**). Let S_Z denote the indices of the subjects who undertake the word study, and let S_{NZ} denote the indices of the subjects who do not. Let DPE denote the difference between the average observed word identification accuracy of the S_Z group and the average observed word identification accuracy of the S_{NZ} group, as follows:

$$DPE_{2K}^{\{1\}} = \frac{1}{K}\left\{\sum_{i\in S_Z} \mathrm{Y}_i\,(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = 1)\right\} - \frac{1}{K}\left\{\sum_{i\in S_{NZ}} \mathrm{Y}_i\,(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = 0)\right\}.$$

Following the definition of the APE in *Definition 1.2*, the average (causal) priming effect across a sample of $2K$ individuals involving one target word is defined as follows (see **Figure 2C**):

$$APE_{2K}^{\{1\}} = \frac{1}{2K}\left\{\sum_{i=1}^{2K} \mathrm{Y}_i\,(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = 1) - \mathrm{Y}_i(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = 0)\right\}.$$

Since $APE_{2K}^{\{1\}}$ is not observable and $DPE_{2K}^{\{1\}}$ is, one would ask if $DPE_{2K}^{\{1\}}$ is close to $APE_{2K}^{\{1\}}$. The answer depends on two factors: (a) how well matched are subjects who conduct the word study and those who do not; (b) if the word study is randomly assigned. We examine these factors in detail as follows:

First, if the S_Z group and the S_{NZ} group are perfectly matched[11] [i.e., for every subject in the S_Z group who receives a word study, there is a subject in the S_{NZ} group who does not receive a word study; and these two (matched) subjects would perform identically if a word study were conducted or if a word study were not conducted[12]], then $DPE_{2K}^{\{1\}} = APE_{2K}^{\{1\}}$. This holds whether the word study Z is randomly assigned or not (Rubin, 1974). Second, if the two groups are not perfectly matched, but before the tests, investigators have controlled all the variables that would affect the performance (e.g., only consider subjects with the same age, gender, and education background), then $DPE_{2K}^{\{1\}}$ is close to $APE_{2K}^{\{1\}}$ (i.e., the subjects are as if matched). Third, if the word study Z is randomly assigned, even if there are unmatched subjects (e.g., subjects have significant different language proficiency). For example, English speakers may perform better than non-English speakers in a word completion test in English; the random assignment is going to balance, in expectation, all observed and unobserved factors that would impact the word identification. To put it more concretely, by randomly assigning a word study to individuals, some English speakers would receive a word study (the rest of the English speakers would not receive one), and some non-English speakers would receive a word study (the rest non-English speakers would not receive one). As a result, the individuals who receive a word test consist of both English and non-English speakers, and the individuals who do not receive a word test also consist of both English and non-English speakers; thus, the bias due to language efficiency is reduced. Randomization becomes increasingly effective when the sample size N increases (Scheffe, 1959; Rubin, 1974; Wu and Hamada, 2000; Hinkelmann and Kempthorne, 2005).

Although matching or randomization makes $DPE_{2K}^{\{1\}}$ a suitable estimator for $APE_{2K}^{\{1\}}$, it remains important to generalize

---

[11] The definition of "match" here is more restricted than it is in the context of propensity score matching, as we consider the matched pair to have identical potential outcomes. We use this term for illustration of causal effect rather than estimation.

[12] We need both potential outcomes to be equal; not just the potential outcome under treatment (*i.e.*, $Z = 1$, namely had a word study been conducted).

it to any $2K$ sample. To that end, we defined the expected priming effect (EPE) (i.e., the expectation of $DPE_{2K}^{\{1\}}$) as follows:

$$EPE_{2K}^{\{1\}} = \mathbb{E}\left\{\frac{1}{K}\sum_{i\in S_Z} I_{\mathrm{Y}_i(\mathrm{X}\,=\,\mathrm{x},\ \mathrm{Z}\,=\,1)\,=\,1} - \frac{1}{K}\sum_{i\in S_{NZ}} I_{\mathrm{Y}_i(\mathrm{X}\,=\,\mathrm{x},\ \mathrm{Z}\,=\,0)\,=\,1}\right\}$$

where $\mathbb{E}$ indicates the expectation operation.

Since $\mathrm{Y}_i = 0$ or 1, then $DPE_{2K}^{\{1\}} = \frac{1}{K}\sum_{i\in S_Z} I_{\mathrm{Y}_i(\mathrm{X}\,=\,\mathrm{x},\ \mathrm{Z}\,=\,1)\,=\,1} - \frac{1}{K}\sum_{i\in S_{NZ}} I_{\mathrm{Y}_i(\mathrm{X}\,=\,\mathrm{x},\ \mathrm{Z}\,=\,0)\,=\,1}$, where $I_{\mathrm{Y}_i(\mathrm{X}\,=\,\mathrm{x},\ \mathrm{Z}\,=\,z)\,=\,1}$ is an indicator function[13] that takes value 1 if $\mathrm{Y}_i(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = \mathrm{z}) = 1$, and takes value 0 if $\mathrm{Y}_i(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = \mathrm{z}) = 0$. Then $EPE_{2K}^{\{1\}}$ reduces to

$$EPE_{2K}^{\{1\}} = \mathbb{P}\{\mathrm{Y}(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = 1) = 1\} - \mathbb{P}\{\mathrm{Y}(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = 0) = 1\} \tag{1}$$

where $\mathbb{P}\left\{\mathrm{Y}(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = \mathrm{z}) = \mathrm{y}\right\}$ denotes the probability of the word identification Y equals to y ($\mathrm{y} = 0$ or 1) given the cue $X = \mathrm{x}$ and the word study Z equals to z ($z = 0$ or 1).

In simple terms (see Remark 3), $EPE_{2K}^{\{1\}}$ means that the expected priming effect estimated from $2K$ subjects regarding one word is the difference between the probability of correctly identifying the target word for all subjects who have taken the word study ($Z = 1$) and the probability of correctly identifying the target word for those who have not participated in the word study ($Z = 0$).

Estimating priming effect involving multiple target words

The variability of individual memory affects the individual priming effect (a treatment of which is to estimate the average priming effect across subjects, as outlined in *Definition 1.2*) and so does the variability of words. Hence, the *APE* estimated using a complicated, uncommon, and non-word is likely to differ from the *APE* estimated using a simple and common word; this is true even when words of similar complexity are considered (because even when we only focus on, say, words of intermediate complexity, there are, potentially, differences in syllable shapes and word patterns). A natural treatment is to conduct tests on multiple words and estimate the average priming effect over these words across subjects.

Definition 1.3 (Average priming effect across multiple studied words)

Consider a wordlist consisting of M target words in a study consisting of a total of N subjects (see **Figure 2D**). Define X_{ij} as a cue given to the i^{th} subject associated with the j^{th} target word. Define Y_{ij} as the outcome of the corresponding word identification. Then, the *average priming effect* of the word study Z across M words on multiple word identifications is defined as follows:

$$APE_N^{\{M\}} = \frac{1}{NM}\left\{\sum_{i\,=\,1}^{N}\sum_{j\,=\,1}^{M}\mathrm{Y}_{ij}\left(\mathrm{X}_{ij} = \mathrm{x}_{ij},\ \mathrm{Z} = 1\right) - \mathrm{Y}_{ij}\left(\mathrm{X}_{ij} = \mathrm{x}_{ij},\ \mathrm{Z} = 0\right)\right\}$$

where $\mathrm{Y}_{ij}\left(\mathrm{X}_{ij} = \mathrm{x}_{ij},\ \mathrm{Z} = \mathrm{z}\right)$ indicates the word identification result of the j^{th} target word from the i^{th} subject after given the cue $\mathrm{X}_{ij} = \mathrm{x}_{ij}$ and the word study Z ($Z = 1$ means after viewing a wordlist consisting of target words and $Z = 0$ means without the word study). The superscript $\{M\}$ indicates that it is the average priming effect for M ($M \geq 2$) target words, and the subscript N indicates that the estimate is obtained from a sample of N subjects.

Again, $APE_N^{\{M\}}$ is not observable. The observable DPE in a study consisting of $N = 2K$ subjects and M target words (between the group given a word study and the group not given a word study) is as follows:

$$DPE_{2K}^{\{M\}} = \frac{1}{KM}\left\{\sum_{i\in S_Z}\sum_{j\,=\,1}^{M}\mathrm{Y}_{ij}\left(\mathrm{X}_{ij} = \mathrm{x}_{ij},\ \mathrm{Z} = 1\right)\right\} - \frac{1}{KM}\left\{\sum_{i\in S_{NZ}}\sum_{j\,=\,1}^{M}\mathrm{Y}_{ij}\left(\mathrm{X}_{ij} = \mathrm{x}_{ij},\ \mathrm{Z} = 0\right)\right\}. \tag{2}$$

Similar to a $2K$ trial study concerning one target word, the expected priming effect reduces to

$$EPE_{2K}^{\{M\}} = \frac{1}{M}\sum_{j\,=\,1}^{M} EPE_{2K,j}^{\{1\}} \tag{3}$$

where $EPE_{2K,j}^{\{1\}}$ refers to $EPE_{2K}^{\{1\}}$ for the j^{th} word.

In simple terms, $EPE_{2K}^{\{M\}}$ means that the expected priming effect estimated from $2K$ subjects across M words is the difference between the probability of corrected identifying *each* of the M target words for all subjects who have participated in the word study ($Z = 1$) and the probability of correctly identifying the corresponding word for all subjects who have not participated in the word study ($Z = 0$) averaged over M words. For simplicity, let us denote $\mathbb{P}\{\mathrm{Y}(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = 1) = 1\}$ as p_1 and $\mathbb{P}\{\mathrm{Y}(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = 0) = 1\}$ as p_0, which can be estimated by $\frac{1}{KM}\left\{\sum_{i\in S_Z}\sum_{j\,=\,1}^{M}\mathrm{Y}_{ij}\left(\mathrm{X}_{ij} = \mathrm{x}_{ij},\ \mathrm{Z} = 1\right)\right\}$ and $\frac{1}{KM}\left\{\sum_{i\in S_{NZ}}\sum_{j\,=\,1}^{M}\mathrm{Y}_{ij}\left(\mathrm{X}_{ij} = \mathrm{x}_{ij},\ \mathrm{Z} = 0\right)\right\}$, respectively.

Remark 3. **Eqs.** (1, 3) are analytical solutions to estimating the priming effect for one target word and M target words, respectively. They have been used intuitively by experimentalists; the aforementioned arguments demonstrate the mathematical validity of such usages in practice.

---

[13] Random variables related to individuals are assumed to be independent and identically distributed (*i.i.d.*). Thus, we can write the indicator function $I_{\mathrm{Y}_i(\mathrm{X}\,=\,\mathrm{x},\ \mathrm{Z}\,=\,z)\,=\,1}$ as $I_{\mathrm{Y}(\mathrm{X}\,=\,\mathrm{x},\ \mathrm{Z}\,=\,z)\,=\,1}$.

Connecting the potential outcomes framework with multinomial processing tree (MPT) models in studying priming

It turns out that the potential outcome framework-based priming study discussed here can be linked to the priming study using the multinomial processing tree (MPT) model (Batchelder and Riefer, 1999; Erdfelder et al., 2009). To see this, consider a word fragment completion test using an MPT diagram (see **Figure 3A**).

Here, let us denote A and $1-A$ as the probabilities of correctly and incorrectly, respectively, identifying the words without a word study (i.e., given $Z = 0$). Let us denote B and $1-B$ as the probabilities of storing (consciously or unconsciously) and not storing, respectively, the studied word after a word study (i.e., given $Z = 1$). Let C and $1-C$ be the probabilities of correctly and incorrectly, respectively, identifying the words if the studied words are stored in the memory; let D and $1-D$ be the probabilities of correctly and incorrectly, respectively, identifying the words if the studied words are not stored in the memory.

Naturally, $p_1 = BC + D(1-B)$ and $p_0 = A$, where p_1 and p_0 are defined previously, and the priming effect estimated using the potential outcomes framework is $p_1 - p_0$. We have $p_1 - p_0 = \{BC + D(1-B)\} - A = B(C-D) + (D-A)$. Note that (1) $B(C-D)$ is the product of consciously or unconsciously storing information from the word study into the memory (i.e., B) and the improvement[14] of word identification accuracy, thanks to the stored information [i.e., $(C-D)$]; (2) $(D-A)$ gives the difference between the probability of correctly uncovering words after a word study, even though no information from the word study has been added into the memory (to correctly identify words, one, therefore, has to either actively retrieve existing knowledge or use guessing), and the probability of uncovering words without a word study (which also relies on either existing knowledge or guessing).

Furthermore, it is not unfair to assume that A equals or is very close to D. Suppose $A = D$, then the relationship between the priming effect identified using the potential outcomes framework and the MPT model simplifies to $p_1 - p_0 = B(C-D)$. In other words, the potential priming effect (i.e., $p_1 - p_0$) chiefly depends on two factors: first, the consciously or unconsciously stored memory from the word study (i.e., B); second, the improvement of word identification accuracy, thanks to the stored information (i.e., $(C-D)$).

The aforementioned argument can be extended to studying multiple, successive word identification phases. We leave this to our readers as an exercise.

Extending the framework to experimental paradigms with exclusion and inclusion instructional conditions

Consider a three-phase study (see **Figure 3B**), where two sets of different items are shown during Phases 1 and 2 for learning purposes, and during Phase 3, the participants are given a list of items consisting of items that have appeared during Phases 1 and 2 and distractor items that have not appeared before. Subsequently, they are asked to classify them into either “old” or “new” following an inclusion instruction or an exclusion instruction (Buchner et al., 1995). Under an inclusion instruction, the participants need to call an item *old* if it has appeared in either Phase 1 or 2 and call a distractor item *new*. Under an exclusion instruction, the participants need to call an item *old* only if it has appeared in Phase 2, and *new* otherwise.

The framework proposed in this article can also be modified to study the experimental paradigm with exclusion and inclusion instructional conditions. To demonstrate this, let us define Z_1 and Z_2 as two lists of items during Phases 1 and 2, respectively. Let X and X' denote the outcomes of the identification during Phase 3 under inclusion and exclusion instructions, where their realizations are either {new} or {old} for each given item. Following similar arguments as before, we define the potential difference between the results from the inclusion and exclusion instructions as follows:

$$\frac{1}{NM}\left\{\sum_{i=1}^{N}\sum_{j=1}^{M} \mathrm{Y}_{ij}(\mathrm{X}_{ij} = \mathrm{x}_{ij}, Z_1 = 1, Z_2 = 1) - \mathrm{Y}_{ij}(X'_{ij} = x'_{ij}, Z_1 = 1, Z_2 = 1)\right\}.$$

Note that here, it is assumed that the inclusion and exclusion instructions are given to the same participants in a group. This is not ideal as repeating Phase 3 under different conditions may bias the results. Using the potential outcome framework, this scenario can be estimated as follows:

$$\frac{1}{KM}\left\{\sum_{i\in S_I}\sum_{j=1}^{M} \mathrm{Y}_{ij}\left(\mathrm{X}_{ij} = \mathrm{x}_{ij}, Z_1 = 1, Z_2 = 1\right)\right\} - \frac{1}{KM}\left\{\sum_{i\in S_E}\sum_{j=1}^{M} \mathrm{Y}_{ij}\left(X'_{ij} = x'_{ij}, Z_1 = 1, Z_2 = 1\right)\right\}$$

where the two parts (before and after the minus sign) are estimated from subjects in groups S_I (following the inclusion instruction) and S_E (following the exclusion instruction), respectively. Note that the aforementioned result equals to $p_{i-}p_e$ in Buchner et al. (1995), which quantifies the probability of consciously recollecting a Phase 1 item.

---

[14] It is natural to assume $C = D$; otherwise, we can replace "improvement" with "the difference."

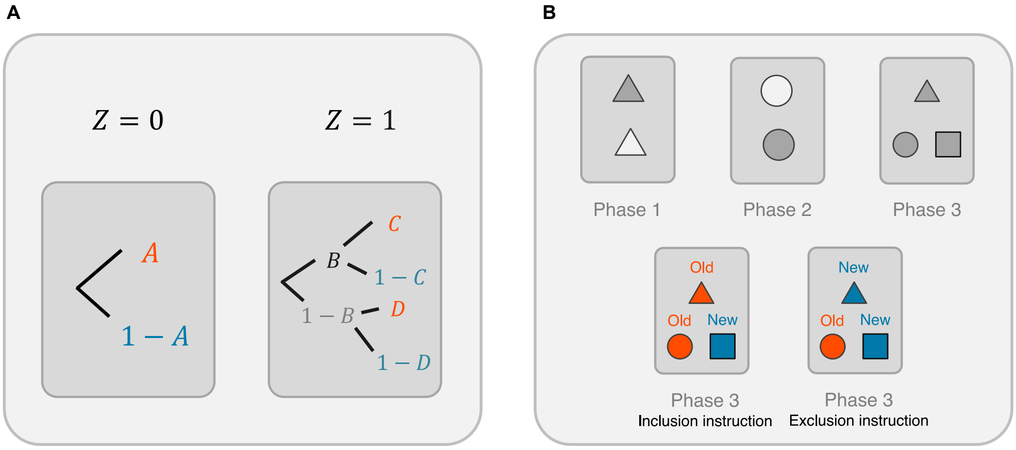


FIGURE 3
Linking the potential outcomes-based priming research with two prominent quantitative psychological methods. **(A)** Multinomial processing tree (MPT) model for studying the priming effect. Here, A and $1 - A$ are the probabilities of correctly and incorrectly, respectively, identifying the words without a word study (i.e., given $Z = 0$); B and $1 - B$ are the probabilities of storing (consciously or unconsciously) and not storing, respectively, the studied word after a word study (i.e., given $Z = 1$); C and $1 - C$ are the probabilities of correctly and incorrectly, respectively, identifying the words if the studied words are stored in the memory; D and $1 - D$ are the probabilities of correctly and incorrectly, respectively, identifying the words if the studied words are not stored in the memory. **(B)** A three-phase experiment with exclusion and inclusion instructional conditions. Top left: A set of items is shown during the Phase 1 study. Top middle: Another set of items (different from those in Phase 1) is shown during Phase 2. Top right: During Phase 3, the participants are given a list of items consisting of those who have appeared during Phases 1 and 2 and distractor items that have not appeared before. Subsequently, they are asked to classify them into either "old" or "new" following an inclusion instruction or an exclusion instruction. Bottom left: Under the inclusion instruction, participants need to call an item old if it has appeared in either Phase 1 or 2 and call a distractor item new. Bottom right: Under the exclusion instruction, participants need to call an item old only if it has appeared in Phase 2, and new otherwise.

Testing the significance of the priming effect

Returning to the priming study, although the focus of this article so far has been to define and quantify the causal priming effect, it may also be important for investigators to test whether a detected causal effect is significant. For example, consider 100 subjects who have undertaken the word study and 100 people who have not undertaken the word study. Suppose the estimated expected priming effect (*EPE*) is 0.1; is 0.1 in a sample of 200 subjects significant (from 0, where 0 indicates no priming effect)? What if the estimated *EPE* is 0.05?

One way to answer this question is to conduct a hypothesis test on whether the estimated priming effect is significant; that is, to verify the (alternative) hypothesis that the *EPE* is significantly greater than zero. Thanks to **Eqs.** 1, 3, the *EPE* can be written in terms of probability and can therefore be examined using a proportion test (Ott and Longnecker, 1980; Bickel and Doksum, 2000).

Formally, the test statistic is defined as follows:

$$z = \frac{DPE}{\sqrt{\hat{p}\,(1-\hat{p})\,(\frac{1}{N_1}+\frac{1}{N_0})}} \qquad (4)$$

where $\hat{p}_1 = \frac{1}{KM}\sum_{i\in S_Z}\sum_{j=1}^{M} \mathrm{Y}_{ij}\left(\mathrm{X}_{ij} = \mathrm{x}_{ij},\ \mathrm{Z} = 1\right)$, $\hat{p}_0 = \frac{1}{KM}\sum_{i\in S_{NZ}}\sum_{j=1}^{M} \mathrm{Y}_{ij}\left(\mathrm{X}_{ij} = \mathrm{x}_{ij},\ \mathrm{Z} = 0\right)$, $DPE = \hat{p}_1 - \hat{p}_0$, $\hat{p} = \frac{1}{2KM}\sum_{i\in S_Z\cup S_{NZ}}\sum_{j=1}^{M} \mathrm{Y}_{ij}(\mathrm{X}_{ij} = \mathrm{x}_{ij},\ Z = \{0, 1\})$, and $N_1 = N_2 = KM$ [15]. Here, the *DPE* is the empirical estimate of the *EPE* obtained from **Eq.** 2 and $\hat{p}$ is the pooled probability (from both groups) of correct word identification (in other words, the overall probability of correctly identifying a word when the group that undertaken a word study and the group that did not undertake a word study are combined). The hat symbol, for example, in $\hat{p}_1$ is an estimate of p_1.

One can then compare the p-value associated with the z score to evaluate the significance. Note that the aforementioned z-test is the same as a Chi-square test, where the z-statistic is equal to the square root of the Chi-square statistic, and the p-values of the two tests are identical. When the word studies are multivariate (e.g., there are more than two types of word study), continuous (e.g., the word study involves words with different degrees of complexity), or time-dependent (e.g., several tests are carried out with large time intervals in between), more advanced statistical tests can be used (see the Discussion section for details).

Subsequently, the 100 $(1 - \alpha)$ percent confidence interval (Wilson, 1927; Newcombe,

---

[15] Readers can relatively easily extend it to more complicated cases involving unbalanced groups, where the numbers of subjects and/or target words in two groups are unequal.

1998) for the estimated priming effect is as follows:

$$\left(DPE - z_{(1-\frac{\alpha}{2})}\sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{N_1}+\frac{\hat{p}_0\left(1-\hat{p}_0\right)}{N_0}},\right.$$

$$\left.DPE + z_{(1-\frac{\alpha}{2})}\sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{N_1}+\frac{\hat{p}_0\left(1-\hat{p}_0\right)}{N_0}}\right).$$

Estimating priming effects with covariates

Although the word study Z is the primary factor that affects the outcome Y, it remains possible that there exist additional variables (denoted as W) that, if not considered, may bias the estimation of the causal priming effect. These variables could either have a causal relationship with the word identification outcome Y (e.g., take W as intelligence) or are spuriously (i.e., by chance) correlated with the outcome in the sample (e.g., one's height). Randomization only ensures that in expectation, the covariates are balanced between the two treatment groups. There, however, could still be chance imbalances in the covariates between the two treatment groups; in this case, adjusting for the covariates will increase the signal-to-noise ratio[16] and make the priming effect more likely to be detected, if exists.

For example, take W as one's IQ, which may affect word identification. Consider 20 subjects with a mean IQ of 100 (10 with IQ larger than 100 and 10 with IQ less than 100). Certainly, we could create two splits with each split containing five individuals with above-average IQ and five with below-average IQ. Our point is that sometimes, such a balanced sample is difficult to obtain, and thus, protective measures need to be taken instead (a good example here is the proficiency in English language – another variable that may affect word identification). In practice, however, it is difficult and costly for researchers to collect samples that contain subjects that are perfectly matched. Thus, we proceed here assuming such a (not completely matched) case occurs. For example, if we are to randomly assign a word study ($Z = 1$) to 10 subjects and no word study ($Z = 0$) to another 10 subjects, the group with the word study may contain eight subjects with above-average IQ and the other group with two subjects with above-average IQ. Then, the result using **Eq.** 2 could potentially over-estimate the priming effect since there are more people with above-average IQ in the word study group.

The effect of an additional variable can be adjusted in a logistic regression model. Specifically, consider

$$logit(\mathbb{P}\{\mathrm{Y}_i(\mathrm{X} = \mathrm{x},\ \mathrm{Z} = z_i) = 1\}) = \beta_0 + \beta_z z_i + \beta_w w_i$$

where z_i indicates whether the i^{th} subject receives a word study or not, w_i is the IQ for the i^{th} subject, β_0 is the estimated intercept, and β_z and β_w are the estimated parameters for z_i and w_i, respectively. The estimated β_z then indicates the priming effect from Z, when it is adjusted for the IQ effect (W). Specifically, controlling (i.e., removing) the effect from IQ to word identification Y, β_z quantifies the priming effect: the probability of correctly identifying a word increases $\frac{e^{\beta_z}}{1+e^{\beta_z}}$ when an individual conducts a word study versus not conducting a word study. Again, the logistic formula is stated for a word study considering one target word with the same cue $X = x$ and can be relatively easily extended to a study considering multiple words and multivariate covariates.

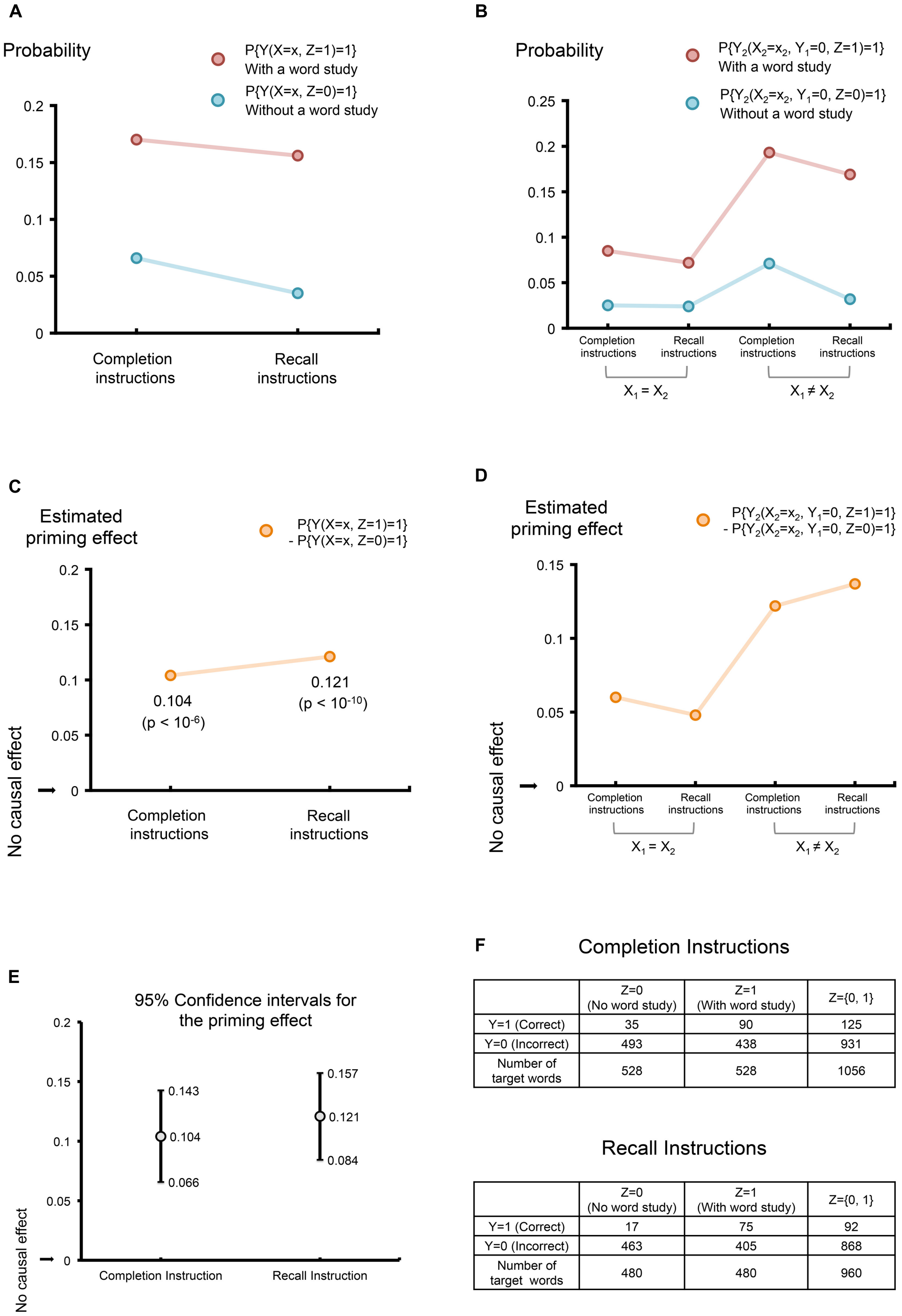
In the following, we will perform data analysis using data from a between-subjects study (Hayman and Tulving, 1989) to demonstrate how to use the framework to study the potential causal priming effect. The advantage of using a between-subject study is that the priming effect can be evaluated when the same experiment cannot be run on the same subjects more than one time; it may also reduce the likelihood of carryover or learning effect in a repeated-measures design (see section "Discussion").

Results

Consider a between-subjects WFC test. A total of 84 students enrolled in a second-year psychology course at the University of Toronto were randomly divided into two groups (one with the last name A-K and the other with the last name L-Z). A set of 48 target English words of intermediate difficulty was selected from a word pool and divided into two wordlists (A and B), with 24 target words in each list. An additional 64 (non-target) English words were used as buffer words. During the study phase, the first group studied wordlist A and the second wordlist B; the wordlist B thus served as non-studied words for the first group, and wordlist A served as non-studied words for the second group. During the test phase, there were two test instructions: the subjects with completion instructions were asked to complete the fragment with any word that comes to mind; subjects with recall instructions were asked to complete the fragment only with studied words. All subjects are randomized into four groups, each to take two tests. Specifically, participants in Group 1 ($N = 22$) conducted two tests under the completion instructions with the same fragment cues during the two tests; participants in Group 2 ($N = 22$) conducted two tests under the completion instructions with different fragment cues during the two tests; participants in Group 3 ($N = 20$) conducted two tests under the recall instructions with the same fragment cues during the two tests; participants in Group 4 ($N = 20$) conducted two tests under the recall instructions with different fragment cues during the two tests. Full data description is available in Experiment 2 in Hayman and Tulving (1989) with study data summarized in **Figure 4F**.

---

[16] Statistically speaking, adjusting for covariates in a randomized study improves the precision (reduces the variance) of the treatment effect estimator (see Moore et al., 2011; Qian et al., 2018).



Completion Instructions

	Z=0 (No word study)	Z=1 (With word study)	Z={0, 1}
Y=1 (Correct)	35	90	125
Y=0 (Incorrect)	493	438	931
Number of target words	528	528	1056

Recall Instructions

	Z=0 (No word study)	Z=1 (With word study)	Z={0, 1}
Y=1 (Correct)	17	75	92
Y=0 (Incorrect)	463	405	868
Number of target words	480	480	960

FIGURE 4

Estimating the causal priming effect. **(A)** Probabilities of correctly uncovering target words given fractional cues, under the completion instructions and the recall instructions. The color red is used to indicate experiments involving a word study including target words; the color blue is used to indicate experiments involving a word study without target words (abbreviated as "without a word study"). **(B)** The probabilities of correctly uncovering target words during a second cue-based test, given that the first test failed to uncover the same word. Four experimental sceneries were considered, combining two instruction strategies: the completion instructions and the recall instructions, and two types of cues: the same cues and different cues during two tests. X_1 and X_2 refer to cues from the first test and the second test, respectively; $X_1 = X_2$ indicates the same cues were used in the two tests; $X_1 \neq X_2$ indicates different cues were used in the two tests. Again, the color red is used to indicate experiments involving a word study including target words; the color blue is used to indicate experiments involving a word study without target words (abbreviated as "without a word study"). **(C)** The estimated causal priming effects and their *p*-values correspond to **(A)**. **(D)** The estimated causal priming effects correspond to **(B)**. **(E)** The estimated 95% confidence intervals for the priming effect in **(C)**. **(F)** Data used in estimating the priming effect corresponding to **(A,C,E)**. Data adapted by permission from RightsLink Permissions American Psychological Association "Is priming in fragment completion based on a 'traceless' memory system?" by Hayman and Tulving (1989), American Psychological Association.

The potential priming effect of the WFC test is displayed in **Figure 4**. Using **Eqs.** 2, 4, the *DPE* under completion instruction is 0.104 ($z = 5.23$, $p = 10^{-6}$), with a 95% confidence interval (0.066, 0.143); the *DPE* under recall instructions is 0.121 ($z = 6.36$, $p = 10^{-10}$), with a 95% confidence interval (0.084, 0.157). The corresponding estimated probabilities (of correct word identification with or without a word study), priming effects and their confidence intervals are shown in **Figures 4A,C,E**, respectively.

Next, we consider the *DPEs* where the word study primes the identification of a target word in the second test, given the identification of the same target word failed during the first test (see **Figures 4B,D**). Although the number of the words that failed to be identified during the first test was not reported in Hayman and Tulving (1989) (hence we cannot compute the exact p-values), readers could follow the previous example and use **Eq.** 4 in their research when data are available. Nevertheless, we will report the *DPEs* without p-values. There are two reasons for this. First, since the total number of the target words studied is large [i.e., 480 words (20 subjects each with 24 target words) and 528 words (22 subjects each with 24 target words) in our case], a positive *DPE* is likely to yield a significance non-zero priming effect. Second, it allows us to numerically compare the priming effects under different experimentation strategies. Specifically, using the same cues in the two tests, the *DPE* under the completion instructions is 0.06; the *DPE* under the recall instructions is 0.048; meanwhile, using the different cues in the two tests, the *DPE* under the completion instructions is 0.122; the *DPE* under the recall instructions is 0.137 (see **Figure 4D**). The much stronger priming effect observed in both experiments where different cues are provided suggests that given a failed attempt using one cue during the first test, information has potentially been learned by combining the first cue and a *different* cue during the second test.

Extensions, limitations, future directions, and final remarks

In this article, we defined, quantified, and tested the priming effect using the *potential outcomes framework*. Although we only considered cases involving a binary exposure (having a word study versus not having a word study), the framework can be extended to categorical exposures (e.g., we can code an exposure that does not consist of a word study as $Z = 0$, one that consists of a word study including short words as $Z = 1$, and one that consists of a word study including long words as $Z = 2$). In addition, the framework can be extended to continuous exposures (e.g., when a word study consists of words with different degrees of complexity, we can allow Z to take any value between 0 and 100 to indicate complexity of each word). Furthermore, it can be extended to time-dependent exposures [e.g., we can write word studies conducted at different times as $Z(t)$, for each time t]. Finally, it can also be extended to cases where several exposures are concerned (e.g., let $Z_1 =$ reading a list of words, $Z_2 =$ viewing a list of non-word symbols, and $Z_3 =$ listening to a list of words), where the priming effect for each exposure can be estimated when the other exposures are controlled. For example, when estimating the priming effect of symbol recognition ($Z_2 = 1$ versus $Z_2 = 0$), one could fix Z_1 and Z_3; namely, the priming effect can be estimated, for example, using $Y(\mathbf{X} = x, Z_1 = 0, Z_2 = 1, Z_3 = 0) - Y(\mathbf{X} = x, Z_1 = 0, Z_2 = 0, Z_3 = 0)$, where bold X indicates all cues used for three studies.

Eq. 4 is used to test the significance of the priming effect with binary exposures. When the exposures are multivariate, continuous, or time-dependent, the test can be carried out by first arranging the exposure and outcome as explanatory and dependent variables in a regression setting, and then testing the exposure effect by examining the significance of the (regression) parameters. For example, when there are three types of word studies (no word study, a word study with short words, and a word study with long words), one can consider a regression model with a block design, where each block consists of subjects from one of the three groups. The estimated regression parameter for the block variable then indicates the priming effect between two paired groups. When the exposure is time-dependent[17], one could refer to functional regression models, wherein $Z(t)$ and $Y(t)$ are treated as functional regressors and responses, respectively (Ramsay and Silverman, 1997).

The proposed method aimed at providing a framework that could estimate and validate analytically priming in between-subjects designs. It nonetheless has a few limitations. First, we demonstrated the method using data from previous experiments (that are not primarily intended to evaluate the priming effect but to assess the independence of successive tests). Inevitably, this restricted our arguments; future research may verify and expand our analysis to general priming research. Future studies may also extend to cases with a larger sample, and non-twin studies need to examine covariant control under the potential outcomes framework (see section "Estimating Priming Effects With Covariates") and its utility on providing an estimated priming effect that is less biased. In parallel, future research may further consider twin studies where the subjects are nearly perfectly matched. Second, the method we introduced rests on the Neyman–Rubin potential outcomes framework. There is, however, on the one hand, not as of yet a consensus that one causal framework is better than others, although we have discussed the advantages of the Neyman–Rubin framework in estimating potential priming effect (especially its mathematical representations). On the other hand, we recognize that despite

---

[17] For example, at each time t, a study $Z(t)$ is assigned to individuals, and a word identification $Y(t)$ is observed. This is particularly useful when the sample consists of subjects whose implicit memory degenerates over time.

differences and disagreements [e.g., see Gelman's blog post (Gelman, 2009) and discussion under the post], there is some commonality between Neyman–Rubin's and Pearl's frameworks (Section 7.4.5 of Pearl, 2009b), and there exists "a happy symbiosis between graphs and counterfactual notation" (Section 7.4.4 of Pearl, 2009b). In this study, while we present the arguments using the Neyman–Rubin model, we have adopted Pearl's diagram representation (although without graphic notations) to illustrate the experiments. We do so without implying that one framework is superior to the other. Future studies may theoretically compare the Neyman–Rubin approach with Pearl's approach in detail for studying potential causal priming (e.g., their mathematical or empirical equivalence or difference). Further research may also incorporate Campbell's approach to identify potential threats that may impair the validity of inferences made on the estimated priming effect.

Sometimes, investigators studying the priming effect may observe post-treatment variables (i.e., variables obtained after the word study Z is assigned). Examples of post-treatment variables are (a) a measure of subjects' compliance to the originally assigned word study – a subject chooses not to take the word study after it is assigned; (b) in studies with a long time interval between two priming tests, whether or not the subject drops out is a post-treatment variable (missingness of outcome); (c) in longitudinal (priming) studies involving patients with severe amnesia, the outcome can be censored (i.e., not recorded due to death); (d) in studies investigating priming effects for patients with brain disorders, surrogate variables of disease progression and fluctuation, such as the degree of memory loss, are post-treatment variables. The estimators provided in this article can only be used to adjust for pre-treatment variables; if one adjusts for post-treatment directly using the framework outlined in this article, the estimated effects are no longer causal (Frangakis and Rubin, 2002).

It is worthwhile noting that besides the potential outcome framework (by comparing outcomes on randomly selected or matched subjects, or subjects with covariates adjusted), priming can also be estimated using a repeated-measures design, in which *all* subjects are exposed first to half of the target words and then another half of the target words (e.g., see Challis and Brodbeck, 1992). The priming effect can then be estimated as the difference between the proportion of fragments of studied words completed and the proportion of fragments of non-studied words completed for the same subject (and averaged across all subjects). Instead of matching two groups of subjects as proposed in this study, the key to using the repeated-measures designs is to match the length, frequency, etc., of the words and randomize the words employed. Whereas this indeed provides an alternative (and potentially convenient[18]) approach to assess the priming effect, and we welcome future research to compare this approach with the potential outcomes framework; a key concern with this method is the carryover or learning effect. The carryover or learning effect here is not necessarily the phenomenon where after studying the same (or similar) words multiple times, the earlier word study and identification may improve the same subject's later word identification; rather, it also includes the phenomenon where the experiment mechanism of the first repeated-measures study may improve learning during the second repeated-measures study. We have seen such a carryover or learning effect during a smartphone-based cognitive test, where even though different tests (e.g., drawing different shapes) were given to the same subjects over time, their performance improved. For the WFC test, it may be possible that the subjects learned some rules (despite not being informed) during the first half of the experiment or became more focused during the second half either because they had guessed the approximate rule or because they had realized that the word study may be an important part (since, for example, two wordlists had been given sequentially) to their performance of the experiment. Future studies could examine the existence of such a learning or carryover effect, and if exists, whether and how it would affect estimating the priming effect.

There are times where even randomization becomes impossible. For example, suppose one is interested in studying how a new medicine affects priming; in this case, we have two potential causes: a word study (Z) and medication (Med). It is unethical to assign a group of 45-year-old healthy subjects to take a new drug to investigate whether the drug improves priming at 50. In addition, there is likely another source, say, the socioeconomic status (which may be related to the affordability of new drugs) or genetics (if there is a family history of memory problems, one may be more willing to take the drug), that may be associated with taking the drug and/or developing memory problems at 50. Similarly, it would be difficult to estimate the effect of taking the drug on improving priming by comparing the performance of an individual at 50 who had taken the drug with his or her performance at 50 had he or she not taken the drug. To solve these issues, the propensity score matching (PSM) estimates the treatment effect by comparing the outcomes of the subjects under treatment (e.g., taking the drug) with a set of "matched" subjects without treatment (e.g., having not taken the drug) (Rosenbaum and Rubin, 1983; Dehejia and Wahba, 1999, 2002; Caliendo and Kopeinig, 2008). More concretely, one could first compute the propensity score of A's and B's taking the drug based on their gender, economic, social, genetic, and demographic backgrounds, and choose two individuals C and D from a group of 50-year-olds who had not taken the drug but have propensity scores (of taking the drug during their younger years) closest to A's and B's, respectively. Subsequently, A and C will receive a word study, and B and D will not. Following

---

[18] For example, when there are no matched samples (but see Propensity Score Matching (PSM) and covariates adjustment discussed in this paper), and that the carryover or learning effect is ignorable, a repeated-measures design is attractive.

the previous notations, we have $Y_A\left(X_j, Z = 1, Med = 1\right)$, $Y_B(X_j, Z = 0, Med = 1)$, $Y_C(X_j, Z = 1, Med = 0)$, and $Y_D\left(X_j, Z = 0, Med = 0\right)$, for $1 \leq j \leq M$, where M target words are considered. Then, the priming effects for the group taking the drug and the one not taking the drug are $DPE_{Med=1} = \frac{1}{M}\left\{\sum_{j=1}^{M} Y_A\left(X_j, Z = 1, Med = 1\right) - \sum_{j=1}^{M} Y_B\left(X_j, Z = 0, Med = 1\right)\right\}$ and $DPE_{Med=0} = \frac{1}{M}\left\{\sum_{j=1}^{M} Y_C\left(X_j, Z = 1, Med = 0\right) - \sum_{j=1}^{M} Y_D\left(X_j, Z = 0, Med = 0\right)\right\}$, respectively. Subsequently, we can estimate the drug effect on priming using $DPE_{Med=1} - DPE_{Med=0}$. Note that for simplicity, only one individual is considered for each of the 2 × 2 factors; one can relatively easily extend the above to include multiple subjects in each group.

Although we have throughout focused on a type of non-semantic priming, other studies have reported that new semantic knowledge can be acquired among (even) patients with amnesia. For example, the learning of specified target words in meaningful texts, statements of facts about people and places, specified target words as parts of meaningful sentences, new computer-related vocabulary, computer commands, semantic interpretations of ambiguous descriptions of situations and events, and production of words to cues consisting of the initial letters of words (see Hayman et al., 1993 for a summary of studies). Future studies should independently verify the extent to which the framework introduced in this article can be used to estimate causal semantic priming. A beginning can, perhaps, be made by reporting the individual ratings of the meaningfulness of the target words (e.g., during a word study, every participant is to rate on a scale of 0–10, the meaningless of each studied word), and subsequently treating the ratings as covariates.

In conclusion, we define, quantify, and test the causal priming effect using the *potential outcomes* framework. Applying data from a between-subjects word completion test, we demonstrate that the framework identifies a significant priming effect from a word study to cue-based word identification, under both completion and recall instructions; the priming effect under the recall instructions is more significant than that under the completion instructions. Furthermore, when there are two consecutive tests, the framework shows that even if the word identification failed during the first test, there is likely a priming effect from the initial word study to the second word identification, regardless of the type of instructions and whether the same or different cues are used in the two tests. In addition, there is a stronger priming effect in experiments where different cues are provided, suggesting that given a failed attempt using one cue during the first test, additional information may have been learned by combining the first cue and a different cue during the second test. Finally, our explorations show that what has been intuitively used by scholars to estimate the priming effect in the past has a meaningful mathematical basis.

Data availability statement

Publicly available datasets were analyzed in this study. These data can be found here: Hayman and Tulving (1989).

Ethics statement

The studies involving human participants were reviewed and approved by the University of Toronto. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

OC designed the potential causal priming framework (based on Neyman and Rubin's works on causal inference and Tulving and Schacter's works on priming) and performed the analysis. HP provided computational support. HC and GN provided neurobiological and psychological interpretations. TQ provided statistical support. MV provided funding, support, and guidance. OC wrote the manuscript, with comments from all other authors. All authors contributed to the article and approved the submitted version.

Acknowledgments

We are grateful to Semir Zeki for his useful comments on earlier versions of the article. We thank the two reviewers for their insightful comments and suggestions, which have significantly improved the range, depth, and quality of the article.

Conflict of interest

Author GN is Medical Director of Neurology at, and minority shareholder of, icometrix. OC had consulted for F. Hoffmann-La Roche.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Baddeley, A. D., and Hitch, G. (1974). Working memory. *Psychol. Learn. Motiv.* 8, 47–89. doi: 10.1016/S0079-7421(08)60452-1

Baddeley, A., Lewis, V., and Vallar, G. (1984). Exploring the articulatory loop. *Q. J. Exp. Psychol. Sect. A* 36, 233–252. doi: 10.1080/14640748408402157

Batchelder, W. H., and Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychon. Bull. Rev.* 6, 57–86. doi: 10.3758/BF03210812

Bickel, P. J., and Doksum, K. A. (2000). *Mathematical statistics: Basic ideas and selected topics, Vol I*, 2nd Edn, Boca Raton, FL: CRC Press.

Buchner, A., Erdfelder, E., and Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *J. Exp. Psychol. Gen.* 124, 137–160. doi: 10.1037/0096-3445.124.2.137

Caliendo, M., and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *J. Econ. Surv.* 22, 31–72. doi: 10.1111/j.1467-6419.2007.00527.x

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychol. Bull.* 54, 297–312. doi: 10.1037/h0040950

Cermak, L. S., Talbot, N., Chandler, K., and Wolbarst, L. R. (1985). The perceptual priming phenomenon in amnesia. *Neuropsychologia* 23, 615–622. doi: 10.1016/0028-3932(85)90063-6

Challis, B. H., and Brodbeck, D. R. (1992). Level of processing affects priming in word fragment completion. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 595–607. doi: 10.1037/0278-7393.18.3.595

Cochran, W. G., and Chambers, S. P. (1965). The planning of observational studies of human populations. *J. R. Stat. Soc. Ser. A* 128, 234–266. doi: 10.2307/2344179

Cohen, N. J., and Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science* 210, 207–210. doi: 10.1126/science.7414331

Cooper, G., Tindall-Ford, S., Chandler, P., and Sweller, J. (2001). Learning by imagining. *J. Exp. Psychol. Appl.* 7, 68–82. doi: 10.1037/1076-898X.7.1.68

Dehejia, R. H., and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *J. Am. Stat. Assoc.* 94, 1053–1062. doi: 10.1080/01621459.1999.10473858

Dehejia, R. H., and Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Rev. Econ. Stat.* 84, 151–161. doi: 10.1162/003465302317331982

Ding, P. (2017). A paradox from randomization-based causal inference. *Stat. Sci.* 32, 331–345. doi: 10.1214/16-STS571

Ding, P., Feller, A., and Miratrix, L. (2016). Randomization inference for treatment effect variation. *J. R. Stat. Soc. Ser. B* 78, 655–671. doi: 10.1111/rssb.12124

Erdfelder, E., Auer, T. S., Hilbig, B. E., Aßfalg, A., Moshagen, M., and Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychol.* 217, 108–124. doi: 10.1027/0044-3409.217.3.108

Frangakis, C. E., and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* 58, 21–29. doi: 10.1111/j.0006-341X.2002.00021.x

Funnell, E. (1983). Phonological processes in reading: New evidence from acquired dyslexia. *Br. J. Psychol.* 74, 159–180. doi: 10.1111/j.2044-8295.1983.tb01851.x

Gelman, A. (2009). *Resolving disputes between J. Pearl and D. Rubin on causal inference*. Available online at: https://statmodeling.stat.columbia.edu/2009/07/05/disputes_about/ (accessed September 26, 2021).

Goldberger, A. S. (1972). *Selection bias in evaluating treatment effects: Some formal illustrations*. Discussion Paper 123–172. Madison, WI: Institute for Research on Poverty.

Graf, P., Squire, L. R., and Mandler, G. (1984). The information that amnesic patients do not forget. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 164–178. doi: 10.1037/0278-7393.10.1.164

Hanley, J. R., and Broadbent, C. (1987). The effect of unattended speech on serial recall following auditory presentation. *Br. J. Psychol.* 78, 287–297. doi: 10.1111/j.2044-8295.1987.tb02247.x

Hashtroudi, S., Parker, E. S., DeLisi, L. E., Wyatt, R. J., and Mutter, S. A. (1984). Intact retention in acute alcohol amnesia. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 156–163. doi: 10.1037/0278-7393.10.1.156

Hayman, C. A. G., and Tulving, E. (1989). Is priming in fragment completion based on a "traceless" memory system? *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 941–956. doi: 10.1037/0278-7393.15.5.941

Hayman, C. A. G., Macdonald, C. A., and Tulving, E. (1993). The role of repetition and associative interference in new semantic learning in amnesia: A case experiment. *J. Cogn. Neurosci.* 5, 375–389. doi: 10.1162/jocn.1993.5.4.375

Hill, A. B. (1965). The environment and disease: Association or causation? *Proc. R. Soc. Med.* 58, 295–300. doi: 10.1177/003591576505800503

Hinkelmann, K., and Kempthorne, O. (2005). *Design and analysis of experiments*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/0471709948

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *J. Mem. Lang.* 30, 513–541. doi: 10.1016/0749-596X(91)90025-F

Johnson, M. K. (1983). "A multiple-entry, modular memory system," in *The psychology of learning and motivation*, ed. G. H. Bower (New York, NY: Academic Press), 81–123. doi: 10.1016/S0079-7421(08)60097-3

Kopelman, M. D., and Corn, T. H. (1988). Cholinergic "blockade" as a model for cholinergic depletion: A comparison of the memory deficits with those of alzheimer-type dementia and the alcoholic korsakoff syndrome. *Brain* 111, 1079–1110. doi: 10.1093/brain/111.5.1079

Light, L. L., Singh, A., and Capps, J. L. (1986). Dissociation of memory and awareness in young and older adults. *J. Clin. Exp. Neuropsychol.* 8, 62–74. doi: 10.1080/01688638608401297

Mishkin, M., Malamut, B., and Bachevalier, J. (1984). "Memories and habits: Two neural systems," in *The neurobiology of learning and memory*, eds G. Lynch, J. L. McGaugh, and N. M. Weinberger (New York, NY: Guilford press), 65–77.

Moore, K. L., Neugebauer, R., Valappil, T., and van der Laan, M. J. (2011). Robust extraction of covariate information to improve estimation efficiency in randomized trials. *Stat. Med.* 30, 2389–2408. doi: 10.1002/sim.4301

Newcombe, R. G. (1998). Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Stat. Med.* 17, 873–890. doi: 10.1002/(SICI)1097-0258(19980430)17:8<873::AID-SIM779>3.0.CO;2-I

Neyman, J. (1923). On the application of probability theory to agricultural experiments Appeared (in Polish). *Rocz. Nauk Rolniczych.* 10, 1–51.

Nissen, M. J., Knopman, D. S., and Schacter, D. L. (1987). Neurochemical dissociation of memory systems. *Neurology* 37, 789–789. doi: 10.1212/WNL.37.5.789

Ott, R. L., and Longnecker, M. T. (1980). *An introduction to statistical methods and data analysis*, 6th Edn. Boston, MA: Cengage.

Parker, E. S., Schoenberg, R., Schwartz, B. L., and Tulving, E. (1983). "Memories on the rising and falling blood-alcohol curve," in *Paper presented at the twenty-fourth annual meeting of the Psychonomic Society*, San Diego, CA.

Parkin, A. J., and Streete, S. (1988). Implicit and explicit memory in young children and adults. *Br. J. Psychol.* 79, 361–369. doi: 10.1111/j.2044-8295.1988.tb02295.x

Pearl, J. (1993). Comment: Graphical models, causality and intervention. *Statist. Sci.* 8, 266–269. doi: 10.1214/ss/1177010894

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika* 82, 669–688. doi: 10.1093/biomet/82.4.669

Pearl, J. (2001). "Direct and indirect effects," in *Proceedings of the seventeenth conference on uncertainty in artificial intelligence*, (San Francisco, CA: Morgan Kaufmann Publishers Inc), 411–420.

Pearl, J. (2009a). Causal inference in statistics: An overview. *Stat. Surv.* 3, 96–146. doi: 10.1214/09-SS057

Pearl, J. (2009b). *Causality: Models, reasoning, and inference*, 2nd Edn. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511803161

Peters, C. C. (1941). A method of matching groups for experiment with no loss of population. *J. Educ. Res.* 34, 606–612. doi: 10.1080/00220671.1941.10881036

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., and Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* 331, 585–589. doi: 10.1038/331585a0

Pribram, K. H. (1984). "Brain systems and cognitive learning process," in *Animal cognition*, eds H. L. Roitblat, T. G. Bever, and H. S. Terrace (Mahwah, NJ: Lawrence Erlbaum Associates. Inc), 627–656.

Qian, T., Rosenblum, M., and Qiu, H. (2018). *Improving power in group sequential, randomized trials by adjusting for prognostic baseline variables and short-term outcomes*. Baltimore, MD: Johns Hopkins University, Dept. Available online at: https://biostats.bepress.com/jhubiostat/paper285/ (accessed Septermber 21, 2022).

Ramsay, J. O., and Silverman, B. W. (1997). *Functional data analysis*. New York, NY: Springer-Verlag New York. doi: 10.1007/978-1-4757-7107-7

Rosenbaum, P. R., and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41–55. doi: 10.1093/biomet/70.1.41

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* 66, 688–701. doi: 10.1037/h0037350

Rubin, D. B. (1977). Assignment to a treatment group on the basis of a covariate. *J. Educ. Stat.* 2, 1–26. doi: 10.3102/10769986002001001

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Stat.* 6, 34–58. doi: 10.1214/aos/1176344064

Sartori, G., Masterson, J., and Job, R. (1987). "Direct-route reading and the locus of lexical decision," in *The cognitive neuropsychology of language*, eds M. Coltheart, G. Sartori, and R. Job (Mahwah, NJ: Lawrence Erlbaum Associates. Inc), 59–77.

Schacter, D. L., Cooper, L. A., and Delaney, S. M. (1990). Implicit memory for unfamiliar objects depends on access to structural descriptions. *J. Exp. Psychol. Gen.* 119, 5–24. doi: 10.1037/0096-3445.119.1.5

Scheffe, H. (1959). *The analysis of variance*. New York, NY: Wiley.

Schwartz, M. F., Saffran, E. M., and Marin, O. S. M. (1980). "Fractionating the reading process in dementia: Evidence for word-specific print-to-sound association," in *Deep dyslexia*, eds M. Coltheart, J. C. Marshall, and K. E. Patterson (London: Routledge & Kegan Paul) 259–269.

Shimamura, A. P. (1986). Priming effects in amnesia: Evidence for a dissociable memory function. *Q. J. Exp. Psychol. Sect. A* 38, 619–644. doi: 10.1080/14640748608401617

Squire, L. R. (1987). *Memory and brain*. Oxford: Oxford University Press.

Tulving, E. (1985). How many memory systems are there? *Am. Psychol.* 40, 385–398. doi: 10.1037/0003-066X.40.4.385

Tulving, E., and Schacter, D. L. (1990). Priming and human memory systems. *Science* 247, 301–306. doi: 10.1126/science.2296719

Tulving, E., and Schacter, D. L. (1992). "Priming and memory systems," in *Neuroscience year: Supplement 2 to the encyclopedia of neuroscience*, eds B. Smith and G. Adelman (Boston, MA: Birkhäuser Boston).

Warrington, E. K., and Weiskrantz, L. (1974). The effect of prior learning on subsequent retention in amnesic patients. *Neuropsychologia* 12, 419–428. doi: 10.1016/0028-3932(74)90072-4

Warrington, E., and Taylor, A. (1978). Two categorical stages of object recognition. *Perception* 7, 695–705. doi: 10.1068/p070695

West, S. G., and Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychol. Methods* 15, 18–37. doi: 10.1037/a0015917

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *J. Am. Stat. Assoc.* 22, 209–212. doi: 10.1080/01621459.1927.10502953

Wright, P. (1978). Feeding the information eaters: Suggestions for integrating pure and applied research on language comprehension. *Instr. Sci.* 7, 249–312. doi: 10.1007/BF00120935

Wright, P., and Wilcox, P. (1978). "Following instructions: An exploratory trisection of imperatives," in *Studies in the perception of language*, eds W. J. M. Levelt and G. B. F. d'Arcais (New York, NY: John Wiley & Sons), 129–153.

Wu, C. F. J., and Hamada, M. (2000). *Experiments: Planning, analysis, and parameter design optimization*. New York, NY: John Wiley & Sons.

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 07 December 2022
DOI 10.3389/fnins.2022.921058

Study on the diversity of mental states and neuroplasticity of the brain during human-machine interaction

Teng Zhang[1], Xiaodong Zhang[1,2]*, Wenjing Zhu[1], Zhufeng Lu[1], Yu Wang[1] and Yingjie Zhang[1]

[1]School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China, [2]Shaanxi Key Laboratory of Intelligent Robot, Xi'an Jiaotong University, Xi'an, China

Introduction: With the increasing demand for human-machine collaboration systems, more and more attention has been paid to the influence of human factors on the performance and security of the entire system. Especially in high-risk, high-precision, and difficult special tasks (such as space station maintenance tasks, anti-terrorist EOD tasks, surgical robot teleoperation tasks, etc.), there are higher requirements for the operator's perception and cognitive level. However, as the human brain is a complex and open giant system, the perception ability and cognitive level of the human are dynamically variable, so that it will seriously affect the performance and security of the whole system.

Methods: The method proposed in this paper innovatively explained this phenomenon from two dimensions of brain space and time and attributed the dynamic changes of perception, cognitive level, and operational skills to the mental state diversity and the brain neuroplasticity. In terms of the mental state diversity, the mental states evoked paradigm and the functional brain network analysis method during work were proposed. In terms of neuroplasticity, the cognitive training intervention paradigm and the functional brain network analysis method were proposed. Twenty-six subjects participated in the mental state evoked experiment and the cognitive training intervention experiment.

Results: The results showed that (1) the mental state of the subjects during work had the characteristics of dynamic change, and due to the influence of stimulus conditions and task patterns, the mental state showed diversity. There were significant differences between functional brain networks in different mental states, the information processing efficiency and the mechanism of brain area response had changed significantly. (2) The small-world attributes of the functional brain network of the subjects before and after the cognitive training experiment were significantly different. The brain had adjusted the distribution of information flow and resources, reducing costs and increasing efficiency as a whole. It was demonstrated that the global topology of the cortical connectivity network was reconfigured and neuroplasticity was altered through cognitive training intervention.

OPEN ACCESS

EDITED BY
Hong Xu,
Nanyang Technological University,
Singapore

REVIEWED BY
Alexandre Marois,
Thales Research and Technology
Canada, Canada
Rui Li,
Xi'an University of Technology, China

*CORRESPONDENCE
Xiaodong Zhang
xdzhang@mail.xjtu.edu.cn

SPECIALTY SECTION
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

RECEIVED 15 April 2022
ACCEPTED 22 November 2022
PUBLISHED 07 December 2022

CITATION
Zhang T, Zhang X, Zhu W, Lu Z,
Wang Y and Zhang Y (2022) Study on
the diversity of mental states
and neuroplasticity of the brain during
human-machine interaction.
Front. Neurosci. 16:921058.
doi: 10.3389/fnins.2022.921058

Discussion: In summary, this paper revealed that mental state and neuroplasticity could change the information processing efficiency and the response mechanism of brain area, thus causing the change of perception, cognitive level and operational skills, which provided a theoretical basis for studying the relationship between neural information processing and behavior.

KEYWORDS

human-machine collaboration, functional brain network, EEG, mental state diversity, neuroplasticity

Introduction

In the actual human-machine collaboration environment or daily life when using computers and various other large, medium and small machines (including the operation of aircraft, cars, trains, and boats), human perception, cognitive level and operational skills is in a dynamic change in real-time, which will seriously affect the performance and safety of the system (Zhang et al., 2022). Studies have shown that (1) the high mental workload of the human brain will cause rapid fatigue, reduced flexibility, stress response, increased human error, and frustration, resulting in errors in information acquisition and analysis and decision-making errors, and then lead to decreased performance. Therefore, it is a significant cause of human accidents. However, the too low mental workload will cause waste of human resources and other resources, cause disgust, and also lead to the decline of operational performance (Wilson, 2005). (2) Fatigue causes a decrease in individual feeling, perception and reaction ability, slow thinking, memory loss, inattention, reduced alertness, insensitivity to external environmental stimuli, and insufficient work motivation (Boksem et al., 2005, 2006; Zhang et al., 2019). In addition, fatigue reduces the brain's ability to process automatic and conflicting information, resulting in reduced cognitive performance (Yang, 2009). Fatigue can weaken the executive function of the frontal lobe of the brain and decrease involuntary attention and selective attention (Boksem and Tops, 2008). Fatigued people are also less likely to correct behavioral errors and show a reduced ability to monitor and regulate behavior (Lorist et al., 2005). In a complex human-machine system, most people and systems are in a special and extreme working environment, coupled with high-intensity workloads, which can easily lead to mental fatigue of operators, resulting in reduced performance and increased errors (Ricci et al., 2007). (3) Emotions affect people's perception and cognitive level and affect the processes of attention, memory, and reaction (Reeves et al., 1991; Stanley and Larsen, 2021), thereby determining the ability of situational awareness (John and Gross, 2004). In a positive emotional state, people tend to pay more attention to the overall situation of things and can see or remember the main outline of things. In contrast, people are more likely to remember the details of things in a negative emotional state. Moreover, all kinds of emotional experiences can play a role in our decision-making process (Peters et al., 2006). For example, positive emotions play a benign role in the decision-making process of switching between automated driving and manual driving (Du et al., 2020). (4) As vigilance decreases, operational performance decreases (Langner and Eickhoff, 2013). Closely related to vigilance, attention refers to the ability to focus cognitive resources on a particular stimulus (Kivikangas et al., 2011, Hancock, 2013). Insufficient attention level will lead to difficulty in completing tasks, while too great attention or too narrow attention range will affect the progress of subtasks (Frey et al., 2013). The above mental workload, fatigue, emotion, and vigilance are all classified as mental states in this paper. The topology of the functional brain networks underlying each mental state varies significantly in spatial dimensions and is diverse. We call this characteristic of the human brain in the spatial dimension the mental state diversity. which is one of the main reasons for real-time changes in human perception, cognitive level and operational skills.

Moreover, under the same mental state, compared with novices, experienced drivers or operators of special machinery and equipment have higher operational proficiency, which can effectively improve operational performance and reduce the occurrence rate of dangerous accidents caused by human error. The reason is repeated cognitive training intervention can cause changes in brain activity and even long-term changes in brain structure, which is manifested in the improvement of operational skills (Taya et al., 2015). For example, the study found that London taxi drivers with more spatial navigation experience in a complex city have larger gray matter volumes in the hippocampus (Maguire et al., 2006). Professional typists who focus on long-term typing practice have increased gray matter volume in brain regions associated with motor tasks, such as the supplementary motor area, prefrontal cortex, and cerebellum (Cannonieri et al., 2007). Since violinists and other string players use the second to fifth fingers of the left hand to play the strings, this results in a greater representation of the left hand fingers in the primary somatosensory cortex (Elbert et al., 1995). Experiments have also shown that after cognitive training interventions, the connectivity, activation, and reorganization properties of brain regions responsible for corresponding functions can be altered, such as athletes who undergo regular training can not only change their brains at the structural level, but also change the brain processing and activation patterns in the context of sports (Seidel-Marzi and Ragert, 2020). Rehabilitation and intensive training intervention can improve the motor ability of people with related functional disabilities, which is inherently manifested as changes in the neural pattern of activation and reorganization of the ipsilateral or contralateral hemisphere of the brain (Feitosa et al., 2022). We call the changes in brain neural activity and morphology caused by repeated and regular reinforcement training in the time dimension as neuroplasticity, that is, the characteristics of the brain in the time dimension. This is also the second main reason for the changes in human perception, cognitive level and operational skills.

To sum up, two main reasons for the dynamic changes of perception, cognitive level and operational skills in the actual operating environment are the mental state diversity and neuroplasticity which represent two significant properties of the brain in spatial dimension and time dimension, respectively. More specifically, the brain is a complex network consisting of spatially distributed regions dedicated to different functions, and it is proposed that mental states functions emerge from dynamic interactions of several brain areas, not from activation of a single brain region. Because in different mental states, the topological structure of the functional connection network of the human brain shows different forms in space, so the diversity of mental states is the attribute of the human brain in the spatial dimension. Moreover, in the process of evolution, development, and remodeling of living organisms, the strength of synaptic connections between neurons, internal activation of neurons, physical structure, and other aspects are shaped by the constantly changing internal and external environment all the time (Wang, 2020). Neuroplasticity changes can also be triggered in the adult brain through two fundamental processes, learning and cognitive training, although changes in brain structure are thought to be limited to critical periods of development (Draganski and May, 2008). Whether the period is years, months, days, hours, or minutes, such neuroplasticity changes over time are attributes of the brain on the time dimension. In summary, mental state diversity and neuroplasticity can be described as two major properties of the brain in spatial and temporal dimensions, respectively. However, few people have comprehensively studied the dynamic changes in the spatial and temporal dimensions of the human brain during work.

Moreover, with the emergence and development of non-invasive brain function monitoring technologies, such as electroencephalogram (EEG) (Li et al., 2022), magnetoencephalography (MEG) (Baillet, 2017), functional magnetic resonance imaging (fMRI) (Power et al., 2017), the interrelationship between human perception, cognition and performance, systems, and technology can be studied from many perspectives. In terms of mental state, for example, Li et al. (2019) designed a mental arithmetic task to induce mental fatigue in the subjects. Significant differences were found between EEG-based functional brain networks before and after the task, with marked changes in their small-world properties. Ghassemzadeh et al. (2019) proposed that the regulation and improvement of attention can be observed through changes in brain networks. Liu et al. (2019) and Wu et al. (2022), respectively proposed that the use of EEG-based functional brain network features has advantages in recognizing emotions. And Liu et al. (2019) found that the spatiotemporal topology of dynamic functional connectivity shows small-world structure. In terms of neuroplasticity, for example, Romero et al. (2008) trained subjects by designing an alphabet addition task, found that EEG of subjects before and after training was significantly different, and proved that EEG can be used as an electrophysiological marker of skill-related neuroplasticity. Seppanen et al. (2012) also observed neuroplasticity using EEG by designing a music training task. The feature of the study was to demonstrate that short-term (within tens of minutes) training tasks also triggered neuroplasticity in the subjects. In conclusion, EEG-based brain network features can effectively reflect different mental states, and EEG can also be used as an electrophysiological marker for judging neuroplasticity (Thibaut et al., 2017). Human-centered cognitive state monitoring, cognitive enhancement, closed-loop adaptive human-machine interaction and other technologies have become research hotspots. Parasuraman (2011) defined this discipline mainly focused on studying brain and behavior in work as neuroergonomics. Compared with traditional ergonomics, modern ergonomics faces increased system complexity, and the relationship between people and systems is nonlinear and fuzzy. Humans' ability to understand and simulate complex human-system interactions at work depends on their knowledge of the complexity of neural information processing, rather than solely on measures of workers' explicit behavior and subjective perceptions. Therefore, this paper uses EEG and functional brain network techniques to demonstrate that changes in mental state and neuroplasticity occur during human-machine collaboration by designing a typical mental state evoked paradigm and cognitive training intervention paradigm. The functional brain network topology formed by the interaction of various functional areas in the spatial and temporal dimensions of the brain is analyzed, which may help people to further study the information processing mechanism and mental expression mechanism in the brain. The reasons for the mental states diversity and neuroplasticity changes induced during work process are explored to provide theoretical support for behavioral approaches to improving performance. Further, by analyzing the neurobiological mechanism behind it, it can promote the improvement of neural enhancement technology, adaptive automation technology and enhanced cognition technology, which is of great significance for the development of brain enhancement systems for the field of human-machine interaction.

This article is organized as follows: in section "Materials and methods", the typical mental state evoked paradigm during work and the cognitive training intervention paradigm are designed, and the functional brain network analysis method is developed. Then the participant and experimental materials are described. In section "Results," (1) the differences between the functional brain networks of typical mental states, the transmission efficiency and the response mechanism of the related brain regions are analyzed, and the characteristics of the dynamic changes of the mental state of the human brain during work are revealed. (2) The differences in the small-world topological properties of the user's brain network before and after cognitive training caused by neuroplasticity and the neurobiological mechanisms behind it are analyzed. Section "Discussions" discusses the results, their implications, limitations, and future research efforts. Finally, section "Conclusion" summarizes the main conclusions of this paper.

Materials and methods

Analysis method for mental state diversity

Description of mental state evoked system

The proposed mental state evoked system mainly included the EEG acquisition module, the computer, and the operation module. The EEG acquisition module was mainly responsible for acquiring, amplifying, and transmitting EEG to the computer. The EEG-W32 model equipment produced by Neuracle Technology Co., LTD. was used, the sampling frequency was 1000 Hz, and the communication method was WiFi. This device consisted of 30 measuring electrodes, one reference electrode (REF) and one ground electrode (GND). The impedance level of all measuring electrodes were kept below 10 kΩ in each experiment. Electrode distribution conformed to international 10–20 standards (**Figure 1**). The computer module consisted of two computers. Computer I was used to record and analyzed EEG, and computer II was used to run virtual tasks. Microprocessor with Intel (R) Core (TM) i7-10710 CPU and i5-4590 CPU were employed in the computer I and computer II, respectively. This module used MATLAB and Robot Operating System melodic. The operation module was used to realize the human-machine interaction function, and it mainly included the joystick. The F710 wireless joystick developed by Logitech (China) Technology Co., LTD. was used, and it adopted 2.4 GHz wireless technology. The overview of the system is illustrated in **Figure 2**, when the operator played games through the joystick, the EEG was collected and transmitted to the computer in real-time.

Subjects and mental state evoked experimental procedure

Twenty-six healthy subjects participated in the experiment (the age range was 20–35, seven female and 19 male subjects), of which 13 participants participated in the mental state evoked experiment, and the remaining 13 participants participated in the cognitive training intervention experiment. All participants reported normal or corrected-to-normal vision and had no previous experience with the mental state evoked and cognitive training interventions systems. Written informed consent was obtained from each participant before the experiment. The Institutional Review Board of Xi'an Jiaotong University approved the proposed experiment, and all experiments were conducted following the Declaration of Helsinki.

In this experiment, subjects were required to operate a virtual robot to perform virtual manipulation tasks. By setting five different operation tasks and adding corresponding stimulation conditions during the operation, various typical mental states (there are mainly resting state, fatigue state, attentive state, inattentive state, positive state, and negative state) were induced. The EEG acquisition module collected EEG in real time and made a marking process. The mental state label was obtained according to the subjective evaluation of the subjects. Among them, the interface of five kinds of operation tasks is shown in **Figure 3A**, and the specific requirements of each task are as follows: in task 0, the subjects need to control the robot to move in a clockwise circle in a rectangular room, without restricting the robot's movement trajectory, as shown by the dotted line in the figure. In task 1, the subjects need to control the robot to walk along the prescribed trajectory, namely the sides a and c of the triangle, as shown by the thin solid lines in the figure. In task 2, the subjects were required to control the robot to grab the small cube on the table, and then place the cube in the white tray. Task 3 is a time-limited task in which the subjects need to control the robot to walk out of the maze within 1 min. In task 4, compared to task 2, the small cube was replaced by a cylinder bottle, and the tray was placed on a table in another room. The subjects were asked to control the robot to grab the cylindrical bottle and move it to another room, placing the cylindrical bottle in a white tray. The difficulty level of task 0 to task 4 is from 1 to 5, the higher the number, the higher the difficulty. The stimulation conditions in the experiment mainly include (1) the operation interface is blocked or the background is blurred, so that the subjects can improve their attention. (2) Repeatedly perform a single task for a long time, so that the subjects lose concentration. (3) Performing complex tasks with high mental workload for a long time, thus making the subjects mentally fatigued. (4) Prompts for correct operation and reward prompts appear, thus inducing a positive state of the subjects.

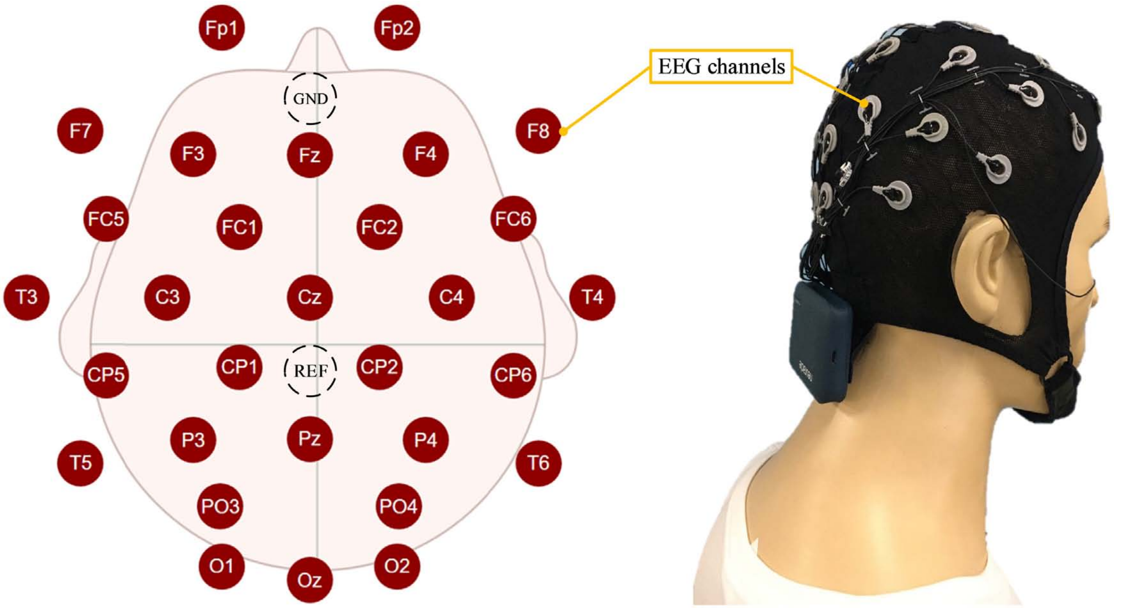


FIGURE 1
The distribution of electroencephalogram (EEG) electrodes and how the participant wore them.

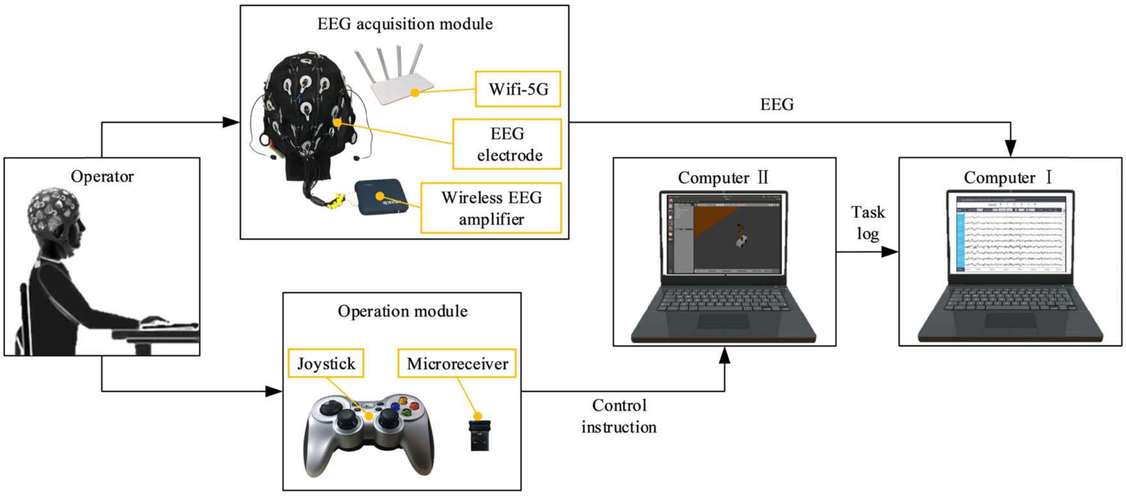


FIGURE 2
An overview of the mental state evoked system.

(5) Prompts of operation errors and task countdowns appear, thereby inducing a negative state of the subjects.

The correspondence between each task and the mental state expected to be evoked is as follows: Task 0 had the lowest difficulty level and had no time requirement, so it was used to induce the subject's inattentive state at the end of the task; The difficulty level of Task 1 was slightly higher than that of Task 0, so it was used to induce the attentive state of the subjects; The difficulty level of Task 2 was medium (level 3). According to the real-time performance of the subjects when performing the task, with the stimulation conditions preset in the experiment, it could be used to induce positive and negative states of the subjects; Task 3 set the condition of time limit, which could increase the subject's concentration or sense of urgency, so it was used to induce the subject's attentive or tension state; Task 4 had the highest difficulty level and could induce fatigue in the subjects by performing difficult tasks for a long time.

The experimental steps are as follows: firstly, the subjects sit quietly in front of the computer screen wearing an EEG cap, so that their hands can comfortably control the handle,

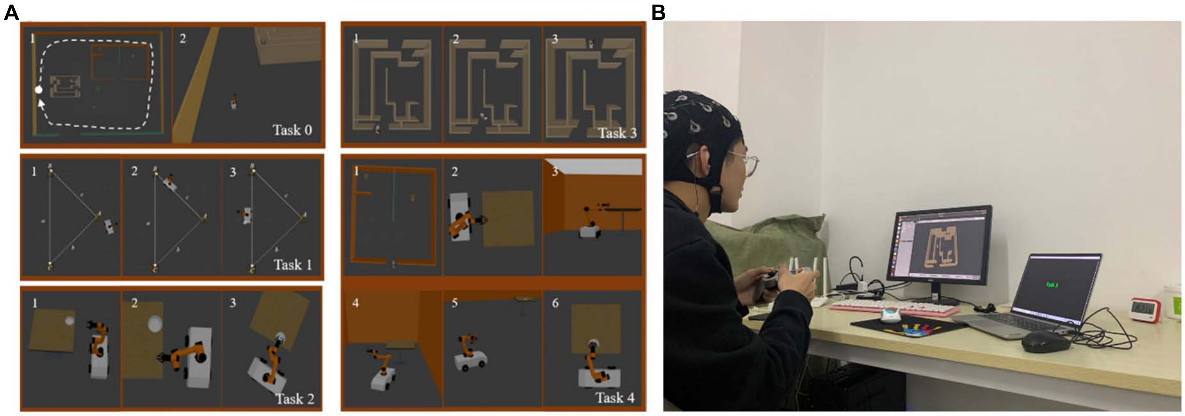


FIGURE 3
(A) The task interface of a typical mental state evoked system. **(B)** The experimental scenario.

and the experimental scene is shown in **Figure 3B**. Before the experiment, the subjects were asked to be familiar with the experimental requirements and the experimental procedure. Then, the experiment officially started, and each subject was asked to perform two rounds of the task. The execution sequence of each round is Task 1 → Task 2 → Task 3 → Task 0 → Task 4. Stimulation conditions were preset, and when a specific task was performed, a specific stimulus condition would appear, in order to induced a related mental state. When the subjects performed the task, the EEG acquisition module would collect EEG in real time and do marking processing. Each time the EEG data was recorded, the subjects were required to complete a subjective evaluation. The total experimental duration of each subject was 2 h.

In order to make the subjects complete the subjective evaluation simply, intuitively and efficiently, we designed a subjective evaluation scale for the operator's mental state based on the SAM scale (Balasubramanian et al., 2018) and the NASA-TLX scale (Harry et al., 2021; **Figure 4**). It included six evaluation indicators, and the degree of each evaluation indicator was divided into five grades, and the combination of multiple evaluation indicators was used to judge the type of mental state. The corresponding relationship between the index parameters in the subjective evaluation scale and the typical mental state is shown in **Table 1**, where "/" stands for unlimited.

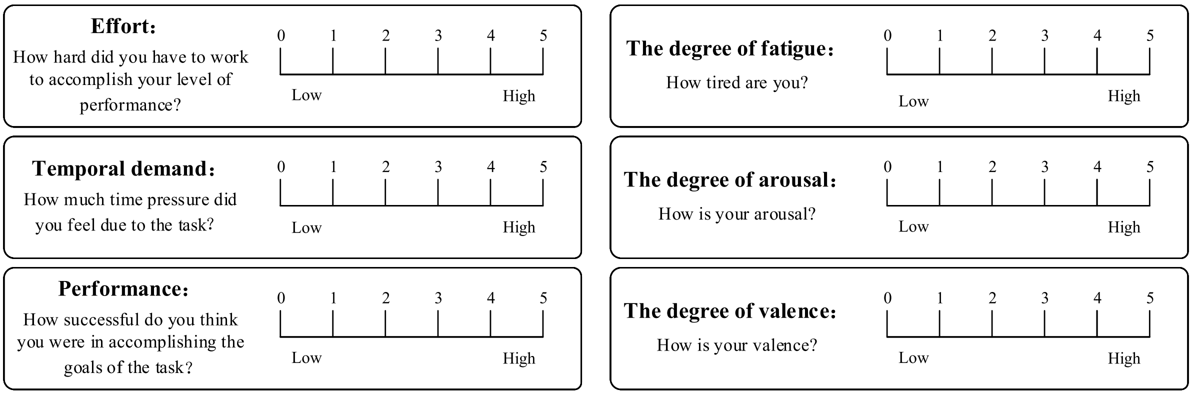
Data processing

The data processing mainly includes four steps. Step 1: filter noise, remove baseline and ocular electrical signals from the collected EEG, detailed steps can be found in our previous study (Zhang et al., 2021). Step 2: on the whole-brain scale, the location of the 30 channels in the EEG acquisition device is used as the node of the brain network. Then, based on the EEG data, calculate the phase lag index (*PLI*) between each channel at low frequency band (that is, the frequency bands of theta and alpha rhythm waves are 4.0∼13.0 Hz) and high frequency band (that is, the frequency band of the beta rhythm wave is 14.0∼30.0 Hz), respectively, since the delta rhythm wave appears only during sleep, deep anesthesia, hypoxia and organic brain lesions, this rhythm wave is not considered in this paper. Step 3: according to the size of the *PLI*, the functional connections between nodes in the brain network are measured, and a series of threshold correlation matrices are generated by setting parameters such as connection thresholds to describe the functional brain network. Step 4: the small-world network attribute parameters of functional brain networks are calculated according to graph theory, and the statistical differences between the indicators of various typical mental states are analyzed by using the Two-sample *T*-test method (Xu et al., 2017). The following is a detailed introduction.

The brain is an extremely complex network, which is interrelated on different temporal and spatial scales. By studying the brain's neurons and the connections between them, we can understand the coordination between brain regions and the functional cognitive principles of the brain. A large number of research results have proved that the brain is neither a completely random network nor a completely regular network, but an "economic" small-world topology network (Sun et al., 2014). The so-called small-word network refers to that it has a small characteristic path length L and a large clustering coefficient C. In general, the clustering coefficient C_i of a node i with degree K_i is defined as the ratio of the number of existing edges ($\#_i$) between neighbors of i (a node j is called a neighbor if $A_{ij} = 1$) and the maximum possible number of edges. This can be formalized as (Ponten et al., 2010),

$$C_i = \frac{2\#_i}{K_i\left(K_i - 1\right)} = \frac{1}{K_i\left(K_i - 1\right)} \sum_{j=1}^{N} \sum_{o=1}^{N} A_{ij} A_{io} A_{oj} \qquad (1)$$

where C_i is an index of local structure, which has been interpreted as a measure of resilience to random error (in case



Effort: How hard did you have to work to accomplish your level of performance?

Temporal demand: How much time pressure did you feel due to the task?

Performance: How successful do you think you were in accomplishing the goals of the task?

The degree of fatigue: How tired are you?

The degree of arousal: How is your arousal?

The degree of valence: How is your valence?

FIGURE 4
Subjective evaluation scale for typical mental states.

TABLE 1 Correspondence between the indicators parameters and typical mental states in the subjective evaluation scale.

Indicators of subjective evaluation scale						Typical mental state
Effort	**Temporal demand**	**Performance**	**The degree of fatigue**	**Arousal**	**Valence**	
/	Medium	/	/	≥Medium	Medium	Attentive state
Low	Low	/	/	Low	Low	Inattentive state
/	/	/	High	Low	Medium	Fatigue state
/	/	High	/	≥Medium	High	Positive state
/	/	Low	/	≥Medium	Low	Negative state

node *i* is lost its neighbors remain connected if C_i is large). C is the mean clustering coefficient of the graph. That is,

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{2}$$

The path length L_{ij} between two nodes *i* and *j* is the minimal number of edges that have to be passed to connect *i* and *j*. The mean shortest path length L of a graph is the mean L_{ij} between all possible pairs of nodes. We note that using the harmonic mean allows for inclusion of isolated nodes (for which $L_{ij} \to \infty$). That is,

$$L = \frac{1}{N(N-1)}\sum_{i \neq j} L_{ij} \tag{3}$$

The small-world coefficient σ is an indicator that indicates whether a network has the small-world topological attribute. σ can be calculated according to C and L. That is,

$$\left.\begin{array}{l} \gamma = C/C_{\text{random}} \\ \lambda = L/L_{\text{random}} \\ \sigma = \gamma/\lambda \end{array}\right\} \tag{4}$$

C_{random} and L_{random} are the clustering coefficient and characteristic path length of the random network, respectively. If $\sigma > 1$, it means that the network has the small-world topology property, otherwise it does not have the small-world topology property. Another commonly used measure is network global efficiency E_g, which is the average of the reciprocal of L_{ij}. Both indicators E_g and L_{ij} can better measure the global information processing and transmission capabilities of the network, as well as the degree of integration of the network. This can be formalized as,

$$E_{\text{g}} = \frac{1}{N(N-1)}\sum_{i \neq j}\frac{1}{L_{ij}} \tag{5}$$

In addition, network local efficiency E_{loc} is the average of all node local efficiencies NE_{loc}, which can measure the degree of differentiation of the network, so that it can effectively characterize the local characteristics of the brain network. The parameters introduced above are global attribute parameters. This paper also uses a local attribute parameter called NE_{loc}, which characterizes the efficiency of parallel information transmission of the node in the network. This can be formalized as,

$$NE_{\text{loc}}(i) = \frac{1}{N_{G_i}(N_{G_i}-1)}\sum_{j \neq k}\frac{1}{L_{jk}} \tag{6}$$

where G_i refers to the subgraph formed by the adjacent nodes of node *i*, and L_{jk} represents the length of the shortest path between nodes *j* and *k*.

When we construct a functional brain network, the definition of the edges between nodes adopts the *PLI*, which can better avoid the effect of volume conductors. It is an index to detect the asymmetry of phase difference distribution between two signals, it can reflect the consistency of phase advance or lag of one signal relative to another, and it is an effective estimate of phase synchronization. The biggest advantage of *PLI* is that it is insensitive to the volume conductor effect of signals (Winter et al., 2007) and can only focus on the coupling relationship between signals (Stam et al., 2009). The specific calculation process is as follows, assuming that φ_n and φ_m are the phases of two time series and $\Delta\varphi$ is the phase difference between them, then the phase synchronization index between time series *n* and time series *m* (*n*, *m* is an integer) is defined as the following formula,

$$\left|\Delta\varphi_{nm}^{PQ}\right| = |P\varphi_n - Q\varphi_m| < \text{constant} \tag{7}$$

we qualified $P = Q = 1$. To calculate phase synchronization, we need to know the instantaneous phase of the two signals. It can be obtained by the Hilbert transformation operation of the analytic signal $\psi(t)$. The $\psi(t)$ can be obtained from a real time series $S(t)$ and its Hilbert transform $\tilde{S}(t)$, as shown in formula,

$$\left.\begin{aligned}
&\psi(t) = S(t) + i\tilde{S}(t) = A(t)e^{i\varphi(t)} \\
&\tilde{S}(t) = \pi^{-1}\int_{-\infty}^{\infty}\frac{S(\tau)}{t-\tau}d\tau \\
&A(t) = \sqrt{S(t)^2 + \tilde{S}(t)^2} \\
&\varphi(t) = \arctan\frac{\tilde{S}(t)}{S(t)}
\end{aligned}\right\} \tag{8}$$

where $A(t)$ represents the instantaneous amplitude and $\varphi(t)$ represents the instantaneous phase. *PLI* estimates the insensitive phase synchronization from the same source by calculating the asymmetry of the phase difference distribution law, as shown in formula,

$$PLI = \left|\left\langle sign\left(\Delta\varphi(t)\right)\right\rangle\right| = \left|\frac{1}{N}\sum_{n=1}^{N} sign\left(\Delta\varphi\left(t_n\right)\right)\right| \tag{9}$$

where *sign* is a sign function. The variation range of *PLI* is 0–1. When *PLI* is 0, it indicates that there is no coupling relationship, and when *PLI* is 1, it indicates that there is a perfect phase locking relationship.

Analysis method for neuroplasticity of the brain

Description of cognitive training interventions system

Similar to the mental state evoked system mentioned above, cognitive training interventions system mainly includes three modules: the EEG acquisition module, the computer, and the operation module. The difference is that the operation module is composed of mouse and keyboard, because the virtual task has changed.

Cognitive training interventions experimental procedure

Studies have found that one of the best approach is to use video games to restore the brain or improve functional plasticity (Bavelier and Green, 2003; Ballesteros et al., 2015; Valentin, 2017). Because of its ease of use and the numerous potential applications, the cognitive training has attracted substantial public attention, and a lot of computer software for "brain training" are available on web, PCs or smartphones. Hence, two kinds of game were designed in this paper. One was Breakout game and the other was Snake game, which were tweaked from the original Atari games (Bevilacqua et al., 2016; Cuccu et al., 2021). **Figure 5A** shows that the operator can use the mouse to control the racket to hit the ball and make it hit bricks. The more bricks it hits, the higher the score; on the contrary, the fewer bricks it hits or the game fails, the lower the score. The way to calculate the score is as follows: 1 point for each brick hit, and the racket successfully hits the ball five times in a row, the speed of the ball will increase by one level, at this time, 2 points for each hit of a brick. By analogy, the higher the speed rating, the higher the score for each brick hit. The scoring stops when the game is cleared or failed. **Figure 5B** shows that the operator can control snake's direction by using the arrow keys on the keyboard, so that snake can eat the square. The more the snake eats, the higher the score; on the contrary, the less the snake eats or game fails, the lower the score. The way to calculate the score is as follows: 1 point for each square eaten, and if the snake eats two squares in a row, the speed of the snake will increase by one level, at this time, 2 points will be awarded for each square eaten. By analogy, the higher the speed level, the higher the score for eating a square. The scoring stops when the game is cleared or failed. The computer records the game task score for each session to evaluate the quality of the control. The game task score can be used to comprehensively evaluate the operation accuracy and reaction time in the behavioral data. If the subject's operation accuracy is poor, it is easy to cause the game to fail, resulting in a low score. In addition, if the subject's reaction time is slow, when the speed level increases, the game will also fail and the score will be lower. Conversely, if both the operation accuracy and reaction time have high levels, the game score will be high. The whole experiment process was mainly divided into three stages: control experiment stage, training experiment stage and test experiment stage. The control experiment stage was set before the training experiment stage, while the test experiment stage was set after the training experiment stage, in order to compare the difference between the control experiment and the test experiment results. In the control experiment stage, participants were required to wear an EEG cap, sit quietly in front of the computer screen, and allow their hands to comfortably control the mouse or keyboard, and perform game tasks for 5 min, then the score of the game task was recorded, and the EEG were collected during the 5 min.

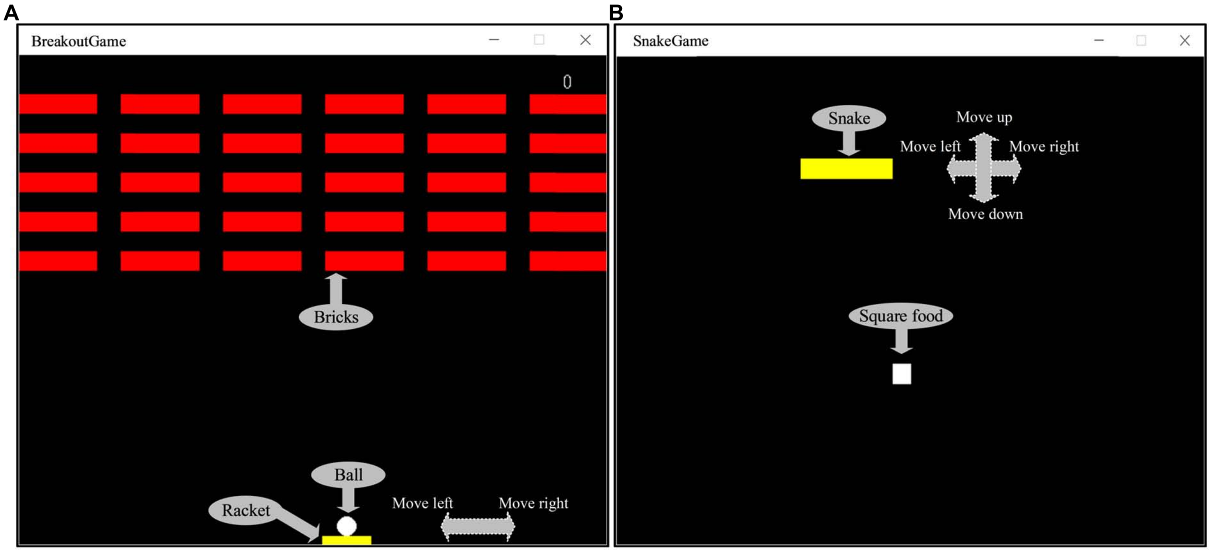


FIGURE 5
(A) The Breakout game interface. **(B)** The Snake game interface.

Participants can choose to rest for 30 s between each game session. Similarly, the same setup was carried out in the test experiment phase. In the training experiment stage, the subjects were required to perform game tasks for about 30 min. If the game was cleared, the experiment could be ended in advance. This stage was the cognitive intervention training stage. Before the control experiment, 30 s of experiment familiarity was set up. In order to control the variables in the control and test experiment stage, we had made efforts in three aspects: (1) In terms of EEG acquisition equipment, the secondary wearing of the EEG cap during the experiment was avoided. Since the re-wearing of the EEG cap will affect the collected signals, we controlled the duration of the experiment within a reasonable range to avoid the situation of wearing the EEG cap twice during the experiment, and ensure that the electrode position and resistance value of the EEG cap remain unchanged. (2) In terms of the subject's mental state, the different mental states of the subjects in the two experimental stages were avoided. The duration of the training experiment stage was no more than 30 min, the task difficulty was moderate, and a rest session was set to avoid the subjects from mental fatigue to the greatest extent. And in the two experimental stages, a subjective questionnaire survey was conducted on the mental state of the subjects, and if there was a difference in the mental state, the data was excluded. (3) The game task parameters of the two experimental stages were set the same, so that the subjects performed the same game tasks in the control and test experimental stages. Figure 6 shows the experimental scenario and the experimental steps. Other requirements are the same as described above, and the data processing steps are also the same as described above.

Results

Result for mental state diversity

In order to analyze the differences of functional brain networks in various typical mental states and the neurobiological mechanisms behind them, three pairs of typical mental states were selected for pairwise comparative analysis: resting state and fatigued state, attentive state and inattentive state, and positive state and negative state. We calculated the *PLI* matrix of each mental state induced by all subjects during the task and calculated its mean value. Then the functional brain network diagram and connection matrix diagram of each typical mental state were drawn. Firstly, the differences among the three pairs of typical mental states are qualitatively analyzed.

Resting state and fatigued state

Figures 7A,C show that the functional brain network connectivity density of low-frequency EEG increased abruptly when the subjects moved from the resting state to the fatigued state. One of the reasons for this phenomenon may be that when adults are fatigued, the slow waves in the EEG gradually increase, while the fast waves gradually decrease (Borghini et al., 2014), so that the functional brain network connection density of low-frequency EEG under fatigue conditions becomes denser. However, **Figures 7B,D** show that, relative to the resting state, the connectivity density of the functional brain network for high-frequency EEG in the fatigued state does not decrease but increases, although the increase is much smaller than in the case of low-frequency EEG. The reason is that the subjects in the fatigued state are in the process

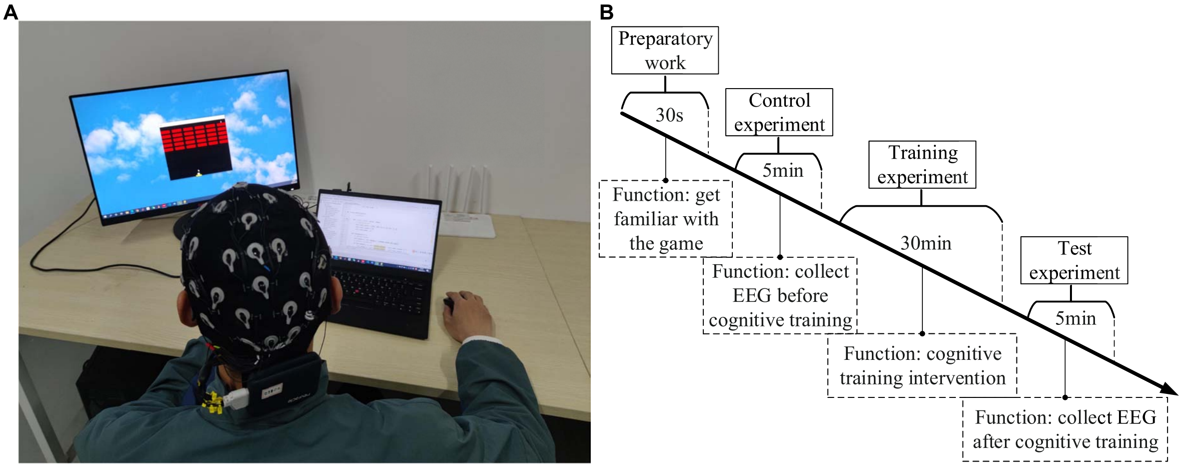


FIGURE 6
(A) The experimental scenario. **(B)** The experimental steps.

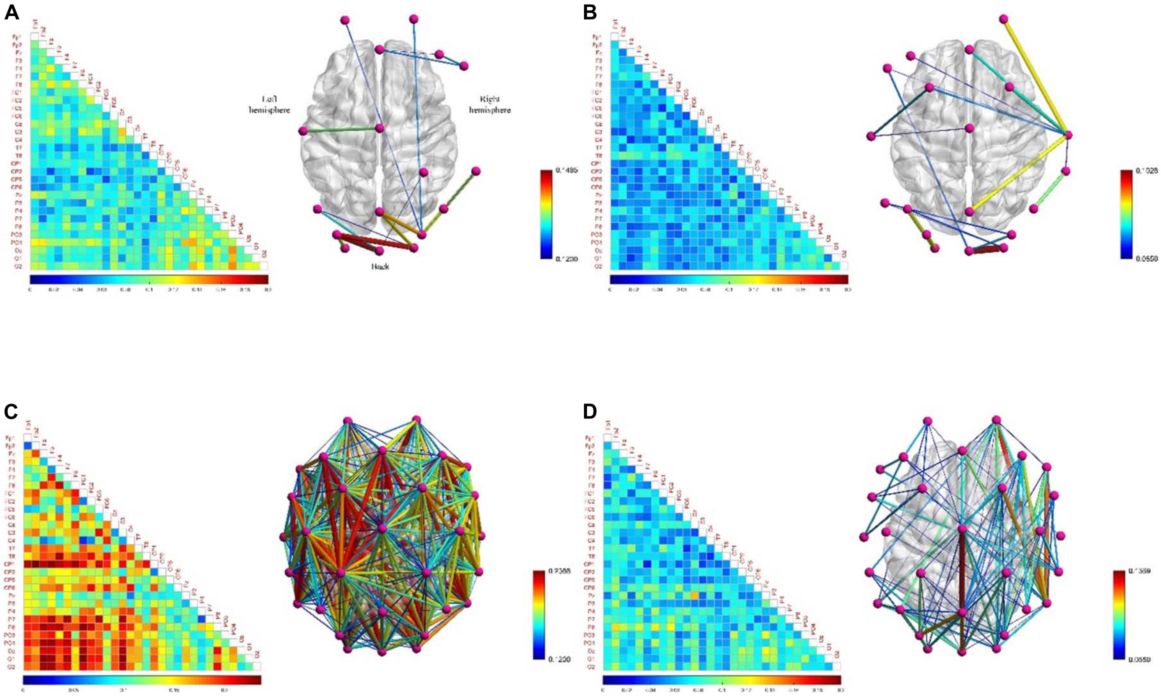


FIGURE 7
The functional brain networks and connectivity matrices in resting and fatigued states. The perspective of the brain network map is a top view, with the left hemisphere of the brain on the left, and the occipital lobe area below. The lower limit of the connection threshold is manually adjusted to better display the characteristics, and the upper limit is the highest value of the network connection strength. The matrix diagram shows the lower triangular area of 30 channels, where the channels are arranged from top to bottom in order from the front (frontal lobe) to the back (occipital lobe) in the brain area, the same below. **(A)** Represents the situation when the electroencephalogram (EEG) is at low frequency in the resting state. **(B)** Represents the situation when the EEG is at high frequency in the resting state. **(C)** Represents the situation when the EEG is at low frequency in the fatigue state. **(D)** Represents the situation when the EEG is at high frequency in the fatigue state.

of work and need to mobilize brain resources to maintain a high level of perception and cognition, so as to effectively perform game tasks. Therefore, this reflects that the functional brain network connection density of both low-frequency and high-frequency EEG increased in the fatigue state compared with the resting state during work, and the connectivity density of the brain network with low-frequency EEG was increased to a higher degree due to further influence by the fatigue state. This phenomenon shows that the fatigue state during work is significantly different from that during non-work. In addition, it is interesting that the functional brain network of high-frequency EEG has lower connectivity density in the left hemisphere than in the right hemisphere under fatigue state, showing an asymmetric connectivity pattern, as shown in **Figure 7D**. This indicates that the processing response of the brain area is biased to the right in the fatigue state, and this phenomenon has also been confirmed in Sun et al. (2014).

The quantitative analysis of resting and fatigued states was carried out below. We calculated the average value of the parameters of the small-world network whose sparsity ranged from 5 to 40% of the functional brain network of all subjects, and used the Two-Sample *T*-test method to statistically analyze the differences between each pair of parameters (the settings for the quantitative analysis below are the same). We compared σ, γ, λ, E_g, and E_{loc} five small-world network parameters. **Figure 8** shows that compared to the resting state, the σ in the fatigued state is significantly reduced (low frequency, $t = 3.873$, $p = 0.002$, effect size of Cohen's $d = 1.937$; high frequency, $t = 4.315$, $p = 0.001$, effect size of Cohen's $d = 2.158$), indicating that the "economy" of the functional brain network in the fatigued state decreases, and the brain needs to expend more resources to maintain a high level of perception and cognition. At the same time, the γ in the fatigue state is significantly reduced (low frequency, $t = 2.486$, $p = 0.026$, effect size of Cohen's $d = 1.243$; high frequency, $t = 2.919$, $p = 0.011$, effect size of Cohen's $d = 1.459$). It indicates that in the process of brain processing information flow, the information processing ability of the local brain network under fatigue decreases, leading to the reduction of cluster characteristics of functional differentiation among the called parts of the brain regions, and the information connectivity and collaborative processing between the local brain regions become weak.

Attentive state and inattentive state

Figure 9A shows that the densely connected areas of the functional brain network of the low-frequency EEG are mostly concentrated near the frontal and occipital lobes when the subjects are in attentive state. This is because the frontal lobe can enhance the excitability of a certain part of the brain area, so that the related stimuli can be noticed, while inhibiting other brain areas, so it can manage attention, concentrate and maintain a high degree of attention. At the same time, the brain regions involved in the regulation of attention also include the inferior parietal cortex, the superior temporal cortex, and the occipital lobe. This phenomenon has also confirmed in the study of Rosenberg et al. (2016). The connectivity density of functional brain networks for low-frequency EEG increased slightly when attention is decreased (**Figure 9C**). Comparing **Figures 7A,C, 9A,C**, it is found that the decrease in concentration and arousal have the same phenomenon in the functional brain network of low-frequency EEG, indicating that there is a certain correlation between attention and arousal. This corroborates the theory that decreased attention is attributed to decreased physiological arousal (Luna et al., 2021). However, the difference degree is much smaller than the difference degree between the fatigue and the resting state, which shows that there is also the difference between the decline of attention and the decline of arousal from another perspective. This may confirm the resource depletion theory proposed by Warm et al. (2008) from the side, that is, the decrease in attention is caused by the lack of replenishment of information processing resources in continuous tasks.

Besides, when the EEG was at high frequency, the degree of difference in connection density between functional brain networks in the attentive and inattentive states was not significant, and the connection patterns shared the same characteristics. For example, in both states, the connection density in the occipital lobe region is high, and the local connection patterns on the P7 channel have the same characteristics, as shown by the arrows in **Figures 9B,D**. The reason may be that the connectivity properties of the functional brain network of high-frequency EEG are mainly task-dominated, that is, regardless of whether the subject is attentive or inattentive state, when performing the same task, the brain maintains the appropriate level of perception and cognition, as well as the level of limb control and decision-making. The activation degree of the brain regions responsible for related functions and the allocation pattern of brain resources are basically consistent under the influence of the same task, thus showing the same connection characteristics on the functional brain network. This shows that compared with low-frequency EEG, high-frequency EEG contains less mental state information, which is mainly affected by the mobilization of brain resources, the activation level of brain regions and the way of information transmission during work. This phenomenon is also found between positive and negative states, as shown in **Figures 11B,D**, so the above inference can be generalized to the case of multiple mental states. This highlights another difference between mental states during work and non-work, and provides theoretical support for the detection and identification of mental states during work.

For attentive and inattentive states, we compared five global attribute parameters and one local attribute parameter of functional brain networks, namely σ, γ, λ, E_g, $E_{\text{loc}0}$, and NE_{loc}. **Figure 10** shows that when the EEG is at a low frequency, compared with the inattentive state, the σ in the attentive state is significantly increased ($t = 2.578$, $p = 0.022$, effect size of Cohen's $d = 1.289$), indicating that maintaining concentration during work improves the "economy" of the small-world attribute of the brain network. The brain optimizes the allocation of resources to maintain high levels of perception and cognition associated with work tasks, thereby making the utilization of brain resources more efficient and the allocation of resources more concentrated. At the same time, the NE_{loc}

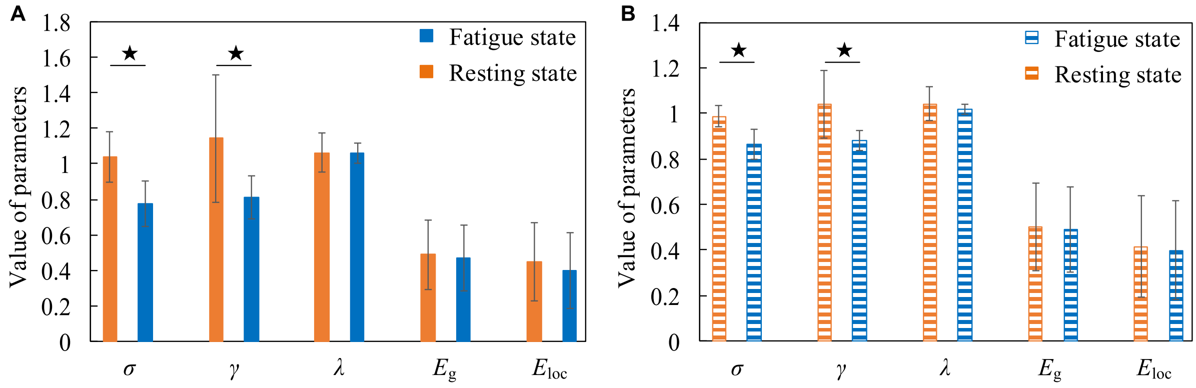


FIGURE 8

The contrasting histograms of five global attribute parameters of functional brain networks in resting and fatigued states. "★" indicates statistical difference, the same as below. **(A)** The case when the electroencephalogram (EEG) is at low frequency. **(B)** The case when the EEG is at high frequency.

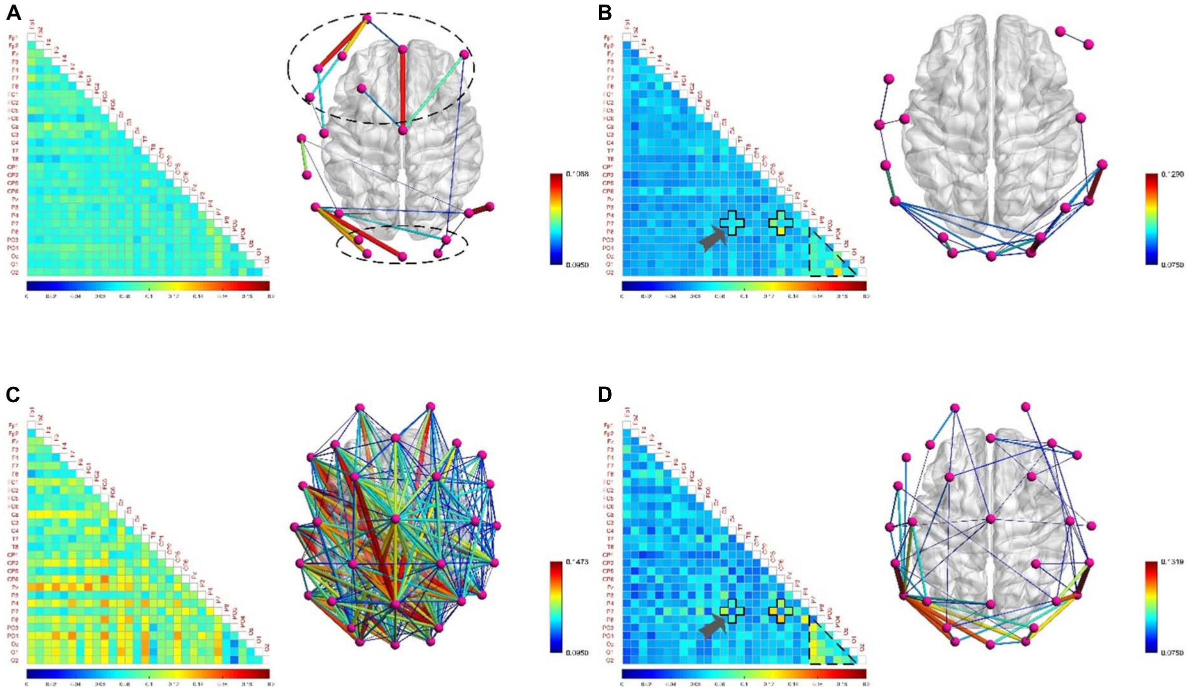


FIGURE 9

The functional brain networks and connectivity matrices in attentive and inattentive states. **(A)** The situation when the electroencephalogram (EEG) is at low frequency in the attentive state. **(B)** The situation when the EEG is at high frequency in the attentive state. **(C)** The situation when the EEG is at low frequency in the inattentive state. **(D)** The situation when the EEG is at high frequency in the inattentive state.

in the occipital lobe area also increased significantly in the attentive state ($t = 2.991$, $p = 0.017$, effect size of Cohen's $d = 1.892$), which proved that the occipital lobe area played a positive role in improving the subjects' attention during work. Besides, no significant differences were found in the functional brain network of high-frequency EEG, further supporting the speculation that the connectivity properties of the brain network of high-frequency EEG are mainly task-dominated.

Positive state and negative state

Figure 11 shows no significant difference and correlation in the connection density of functional brain networks in positive and negative states. The reason may be that the approach motivation system promotes the generation of positive states, while the avoidance motivation system promotes the generation of negative states. Since the approach motivation system and the avoidance motivation system are driven by two independent

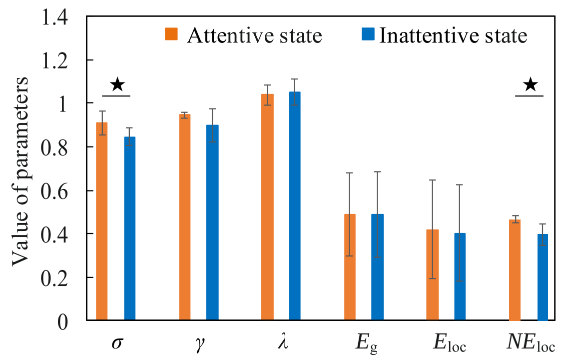


FIGURE 10
The contrast histograms of five global attribute parameters and one local attribute parameter of the functional brain network in the attentive and inattentive states.

neural circuits, and require the coordinated activities of different levels of the nervous system and different brain regions to complete, this may be one of the reasons for this phenomenon (Peng, 2019). In addition, when the brain is in a positive state, the connection density of the functional brain network of the low-frequency EEG in the right hemisphere is denser, and that in the left hemisphere is sparser, as shown in **Figure 11A**; on the contrary, when the brain is in a negative state, the connection density of the functional brain network in the left hemisphere is denser, and that in the right hemisphere is sparse, as shown in **Figure 11C**, showing the asymmetry and difference between the left and right hemispheres. The underlying reason may be that the positive state increases the activation of the left hemisphere cortex, and the negative state increases the activation of the right hemisphere cortex (Peng, 2019), resulting in the difference between the left and right hemispheres in processing positive and negative states. Moreover, the functional brain network of high-frequency EEG not only did not reveal obvious asymmetries and differences between the left and right hemispheres, but also the connectivity patterns shared the same characteristics. For example, in both states, the connection density in the occipital lobe region is high, and the local connection patterns on the P7 channel have the same characteristics, as shown by the arrows in **Figures 11B,D**, which proves the above inference.

For positive and negative states, we also compared five global attribute parameters and one local attribute parameter of functional brain networks, namely σ, γ, λ, E_g, E_{loc}, and NE_{loc}. There was no significant difference in the global properties of functional brain networks in positive and negative states when EEG was at low frequency. However, there were significant

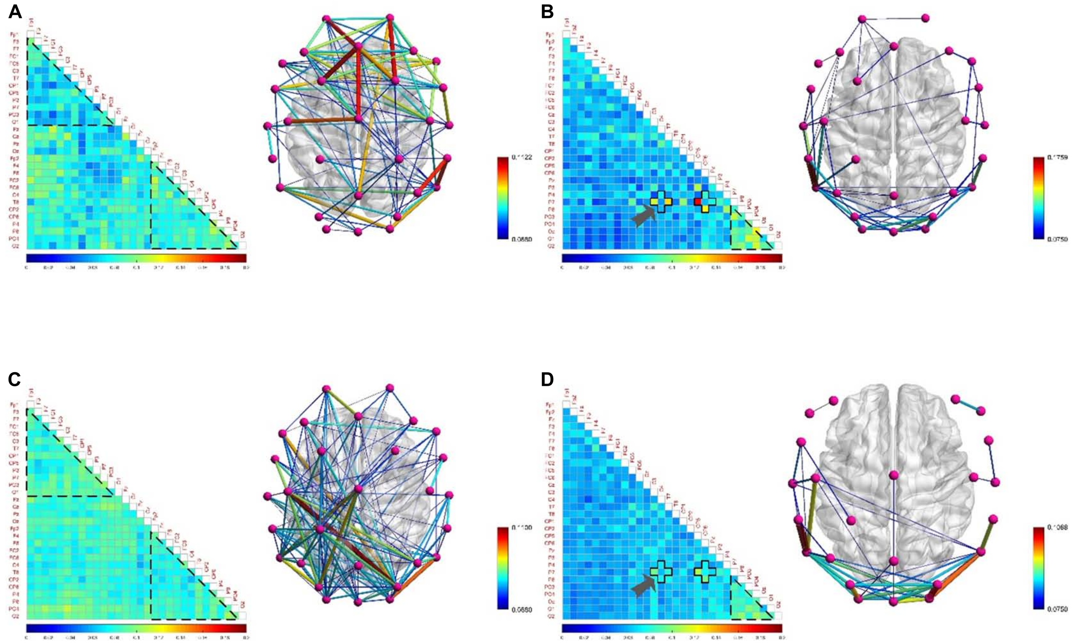


FIGURE 11
The functional brain networks and connectivity matrices in positive and negative states. **(A)** The situation when the electroencephalogram (EEG) is at low frequency in the positive state. In order to observe the difference between the left and right hemispheres from the matrix diagram, the channels in the matrix diagram are arranged in the order of left hemisphere, central axis, and right hemisphere from top to bottom, and other settings are the same as before. **(B)** The situation when the EEG is at high frequency in the positive state. **(C)** The situation when the EEG is at low frequency in the negative state. The channels in the matrix diagram are arranged in the order of left hemisphere, central axis, and right hemisphere from top to bottom, and other settings are the same as before. **(D)** The situation when the EEG is at high frequency in the negative state.

differences in NE_{loc} between the left and right hemispheres of the brain. In **Figure 12A**, the first group of histograms represent the NE_{loc} of the left and right hemispheres in the positive state, respectively. The second group represents the NE_{loc} of the left and right hemispheres in the negative state, respectively. The third group represents the NE_{loc} of the left hemisphere in positive and negative states, respectively. The last group represents the NE_{loc} of the right hemisphere in positive and negative states, respectively. In the negative state, the NE_{loc} in the frontal, central and temporal regions of the left hemisphere (channels are F3, F7, FC1, FC5, C3, and T7) are significantly higher than those in the right hemisphere (channels are F4, F8, FC2, FC6, C4, and T8), indicating that the local information transmission efficiency of the nodes in the left hemisphere was significantly increased in the negative state ($t = 2.896$, $p = 0.016$, effect size of Cohen's $d = 1.672$). At the same time, compared with the negative state, the NE_{loc} in the frontal, central and temporal regions of the right hemisphere increased significantly in the positive state ($t = 2.413$, $p = 0.036$, effect size of Cohen's $d = 1.393$), indicating that the local information transmission efficiency of the nodes in the right hemisphere was significantly increased in the positive state. It should be noted that previous studies have concluded that positive states increase the activation of the left hemisphere cortex, and negative states increase the activation of the right hemisphere cortex (**Peng, 2019**), so it can be concluded that the local efficiency of the node is not positively correlated with the activation degree of the cortex. The common conclusion that can be drawn is that the brain has left and right hemisphere differences in processing positive and negative states. **Figure 12B** shows that σ is significantly higher in the positive state than in the negative state ($t = 2.953$, $p = 0.010$, effect size of Cohen's $d = 1.477$), it shows that the positive state effectively improves the "economy" of the brain's small-world network and optimizes the allocation of resources, so that the brain can more efficiently maintain the level of perception and cognition required for work. At the same time, the γ in the positive state is significantly elevated ($t = 2.613$, $p = 0.020$, effect size of Cohen's $d = 1.306$). It indicates that in the process of brain processing information flow, the information processing ability of the local brain network under positive state increases, leading to the enhancement of cluster characteristics of functional differentiation among the called parts of the brain regions, and the information connectivity and collaborative processing between the local brain regions become stronger. This may explain why subjects have higher quality of work when they are in a positive state.

Result for neuroplasticity of the brain

Firstly, the behavioral data results of the subjects before and after cognitive intervention training were analyzed. We calculated the average value of the game task scores of all subjects before and after training, and used the Two-Sample T-test method to statistically analyze the differences in game task scores. **Figure 13** shows that in both game tasks, the game task scores obtained by subjects after training were significantly higher than those obtained before training, (Breakout game, $t = 2.501$, $p = 0.020$, effect size of Cohen's $d = 0.981$; Snake game, $t = 4.097$, $p < 0.001$, effect size of Cohen's $d = 1.607$), indicating that after the cognitive training intervention, the behavioral data results of the subjects changed, and the operational skills were improved.

In order to verify the changes of neuroplasticity after cognitive training intervention and analyze the underlying neurobiological mechanisms, we analyzed the global and local attributes of functional brain networks before and after training, and there was no significant difference in global attributes, but significant differences in local attributes, mainly on NE_{loc}. The mean values of NE_{loc} in frontal lobe, parietal lobe, temporal lobe and occipital lobe of all subjects participating in this experiment were calculated, respectively, and the difference between each pair of parameters was statistically analyzed by Two-Sample T-test method. In the Breakout game task, **Figure 14A** shows that when the EEG is at a low frequency, NE_{loc} in the frontal lobe is significantly increased after training compared to the situation before training ($t = 3.593$, $p = 0.002$, effect size of Cohen's $d = 1.532$), however, NE_{loc} in the parietal, temporal and occipital lobes are significantly decreased (parietal and temporal lobes, $t = 2.342$, $p = 0.027$, effect size of Cohen's $d = 0.885$; occipital lobe, $t = 4.413$, $p = 0.002$, effect size of Cohen's $d = 2.791$). A similar phenomenon is found in the Snake game task, with **Figure 14B** showing that NE_{loc} in the parietal, temporal, and occipital lobes are significantly reduced after training compared with the pre-training condition (parietal and temporal lobes, $t = 5.765$, $p < 0.001$, effect size of Cohen's $d = 2.179$; occipital lobe, $t = 2.637$, $p = 0.030$, effect size of Cohen's $d = 1.668$). It shows that after the subjects undergo cognitive training intervention, the efficiency of parallel information transmission in the network of nodes in the frontal lobe is improved. This may be related to the fact that the frontal lobe area is responsible for reasoning, calculation, motor control and problem solving (Ming, 2018), which proves that the training of the operation task promotes the improvement of the subjects' operation control ability and task-related auxiliary ability. Conversely, nodes in the parietal, temporal, and occipital lobes are less efficient in parallel information transfer in the network. In addition, when EEG was at high frequency, the difference between NE_{loc} before and after cognitive training was not significant, the reason may be that the connectivity properties of functional brain network of high frequency EEG are mainly task-dominated, which further confirms the above inference.

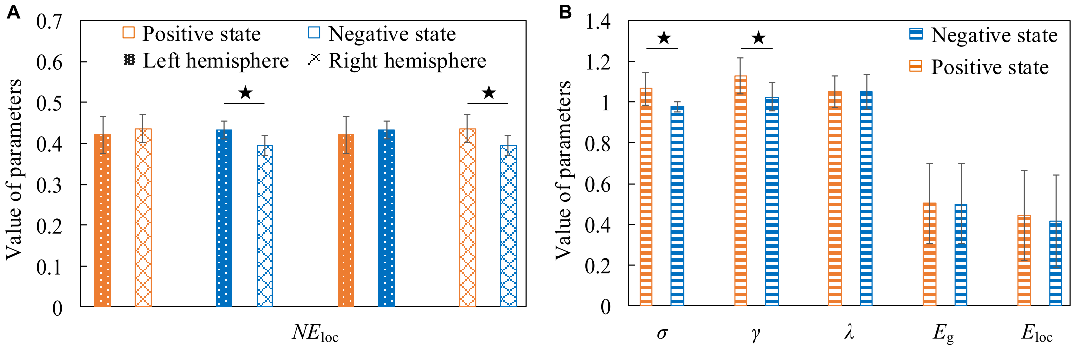


FIGURE 12

The contrast histograms of five global attribute parameters and one local attribute parameter of the functional brain network during positive and negative states. **(A)** The NE_{loc} in different situations when the electroencephalogram (EEG) is at low frequency. **(B)** Five global attribute parameters in two different states when the EEG is at high frequency.

In conclusion, we use Table 2 to summarize the results of the mental state evoked experiment and the cognitive training intervention experiment.

Discussions

In this paper, the dynamic changes of perception, cognitive level and operational skills are mainly attributed to the diversity of mental states and neuroplasticity of the brain, and the phenomenon was innovatively explained from the two dimensions of brain space and time. A typical mental state evoked paradigm and cognitive training intervention paradigm were designed, and a functional brain network analysis method was developed. It was revealed that mental state and neuroplasticity during work can change the information processing efficiency and the response mechanism of brain regions, resulting in changes in perception, cognitive level and operational skills. This provides a theoretical basis for studying the relationship between neural information processing and behavior.

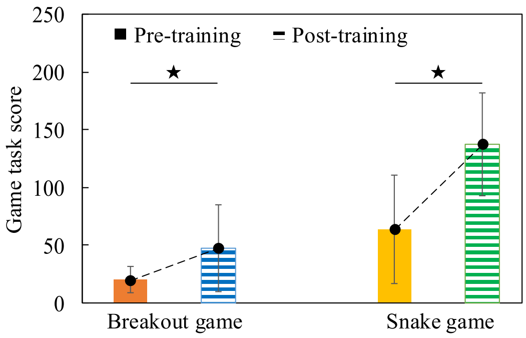


FIGURE 13

The comparison histogram of the scores of the game tasks before and after the cognitive intervention training, the dotted line indicates the change trend.

Changes in mental state during work

Firstly, by setting the task mode and applying stimulation conditions, various mental states will be induced during work, and there is a dynamic transformation between each mental state. Secondly, there is a clear difference between the mental state during work and the mental state during non-work. For example, compared with the resting state, the connection density of the functional brain network of high-frequency EEG did not decrease but increases in the fatigue state during work. The reason is that the subject needs to mobilize more brain resources to maintain a high level of perception and cognition to perform the task. Thirdly, the information of the mental state induced by the task is mainly contained in the low-frequency components of the EEG, and the high-frequency components of the EEG mainly contain task-related information. For example, when comparing resting and fatigued states, attentive and inattentive states, and positive and negative states, significant differences were found in the functional brain networks of low-frequency EEG that were altered by changes in mental state. However, brain networks with high frequency EEG had similar local connectivity features that were not influenced by mental state but were influenced by task. Finally, the mental states induced by work are more complex than those induced by non-work. From the analysis of functional brain network, it was found that the characteristic of opposition between the typical oppositional mental states (such as attentive and inattentive states, positive and negative states) during work is not obvious. The reason is that the degree and perspective of the mental state induced during work and non-work (such as audio, video and other image-induced paradigms or imagination-induced paradigms) are different.

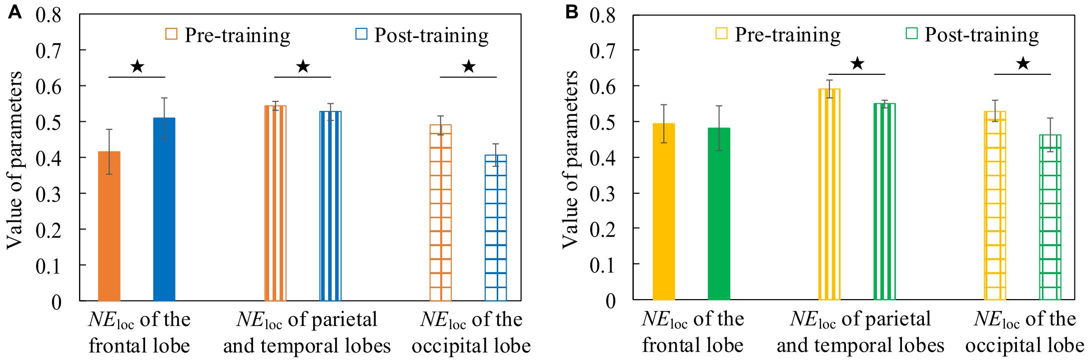


FIGURE 14
The NE_{loc} contrast histograms of functional brain networks in both pre-training and post-training conditions. **(A)** The Breakout game task. **(B)** The Snake game task.

TABLE 2 Overview of experimental results.

	Categories	Results summary
Mental state evoked experiment	Fatigue	The functional brain network connection density increased in fatigue state, and its small-world network parameters σ and γ were significantly reduced. It shows that the "economic" of the brain network in a fatigued state decreases, the information processing and transmission ability declines, and the human brain needs to consume more resources to maintain the level of perception and cognition.
	Attentive state	The σ in the attentive state increased significantly, indicating that maintaining concentration during work can improve the "economy" of the brain network, the utilization of brain resources is more efficient, and the allocation of resources is more concentrated. In addition, the NE_{loc} in the occipital lobe area also increased significantly in the attentive state, which proved that the occipital lobe area played a positive role in improving the subjects' attention.
	Positive and negative states	Functional brain networks in positive and negative states exhibit asymmetry and differences between the left and right hemispheres. The significant increase in σ and γ in the positive state indicates that the "economic" of the brain network has improved, and the information processing and transmission capabilities have become stronger, so that the human brain can more efficiently maintain the level of perception and cognition required for work.
	Common characteristics	The functional brain networks of typical mental states during work and non-work differ in their characteristics.
Cognitive training intervention experiment	Before and after training	No significant differences were found in global properties of functional brain networks before and after training, but significant differences occurred in local properties. For example, the NE_{loc} in the frontal lobe region was significantly increased after training, however, the NE_{loc} in the parietal, temporal, and occipital lobes were significantly decreased, indicating that the efficiency of parallel information transmission in the network by nodes in the brain region has changed.

Furthermore, the human brain needs to perform corresponding memory, decision-making and limb control tasks during work, so the brain needs to regulate resources to maintain a high level of perception and cognition. Therefore, the collected EEG not only contains mental state information, but also contains information about performing tasks.

Changes in neuroplasticity during work

From the perspective of neurobiological research, some researchers pointed out that neuroplasticity mainly includes three types: (1) Synaptic plasticity refers to the strengthening or weakening of synapses that promote the transmission of electrochemical signals between neurons. Changes in synapses may be caused by changes in the concentration of neurotransmitter molecules in synapses and may also be attributed to changes in post-synaptic receptor conduction (Citri and Malenka, 2008). Neurobiological studies have pointed out that the whole life course of synapses will be constantly adjusted and reconstructed to meet the needs of body function due to the changes of internal and external environment. The regulation of synaptic connection strength is controlled by synaptic plasticity in the nervous system, which regulates the synaptic connection strength through the activation state of pre-synaptic and post-synaptic neurons. (2) Intrinsic plasticity refers to the changes in the inner ability of neurons to generate or propagate action potentials. It happens inside neurons rather than at a single synapse between them. The intrinsic plasticity of neurons is usually used to self-regulate neuronal activity, accompanied by a homeostasis mechanism to maintain the dynamic balance of neuronal firing activity (Brown and Randall, 2009). And unlike the regulation of synaptic strength, individual neurons can also change their internal excitability to adapt to different synaptic inputs by changing voltage-gated ion channels (Li et al., 2013). (3) Structural plasticity is associated with the growth and development of synapses and affects the formation and destruction of synapses. structural plasticity observed in serial section electron microscopy is accompanied by synaptic deletion and geometric transformation of the dendritic spine, and that synapses with smaller spine heads are more likely to be deleted (Bourne and Harris, 2011). In fact, in most areas of the human brain, there is a large, sparse, dynamic structure of connections between neurons that generates many potential neural circuits and helps enhance the system's robustness to adapt to changing synaptic inputs (Cui et al., 2017). Changes in neural circuitry regulated by structural plasticity also greatly extend the memory storage capacity of the nervous system to adapt to changing internal and external environments (Berry and Nedivi, 2016). From a data-driven research perspective, we found significant differences in functional brain networks before and after cognitive training. For example, after subjects underwent cognitive training intervention, nodes in the frontal lobe increased the efficiency of parallel information transmission in the network, whereas nodes in the parietal, temporal, and occipital lobes decreased. Accordingly, we propose two hypotheses, the first one is the task relevance hypothesis: the parietal lobe is associated with touch, pressure, temperature, and pain; the temporal lobe is associated with perception, auditory stimulus recognition, and memory; and the occipital lobe is associated with vision (Ming, 2018). Because there are some brain regions whose function is not strongly related to the game task, while others are more strongly related. Therefore, the brain will regulate the distribution of information flow and resources in the whole brain, resulting in a decrease in the activity of nodes with weaker associated regions in the network, thereby reducing the NE_{loc} of the associated regions. The second is the brain resource requirement hypothesis: after cognitive training intervention, when the brain maintains the same level of perception, cognition and limb control, the need for information processing capacity and brain resources is reduced. For example, the subjects played the game for many times, and their sensitivity to a certain important image or sound stimulus in the game would increase with the proficiency of the game task, and the reaction time would become shorter and shorter. The required information flow and brain resources are also less and less, resulting in lower NE_{loc} in the brain regions responsible for the corresponding functions. Through the research on local attributes, it was found that after cognitive training intervention, the brain adjusted the distribution of information flow and resources in the whole brain. The node information processing and transmission capabilities of some brain regions are improved, while others are relatively reduced, which improves the efficiency and economy of functional brain networks as a whole. This further suggests that interventions through cognitive training can induce remodeling of the global topology in cortical connectivity networks and altered neuroplasticity. The findings of this paper echo with the research on neuroplasticity in neurobiological perspectives.

No matter whether the time span is years, months, days, hours, or minutes, in the process of evolution, development, and remodeling of living organisms, The connection strength of synapses between neurons, the intrinsic activation of neurons and the physical structure are all shaped by the changing internal and external environment all the time. Some scholars have proved this phenomenon through experiments. For example, Romero et al. (2008) designed a cognitive training intervention experiment on memory and reasoning. The experiment lasted for about 1 week. And the study found that the subjects had neuroplastic changes, which were reflected in the characteristics of EEG and event-related potentials. Wynn et al. (2019) designed a high-frequency visual stimulation paradigm to induce neuroplastic changes in visual evoked potential components. The experiment lasted for 2 weeks and proved that schizophrenia patients have neuroplastic deficits compared to normal people (Wynn et al., 2019). Seppanen et al. (2012) study found that long-term trained adult musicians have faster auditory perceptual learning than normal people, and music training also modulates rapid neural plasticity for sound encoding. The time span of the cognitive training intervention paradigm we designed is at the hour level, and the designed cognitive training task focuses on the subjects' brain response ability, limb manipulation ability, memory and reasoning ability, and the purpose is to study the neuroplastic changes of users during work. Compared with previous studies, our study fills the gaps in related fields to a certain extent. The purpose of this paper is to demonstrate whether cognitive training interventions during work can induce neuroplastic changes in operators. In the next step, we will further study the effect of cognitive training intervention duration on the degree of changes in user neuroplasticity. By designing an experimental paradigm with a time span of hours, days, weeks, and months, we will study the degree of neuroplasticity and the types of neuroplastic changes that occur in the subjects.

Coupling properties of mental state and neuroplasticity

We attribute the two main reasons for the changes in user perception, cognitive level, and operational skills to changes in mental state diversity and neuroplasticity during work. Because the topology of the functional brain networks underlying each mental state varies significantly in spatial dimensions and is diverse, we classify it as a property of the brain spatial dimension. Since neuroplasticity is a change in brain neural activity and neumorphism that requires repeated and regular reinforcement training in the time dimension, we classify it as a characteristic of the brain time dimension. For the purpose of engineering application, we systematically reveal the changes of mental state and neuroplasticity of the brain induced by work, these two human factors will change the response mechanism of brain regions, the efficiency of information processing and the allocation of brain resources, thereby affecting the user's perception, cognitive level and operational skills. Moreover, most of them appear at the same time, and they accompany and affect each other. For example, professional drivers (or professional operators of special equipment) will experience neuroplastic changes in their brains compared to novices due to long-term and regular training, making their operating skills far superior to novices. When encountering emergencies in the operation, the mental state caused by them will also be different. Novices often have tension and stress reactions, while professional drivers are relatively calm. Therefore, it is necessary to carry out research on the interaction and coupling of mental state and neuroplasticity of the brain during work in the future. When addressing the impact of human factors on human-machine collaboration systems, these two factors are best considered together rather than separately. And it is not only necessary to consider the respective characteristics of these two factors, but also the coupling effect between them should be comprehensively considered.

Future work and limitations

The brain is a complex network consisting of spatially distributed regions dedicated to different functions. It is proposed that cognitive functions emerge from dynamic interactions of several brain areas, not from the activation of a single brain region. We propose here that the brain connectome can be used not only to characterize the diversity of mental states and changes in neuroplasticity, but also to analyze the information processing mechanisms and mental expression mechanisms in the brain, so as to understand the reasons for the changes in perception and cognition levels and the underlying neurobiological mechanisms in the process of human-machine collaboration. This research contributes to the improvement of human-computer interface and the development of new adaptive automation systems. And it helps to drive a new mode of human-machine interaction. This model is trying to transform the way humans and machines work together. It integrates neurophysiological detection techniques in evaluating the performance of human-machine systems, rather than relying solely on the measurement of workers' explicit behaviors and subjective perceptions. At the heart of this model lies the first grasp of how the brain processes perceptual and cognitive information. Then, through the monitoring, learning and inference of continuous neurophysiological signals, we can understand some changing trends, behavior patterns and application scenarios related to the user's job content and goals, so as to be more conducive to application in complex and dynamic human-machine interaction environment. Based on the research conclusions of this paper, we will carry out deeper research in the future, and evaluate the quality and performance of system work information processing by analyzing the perception and cognition level of operators on work tasks and the functional neural network behind them, and compare it with the actual work performance, and then prompt or interfere with the operator to improve the operation or adaptively adjust the system work parameters to improve the system work performance and ensure the safety of personnel and the best performance. Additionally, we will develop brain augmentation systems from a human-centric perspective by dissecting the laws of mental state diversity and neuroplasticity that occur during human-machine interaction. This mainly includes neuro-enhancement technology and augmented cognition technology. Among them, the goal of neuro-enhancement technology is to use current advanced stimulation methods (such as transcranial magnetic stimulation and transcranial direct current stimulation) to enhance people's visual, auditory, sustained attention, positive emotion retention, working memory, logical reasoning, and motor learning ability. The goal of augmented cognition technology is to start from the known limitations of human cognition and use computer-based methods to design to break through the bottleneck of human beings and solve the deviations and deficiencies in human cognition.

On the basis of the research conclusions of this paper, in order to pursue deeper research, the influence of individual and task differences on mental state and neuroplasticity can be the content of further research. For example, in the aspect of neuroplasticity, whether the changes of neuroplasticity have an important relationship with the cognitive level and behavioral ability of the subjects. If the cognitive training task is too difficult or too easy for the subjects, whether it will make it difficult for them to maintain motivation and attention. Such as some people may be exhausted by excessive cognitive workload, while others may be bored by too simplistic cognitive workload. Whether these two extremes negatively affect the user's neuroplasticity. These studies can further explain why functional brain networks in some cases did not differ significantly. Therefore, the next step is to pay attention to the differences of individual users in the human-machine collaboration systems, which needs to be studied urgently.

Despite the significance of this work, several limitations should be considered. First of all, the human factors studied in this paper mainly include mental state and neuroplasticity. It has to be admitted that human factors include many aspects. It is a comprehensive definition, and many human-related factors can be classified into it, and there are also many explanations and descriptions (Wang et al., 2020), and so far there is no accepted name or definition in academia. But this does not affect some of the contributions made by this paper in related fields such as human-machine collaboration, human factors engineering, and neuroergonomics. Secondly, it has to be admitted that compared with MEG and fMRI, EEG has the disadvantage of low spatial resolution, which results in that the spatial resolution of functional brain networks established by EEG cannot be accurate to very fine areas. However, considering its high temporal resolution and the portability of EEG acquisition equipment, it provides feasibility for the study of mental state diversity and neuroplasticity of the brain during work, which is why we choose EEG. Finally, a total of 26 subjects participated in the experiment, which to a certain extent can prove the correctness of the proposed ideas and highlight some important experimental results. Further research will involve testing the proposed hypotheses and results with a larger number of subjects and expanding the research to include studies of inter-individual and gender differences.

Conclusion

In this paper, the evoked paradigm of typical mental state and the analysis method of its functional brain network were proposed, and it was revealed that the mental state of the subjects would change dynamically during the work process and showed the characteristics of diversity. There were significant differences between functional brain networks in different mental states, and the information processing efficiency and the mechanism of brain area response had changed significantly. At the same time, the cognitive training intervention paradigm and its functional brain network analysis method were proposed. It revealed that there was a significant difference between the functional brain networks of the subjects before and after the cognitive training intervention experiment, and the brain adjusted the distribution of information flow and resources in the whole brain. The node information processing and transmission capabilities of some brain regions are improved, while others are relatively reduced, which improves the efficiency and economy of functional brain networks as a whole. This further suggests that interventions through cognitive training can induce remodeling of the global topology in cortical connectivity networks and altered neuroplasticity. To sum up, this paper innovatively and comprehensively studies the dynamic changes of the brain in space and time dimensions, and reveals that mental state and neuroplasticity during work can change the information processing efficiency and the response mechanism of brain regions, resulting in changes in perception and cognitive levels. We expound the complexity of human factors in the field of human-machine interaction and the value behind them from a more comprehensive dimension, which can effectively improve the ability of humans to understand and simulate complex human-machine system interactions at work. The conclusion of this study provides a theoretical basis for the relationship between neural information processing and behavior, and enriches the research connotation in the field of neuroergonomics.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional Review Board of Xi'an Jiaotong University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

TZ proposed and did the research and wrote the manuscript. XZ proposed the research idea, supervised the work, and revised the manuscript. WZ organized and conducted part of the experiment. ZL assisted in processing experimental data. YW assisted in collecting experimental data. YZ revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported in part by the National Key Research and Development Program of China (No. 2017YFB1300303) and partially supported by grants from the Science and Technology Supporting Xinjiang Project (No. 2020E0259).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Baillet, S. (2017). Magnetoencephalography for brain electrophysiology and imaging. *Nat. Neurosci.* 20, 327–339. doi: 10.1038/nn.4504

Balasubramanian, G., Kanagasabai, A., Mohan, J., and Seshadri, G. N. P. (2018). Music induced emotion using wavelet packet decomposition—An EEG study. *Biomed. Signal Process. Control* 42, 115–128. doi: 10.1016/j.bspc.2018.01.015

Ballesteros, S., Mayas, J., Prieto, A., Toril, P., Pita, C., Laura, P. L., et al. (2015). A randomized controlled trial of brain training with non-action video games in older adults: Results of the 3-month follow-up. *Front. Aging Neurosci.* 7:45. doi: 10.3389/fnagi.2015.00045

Bavelier, D., and Green, C. S. (2003). Action video game modifies visual selective attention. *Nature* 423, 534–537. doi: 10.1038/nature01647

Berry, K. P., and Nedivi, E. (2016). Experience-dependent structural plasticity in the visual system. *Annu. Rev. Vis. Sci.* 2, 17–35. doi: 10.1146/annurev-vision-111815-114638

Bevilacqua, V., Brunetti, A., Trigiante, G., Trotta, G. F., Fiorentino, M., Manghisi, V., et al. (2016). *Design and development of a forearm rehabilitation system based on an augmented reality serious game*. (Cham: Springer International Publishing), 127–136. doi: 10.1007/978-3-319-32695-5_12

Boksem, M. A., and Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Res. Rev.* 59, 125–139. doi: 10.1016/j.brainresrev.2008.07.001

Boksem, M. A. S., Meijman, T. F., and Lorist, M. M. (2005). Effects of mental fatigue on attention: An ERP study. *Cogn. Brain Res.* 25, 107–116. doi: 10.1016/j.cogbrainres.2005.04.011

Boksem, M. A. S., Meijman, T. F., and Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biol. Psychol.* 72, 123–132. doi: 10.1016/j.biopsycho.2005.08.007

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003

Bourne, J. N., and Harris, K. M. (2011). Nanoscale analysis of structural synaptic plasticity. *Curr. Opin. Neurobiol.* 22, 372–382. doi: 10.1016/j.conb.2011.10.019

Brown, J. T., and Randall, A. D. (2009). Activity-dependent depression of the spike after-depolarization generates long-lasting intrinsic plasticity in hippocampal CA3 pyramidal neurons. *J. Physiol.* 587, 1265–1281. doi: 10.1113/jphysiol.2008.167007

Cannonieri, G. C., Bonilha, L., Fernandes, P. T., Cendes, F., and Li, L. M. (2007). Practice and perfect: Length of training and structural brain changes in experienced typists. *Neuroreport* 18, 1063–1066. doi: 10.1097/WNR.0b013e3281a030e5

Citri, A., and Malenka, R. C. (2008). Synaptic plasticity: Multiple forms, functions, and mechanisms. *Neuropsychopharmacology* 33, 18–41. doi: 10.1038/sj.npp.1301559

Cuccu, G., Togelius, J., and Cudre-Mauroux, P. (2021). Playing Atari with few neurons; Improving the efficacy of reinforcement learning by decoupling feature extraction and decision making. *Auton. Agents Multi Agent Syst.* 35:17. doi: 10.1007/s10458-021-09497-8

Cui, Y., Ahmad, S., and Hawkins, J. (2017). The HTM spatial pooler-A neocortical algorithm for online sparse distributed coding. *Front. Comput. Neurosci.* 11:111. doi: 10.3389/fncom.2017.00111

Draganski, B., and May, A. (2008). Training-induced structural changes in the adult human brain. *Behav. Brain Res.* 192, 137–142. doi: 10.1016/j.bbr.2008.02.015

Du, N., Zhou, F., Pulver, E. M., Tilbury, D. M., Robert, L. P., Pradhan, A. K., et al. (2020). Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving. *Transp. Res. C Emerg. Technol.* 112, 78–87. doi: 10.1016/j.trc.2020.01.006

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., and Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science* 270, 305–307. doi: 10.1126/science.270.5234.305

Feitosa, J. A., Fernandes, C. A., Casseb, R. F., and Castellano, G. (2022). Effects of virtual reality-based motor rehabilitation: A systematic review of fMRI studies. *J. Neural Eng.* 19:011002. doi: 10.1088/1741-2552/ac456e

Frey, J., Mühl, C., Lotte, F., and Hachet, M. (2013). "Review of the use of electroencephalography as an evaluation method for human-computer interaction," in *Proceedings of the 2013 international conference on physiological computing systems*, Angers.

Ghassemzadeh, H., Rothbart, M. K., and Posner, M. I. (2019). Anxiety and brain networks of attentional control. *Cogn. Behav. Neurol.* 32, 54–62. doi: 10.1097/WNN.0000000000000181

Hancock, P. A. (2013). In search of vigilance the problem of iatrogenically created psychological phenomena. *Am. Psychol.* 68, 97–109. doi: 10.1037/a0030214

Harry, E., Sinsky, C., Dyrbye, L. N., Makowski, M. S., Trockel, M., Tutty, M., et al. (2021). Physician task load and the risk of burnout among US physicians in a national survey. *Jt. Commun. J. Qual. Patient. Saf.* 47, 76–85. doi: 10.1016/j.jcjq.2020.09.011

John, O. P., and Gross, J. J. (2004). Healthy and unhealthy emotion regulation: Personality processes, individual differences, and life span development. *J. Pers.* 72, 1301–1333. doi: 10.1111/j.1467-6494.2004.00298.x

Kivikangas, J. M. A., Chanel, G. B., Cowley, B. A., Ekman, I. A., Salminen, M. A., Järvelä, S. A., et al. (2011). A review of the use of psychophysiological methods in game research. *J. Gam. Virtual Worlds* 3, 181–199. doi: 10.1386/jgvw.3.3.181_1

Langner, R., and Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychol. Bull.* 139, 870–900. doi: 10.1037/a0030694

Li, G., Luo, Y., Zhang, Z., Xu, Y., Jiao, W., Jiang, Y., et al. (2019). Effects of mental fatigue on small-world brain functional network organization. *Neural Plast.* 2019:1716074. doi: 10.1155/2019/1716074

Li, R., Liu, D., Li, Z., Liu, J., Zhou, J., Liu, W., et al. (2022). A novel EEG decoding method for a facial-expression-based BCI system using the combined convolutional neural network and genetic algorithm. *Front. Neurosci.* 16:988535. doi: 10.3389/fnins.2022.988535

Li, Y., Li, C., and Vasilaki, E. (2013). Synergies between intrinsic and synaptic plasticity based on information theoretic learning. *PLoS One* 8:e62894–e62894. doi: 10.1371/journal.pone.0062894

Liu, X., Li, T., Tang, C., Xu, T., Chen, P., Bezerianos, A., et al. (2019). Emotion Recognition and Dynamic Functional Connectivity Analysis Based on EEG. *IEEE Access* 7, 143293–143302. doi: 10.1109/ACCESS.2019.2945059

Lorist, M. M., Boksem, M. A., and Ridderinkhof, K. R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cogn. Brain Res.* 24, 199–205. doi: 10.1016/j.cogbrainres.2005.01.018

Luna, F. G., Barttfeld, P., Martín-Arévalo, E., and Lupiáñez, J. (2021). The ANTI-Vea task: Analyzing the executive and arousal vigilance decrements while measuring the three attentional networks. *Psicológica* 42, 1–26. doi: 10.2478/psicolj-2021-0001

Maguire, E. A., Woollett, K., and Spiers, H. J. (2006). London taxi drivers and bus drivers: A structural MRI and neuropsychological analysis. *Hippocampus* 16, 1091–1101. doi: 10.1002/hipo.20233

Ming, D. (2018). *Neural engineering*. Beijing: Science Press.

Parasuraman, R. (2011). Neuroergonomics: Brain, cognition, and performance at work. *Curr. Dir. Psychol. Sci.* 20, 181–186. doi: 10.1177/0963721411409176

Peng, D. (2019). *General Psychology*. Beijing: Beijing Normal University Press.

Peters, E., Västfjäll, D., Gärling, T., and Slovic, P. (2006). Affect and decision making: A "hot" topic. *J. Behav. Decis. Mak.* 19, 79–85. doi: 10.1002/bdm.528

Ponten, S. C., Daffertshofer, A., Hillebrand, A., and Stam, C. J. (2010). The relationship between structural and functional connectivity: Graph theoretical analysis of an EEG neural mass model. *NeuroImage* 52, 985–994. doi: 10.1016/j.neuroimage.2009.10.049

Power, J. D., Plitt, M., Laumann, T. O., and Martin, A. (2017). Sources and implications of whole-brain fMRI signals in humans. *Neuroimage* 146, 609–625. doi: 10.1016/j.neuroimage.2016.09.038

Reeves, B. R., Newhagen, J., Maibach, E., Basil, M., and Kurz, K. (1991). Negative and positive television messages: Effects of message type and context on attention and memory. *Am. Behav. Sci.* 34, 679–694. doi: 10.1177/0002764291034006006

Ricci, J. A., Chee, E., Lorandeau, A. L., and Berger, J. (2007). Fatigue in the U.S. Workforce: Prevalence and Implications for Lost Productive Work Time. *J. Occup. Environ. Med.* 49, 1–10. doi: 10.1097/01.jom.0000249782.60321.2a

Romero, S. G., McFarland, D. J., Faust, R., Farrell, L., and Cacace, A. T. (2008). Electrophysiological markers of skill-related neuroplasticity. *Biol. Psychol.* 78, 221–230. doi: 10.1016/j.biopsycho.2008.03.014

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., et al. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nat. Neurosci.* 19, 165–171. doi: 10.1038/nn.4179

Seidel-Marzi, O., and Ragert, P. (2020). Neurodiagnostics in sports: Investigating the Athlete's brain to augment performance and sport-specific skills. *Front. Hum. Neurosci.* 14:133. doi: 10.3389/fnhum.2020.00133

Seppanen, M., Hamalainen, J., Pesonen, A. K., and Tervaniemi, M. (2012). Music training enhances rapid neural plasticity of n1 and p2 source activation for unattended sounds. *Front. Hum. Neurosci.* 6:43. doi: 10.3389/fnhum.2012.00043

Stam, C. J., de Haan, W., Daffertshofer, A., Jones, B. F., Manshanden, I., van Cappellen van Walsum, A., et al. (2009). Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease. *Brain* 132, 213–224. doi: 10.1093/brain/awn262

Stanley, E. A., and Larsen, K. L. (2021). Difficulties with emotion regulation in the contemporary U.S. Armed Forces: Structural contributors and potential solutions. *Armed Forces Soc.* 47, 77–105. doi: 10.1177/0095327X19848018

Sun, Y., Lim, J., Kwok, K., and Bezerianos, A. (2014). Functional cortical connectivity analysis of mental fatigue unmasks hemispheric asymmetry and changes in small-world networks. *Brain Cogn.* 85, 220–230. doi: 10.1016/j.bandc.2013.12.011

Taya, F., Sun, Y., Babiloni, F., Thakor, N., and Bezerianos, A. (2015). Brain enhancement through cognitive training: A new insight from brain connectome. *Front. Syst. Neurosci.* 9:44. doi: 10.3389/fnsys.2015.00044

Thibaut, A., Russo, C., Morales-Quezada, L., Hurtado-Puerto, A., Deitos, A., Freedman, S., et al. (2017). Neural signature of tDCS, tPCS and their combination: Comparing the effects on neural plasticity. *Neurosci. Lett.* 637, 207–214. doi: 10.1016/j.neulet.2016.10.026

Valentin, L. S. S. (2017). Can digital games be a way of improving the neuroplasticity in stroke damage? Can the adult brain grow new cells or rewire itself in response to a new experience? *Open J. Med. Psychol.* 06, 153–165. doi: 10.4236/ojmp.2017.62013

Wang, B., Huang, S., and Yi, B. (2020). State-of-art of human factors/ergonomics in intelligent manufacturing. *Chin. J. Mech. Eng.* 56, 240–253.

Wang, X. (2020). *Adaptive learning of echo state networks based on neural plasticity mechanisms*. Ph.D. dissertation. Shang Hai: Donghua University.

Warm, J. S., Parasuraman, R., and Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Hum. Fact.* 50, 433–441. doi: 10.1518/001872008X312152

Wilson, G. F. (2005). Operator functional state assessment for adaptive automation implementation. *Proc. SPIE-Int. Soc. Optical Eng.* 5797, 100–104. doi: 10.1117/12.601806

Winter, W. R., Nunez, P. L., Ding, J., and Srinivasan, R. (2007). Comparison of the effect of volume conduction on EEG coherence with the effect of field spread on MEG coherence. *Stat. Med.* 26, 3946–3957. doi: 10.1002/sim.2978

Wu, X., Zheng, W. L., Li, Z., and Lu, B. L. (2022). Investigating EEG-based functional connectivity patterns for multimodal emotion recognition. *J. Neural Eng.* 19:016012. doi: 10.1088/1741-2552/ac49a7

Wynn, J. K., Roach, B. J., McCleery, A., Marder, S. R., Mathalon, D. H., and Green, M. F. (2019). Evaluating visual neuroplasticity with EEG in schizophrenia outpatients. *Schizophr. Res.* 212, 40–46. doi: 10.1016/j.schres.2019.08.015

Xu, M., Fralick, D., Zheng, J. Z., Wang, B., Tu, X. M., and Feng, C. (2017). The Differences and Similarities Between Two-Sample T-Test and Paired T-Test. *Shanghai Arch. Psychiatry* 29, 184–188. doi: 10.11919/j.issn.1002-0829.217070

Yang, B. (2009). *ERP study of information automatic processing and conflict processing in the mental fatigue.* Ph.D. dissertation. Xi'an: Fourth Military Medical University.

Zhang, T., Zhang, X., Lu, Z., Zhang, Y., Jiang, Z., and Zhang, Y. (2022). Feasibility study of personalized speed adaptation method based on mental state for teleoperated robots. *Front. Neurosci.* 16:976437. doi: 10.3389/fnins.2022.976437

Zhang, T., Zhang, X., Zhang, Y., Lu, Z., and Li, H. (2019). "Effects of user fatigue mental state on the facial-expression paradigm of BCI," in *Proceedings of the 2019 WRC Symposium on Advanced Robotics and Automation (WRC SARA)*, Beijing. doi: 10.1109/WRC-SARA.2019.8931802

Zhang, T., Zhang, X., Zhang, Y., Lu, Z., Zhu, W., and Jiang, Y. (2021). A precise control method for brain-computer cooperation with deep reinforcement learning. *J. Xi'an Jiaotong Univ.* 2021, 1–9.