frontiers | Research Topics

Recent advances and future directions in facial appearance research

Edited by
Ayahito Ito, Koyo Nakamura, Yoshiyuki Ueda and Chihiro Saegusa

Published in
Frontiers in Psychology

ISSN 1664-8714
ISBN 978-2-83251-849-6
DOI 10.3389/978-2-83251-849-6

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

Recent advances and future directions in facial appearance research

Topic editors
Ayahito Ito — Kochi University of Technology, Japan
Koyo Nakamura — University of Vienna, Austria
Yoshiyuki Ueda — Kyoto University, Japan
Chihiro Saegusa — Kao Corporation, Japan

Citation
Ito, A., Nakamura, K., Ueda, Y., Saegusa, C., eds. (2023). *Recent advances and future directions in facial appearance research*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83251-849-6

Table of contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 22 February 2023
DOI 10.3389/fpsyg.2023.1154703

OPEN ACCESS

EDITED AND REVIEWED BY
Barnaby James Wyld Dixson,
The University of Queensland, Australia

*CORRESPONDENCE
Yoshiyuki Ueda
✉ ueda.yoshiyuki.3e@kyoto-u.jp
Koyo Nakamura
✉ koyo.nakamura@univie.ac.at
Chihiro Saegusa
✉ saegusa.chihiro@kao.com
Ayahito Ito
✉ ayahito.ito@gmail.com

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

RECEIVED 31 January 2023
ACCEPTED 09 February 2023
PUBLISHED 22 February 2023

CITATION
Ueda Y, Nakamura K, Saegusa C and Ito A (2023) Editorial: Recent advances and future directions in facial appearance research. *Front. Psychol.* 14:1154703. doi: 10.3389/fpsyg.2023.1154703

Editorial: Recent advances and future directions in facial appearance research

Yoshiyuki Ueda[1]*, Koyo Nakamura[2,3]*, Chihiro Saegusa[4]* and Ayahito Ito[5,6,7]*

[1]Institute for the Future of Human Society, Kyoto University, Kyoto, Japan, [2]Faculty of Psychology, University of Vienna, Vienna, Austria, [3]Japan Society for the Promotion of Science, Tokyo, Japan, [4]Kao Corporation, Tokyo, Japan, [5]Research Institute for Future Design, Kochi University of Technology, Kami, Japan, [6]Department of Psychology, University of Southampton, Southampton, United Kingdom, [7]Faculty of Health Sciences, Hokkaido University, Sapporo, Japan

KEYWORDS

face, impression, judgment, social perception, representation, visualization, fMRI

Editorial on the Research Topic
Recent advances and future directions in facial appearance research

This Research Topic aims to present readers with a group of papers that represent state-of-the-art facial appearance research. Faces are important in identifying individuals. However, they also significantly contribute to our estimation of the kind of a person someone is and a person's current emotions and moods. Owing to its social significance, facial appearance research has long been at the core of psychological research. The papers published in the Research Topic are very diverse and cover a variety of topics in this field. These studies include the internal representation of faces (Elson et al.; Minemoto and Ueda), personal characteristics inferred from faces (Albohn et al.; Ito et al.; Komori et al.; Mattavelli et al.; Sakuta; Ueda), visualization and face modeling (Albohn et al.; Komori et al.; Watson and Johnston), and impressions given by sanitary face masks (Dudarev et al.; Kamatani et al.). These methodologies include standard rating tasks (Dudarev et al.; Kamatani et al.; Mattavelli et al.; Sakuta), adaptation (Minemoto and Ueda), as well as visualization using principal component analysis (PCA; Watson and Johnston), StyleGAN (Albohn et al.), and Bayesian optimization (Komori et al.), and functional magnetic resonance imaging (fMRI; Elson et al.; Ito et al.; Ueda). Some studies have proposed a new methodology for facial appearance research (Albohn et al.; Watson and Johnston), created a new database (Mattavelli et al.), examined young children (Sakuta) and cultural differences (Dudarev et al.; Kamatani et al.; Sakuta), and reviewed previous studies (Minemoto and Ueda; Ueda).

The idea of "face space" (Valentine, 1991; Valentine et al., 2016) has been extensively used to understand the neural representation of faces. To understand the neural representation of faces distant from an average face, Elson et al. investigated the effects of caricaturing faces on neural responses using fMRI. Their findings indicate that the neural processing of caricatured faces recruits an object-selective cortex.

Adaptation and aftereffect methods attempt to examine face space in terms of behavioral indicators. Minemoto and Ueda reviewed recent research and showed whether the norm-based coding hypothesis, a theory for representing faces, can be applied to children and people with impaired face recognition. They also presented studies investigating the representation of ensembles (summary statistics) of faces and social signals via faces, suggesting a new direction for face adaptation paradigms.

Emerging evidence shows that the determination of trait impressions of faces can vary across individuals and cultures to a non-negligible extent (Hehman et al., 2017; Jones et al., 2021). Therefore, it is necessary to rethink the universality assumed in early theories. Sakuta attempted to test the universality of facial trait inferences by investigating young non-Western children and adults. The results showed that trustworthiness and competence inferred from facial appearance are culturally diverse, whereas dominance remains relatively shared across cultures.

Facial appearance influences various social relationships; early romantic feelings are no exception. Ueda marshaled previous literature on how human brains integrate perceived facial attractiveness into initial romantic attraction. He proposed that an important topic for future research is how human brains shape a persistent attraction to a particular person, which should be examined by integrating neuroimaging with ecologically valid tasks.

Regarding "cuteness," another impression perceived from faces, Komori et al. aimed to identify the facial features related to this perception. They used Bayesian optimization, a global sequential optimization method for estimating unknown functions, to search for facial morphological features that enhance the perception of facial cuteness. The results showed that perceived cuteness was linked to a relatively lower position of facial components and narrower jawline, but not to forehead height.

Face evaluation and first impressions are influenced not only by invariant facial features but also by multiple other factors; however, the integrated modulation has not been systematically investigated. Mattavelli et al. created a new database, referred to as the Bi-AGI database, consisting of male and female faces of varying age, gaze direction, and illumination conditions, and found that illumination has a greater effect on face evaluation of younger faces. The new database is freely available.

A key to success in the scientific modeling of faces is to capture the facial features that appear in real faces. Albohn et al. introduced a novel paradigm for generating and visualizing photorealistic face images that correspond to an individual's mental representation. Previous research on faces broadly examined group-level consensus. However, individual differences are larger than the variability in stimulus levels and have limitations in that individual face representations are examined using noisy and blurred visualizations. Their proposed method can generate generalizable faces, thereby resulting in an understanding of how social judgments are formed.

Watson and Johnston proposed a PCA-based active-appearance model to capture spatiotemporal variations in dynamic facial expressions. Extending the existing face caricaturing techniques, this model allows the generation of dynamically caricatured and realistic faces.

To uncover the perceptual mechanisms of facial appearance, the characteristics of the face as well as those of the perceiver should be considered. Ito et al. investigated own-age bias in facial impression formation and noted a higher preference for young faces than for older faces, regardless of the age of participants. This suggests that preferential choice of face is less susceptible to own-age bias across the lifespan of individuals.

The perception of mask-wearing faces is gaining attention owing to their widespread use due to COVID-19. Recent studies have shown mixed results regarding whether wearing a sanitary mask enhances the perceived facial attractiveness. Kamatani et al. proposed a new hypothesis stating that the occluded area of the face is interpolated by a moderately attractive face shaped by each person's experience, and further examined it using the other-race effect.

Dudarev et al. aimed to elucidate the influence of wearing sanitary masks on the perceived attractiveness of the wearer. They showed that pro-mask participants rated individuals with masks as generally more attractive than individuals without masks, whereas anti-mask participants rated vice versa. These results suggest that perceived attractiveness is also affected by perceiver characteristics.

These studies will provide readers with useful contemporary psychological findings regarding facial appearance and the impressions derived from it. Furthermore, they provide not only new scientific insights but also novel and freely available methods and stimulus sets that will further elucidate the mechanisms of facial appearance processing. Facial appearance, for better or worse, governs human society. We hope this Research Topic contributes to clarifying how we make decisions and are biased regarding them.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

CS is employed by Kao Corporation.

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Hehman, E., Sutherland, C. A., Flake, J. K., and Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *J. Pers. Soc. Psychol.* 113, 513. doi: 10.1037/pspa0000090

Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., et al. (2021). To which world regions does the valence–dominance model of social perception apply? *Nat. Hum. Behav.* 5, 159–169. doi: 10.1038/s41562-020-01007-2

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol. A*. 43, 161–204. doi: 10.1080/14640749108400966

Valentine, T., Lewis, M. B., and Hills, P. J. (2016). Face-space: a unifying concept in face recognition research. *Q. J. Exp. Psychol.* 69, 1996–2019. doi: 10.1080/17470218.2014.990392

frontiers in Psychology

ORIGINAL RESEARCH
published: 11 March 2022
doi: 10.3389/fpsyg.2022.822234

Check for updates

The Role of the Ventromedial Prefrontal Cortex in Preferential Decisions for Own- and Other-Age Faces

Ayahito Ito[1]*, Kazuki Yoshida[2]*, Ryuta Aoki[3], Toshikatsu Fujii[4], Iori Kawasaki[5], Akiko Hayashi[5], Aya Ueno[5], Shinya Sakai[2], Shunji Mugikura[6,7], Shoki Takahashi[7] and Etsuro Mori[5]

[1] Research Institute for Future Design, Kochi University of Technology, Kochi, Japan, [2] Faculty of Health Sciences, Hokkaido University, Sapporo, Japan, [3] Graduate School of Humanities, Tokyo Metropolitan University, Tokyo, Japan, [4] Kansei Fukushi Research Institute, Tohoku Fukushi University, Sendai, Japan, [5] Department of Behavioral Neurology and Cognitive Neuroscience, Graduate School of Medicine, Tohoku University, Sendai, Japan, [6] Division of Image Statistics, Tohoku Medical Megabank Organization, Sendai, Japan, [7] Department of Diagnostic Radiology, Graduate School of Medicine, Tohoku University, Sendai, Japan

OPEN ACCESS

Edited by:
Louise A. Brown Nicholls, University of Strathclyde, United Kingdom

Reviewed by:
Mingzhu Hou, The University of Texas at Dallas, United States
Niv Reggev, Ben-Gurion University of the Negev, Israel

****Correspondence:***
Ayahito Ito
ayahito.ito@gmail.com
Kazuki Yoshida
ot-k-yoshida@huhp.hokudai.ac.jp

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

Received: *25 November 2021*
Accepted: *16 February 2022*
Published: *11 March 2022*

Citation:
Ito A, Yoshida K, Aoki R, Fujii T, Kawasaki I, Hayashi A, Ueno A, Sakai S, Mugikura S, Takahashi S and Mori E (2022) The Role of the Ventromedial Prefrontal Cortex in Preferential Decisions for Own- and Other-Age Faces. Front. Psychol. 13:822234. doi: 10.3389/fpsyg.2022.822234

Own-age bias is a well-known bias reflecting the effects of age, and its role has been demonstrated, particularly, in face recognition. However, it remains unclear whether an own-age bias exists in facial impression formation. In the present study, we used three datasets from two published and one unpublished functional magnetic resonance imaging (fMRI) study that employed the same pleasantness rating task with fMRI scanning and preferential choice task after the fMRI to investigate whether healthy young and older participants showed own-age effects in face preference. Specifically, we employed a drift-diffusion model to elaborate the existence of own-age bias in the processes of preferential choice. The behavioral results showed higher rating scores and higher drift rate for young faces than for older faces, regardless of the ages of participants. We identified a young-age effect, but not an own-age effect. Neuroimaging results from aggregation analysis of the three datasets suggest a possibility that the ventromedial prefrontal cortex (vmPFC) was associated with evidence accumulation of own-age faces; however, no clear evidence was provided. Importantly, we found no age-related decline in the responsiveness of the vmPFC to subjective pleasantness of faces, and both young and older participants showed a contribution of the vmPFC to the parametric representation of the subjective value of face and functional coupling between the vmPFC and ventral visual area, which reflects face preference. These results suggest that the preferential choice of face is less susceptible to the own-age bias across the lifespan of individuals.

Keywords: face, subjective value, preference, connectivity, functional magnetic brain imaging, own-age, other-age

INTRODUCTION

Age has a prominent effect on face perception (Korthase and Trenholme, 1982; McLellan and Mckelvie, 1993; Perrett et al., 1998; Zebrowitz et al., 2003; Rhodes, 2006). Own-age bias is a well-known bias that reflects the effects of age on face perception (Anastasi and Rhodes, 2005; Rhodes and Anastasi, 2012). Previous studies have indicated that own-age faces are better recognized and remembered than other-age faces (Wright and Stroud, 2002; Anastasi and Rhodes, 2005). For example, Wright and Stroud (2002) showed that older culprits were better recognized by older persons than by young persons, whereas young culprits were better recognized by young persons than by older persons. Furthermore, a seminal meta-analysis conducted by Rhodes and Anastasi (2012) showed that the recognition memory for own-age faces is better than that for other-age faces across the lifespan of individuals. These results have been explained by the increased contact hypothesis (i.e., a higher frequency of contact with own-group individuals increases the expertise of face perception) (Chiroro and Valentine, 1995) and/or socio-cognitive accounts (i.e., in-group/out-group categorization of faces results in memory bias and own-group faces are better remembered than other-group faces) (Bernstein et al., 2007). Thus, own-age bias is considered to be rooted in the importance of or experience with one's own age group in their daily lives, and the amount of exposure to a certain age group modulates the perceptual expertise of faces within the group (Ebner and Johnson, 2009; He et al., 2011; Macchi Cassia et al., 2012).

However, it remains unclear whether own-age bias exists in facial impression formation. Facial impressions are a fundamental aspect of human society, and their influence varies from mate preference to the results of political elections (Buss, 1985; Todorov et al., 2005; Todd et al., 2007; Berggren et al., 2010). Recently, researchers found that own-age bias existed in visual attention and emotion recognition, which appeared to be related to the own-age bias in face preference (Ebner et al., 2011c, 2013; He et al., 2011). For example, a previous eye-tracking study that employed a passive face viewing task revealed that people see own-age faces longer than other-age faces (He et al., 2011). Another study that used an emotion identification task also supported this finding (Ebner et al., 2011c). Although the own-age bias in visual attention is yet to be determined (Strickland-Hughes et al., 2020), own-age bias may affect facial preference via heightened visual attention toward own-age faces.

When people choose a face that they prefer from two alternatives, the subjective value of each face is computed, and these values are compared and preferential choices are made (Levy and Glimcher, 2011). Previous studies using functional magnetic resonance imaging (fMRI) have shown that the ventromedial prefrontal cortex (vmPFC) computes value signals (Chib et al., 2009; Lebreton et al., 2009; Hare et al., 2010; Ito et al., 2016; Suzuki et al., 2017; Suzuki and O'Doherty, 2020; Pessiglione and Daunizeau, 2021). Specifically, it has been argued that the brain valuation system (BVS), including the vmPFC, parametrically represents subjective values of faces and predicts later preferential choices (Lebreton et al., 2009). Other strands of evidence suggest that gaze biases preferential choice by changing attention-dependent relative value signals of choice options (Shimojo et al., 2003; Armel et al., 2008; Lim et al., 2011). It has been shown that exogenous changes in gaze duration can bias participants' decisions (Shimojo et al., 2003; Armel et al., 2008), suggesting that the brain computes attention-dependent relative value signals. Lim et al. (2011) demonstrated that the vmPFC is associated with this value computation. Thus, different gaze patterns among people for own-age faces and other-age faces may bias subjective ratings and vmPFC activity for own-age faces. Although a previous study suggested the possibility of the existence of own-age bias in the representation of the subjective value of faces (Ito et al., 2016), this question has rarely been explicitly addressed.

To formally investigate whether the own-age bias exists in face preference, we used a drift-diffusion model (DDM), which is a type of sequential sampling model that assumes that decision making is a process comprising a noisy accumulation of evidence from a stimulus (Ratcliff and McKoon, 2008; Ratcliff et al., 2016). The DDM has been widely used to represent perceptual decision making processes, and parameters estimated from the distributions of choice probabilities and reaction times provide deeper insights into choice features among participants (Ratcliff et al., 2009, 2016; Brunton et al., 2013). In recent years, the DDM has been applied to various value-based decision making tasks, such as preferential choice of food items, and it elaborates the neurocognitive mechanisms underlying the choice processes (Milosavljevic et al., 2010; Krajbich and Rangel, 2011; Polanía et al., 2014; Lopez-Persem et al., 2016). Previous studies that employed DDMs have repeatedly reported that the vmPFC is involved in evidence accumulation during value-based choices (Basten et al., 2010; Lim et al., 2011; Hunt et al., 2012; Gherman and Philiastides, 2018). Intriguingly, the contribution of the vmPFC to face preference has also been reported in recent fMRI studies (Lebreton et al., 2009; Ito et al., 2016; Murakami et al., 2018). Thus, a DDM analysis can clarify not only the psychological processes involved in value-based choice but also whether the role of the vmPFC in evidence accumulation can be affected by the own-age bias.

We re-analyzed three datasets from two published and one unpublished fMRI study, all of which employed the same task (Ito et al., 2016; Murakami et al., 2018). The task consisted of a pleasantness-rating task during fMRI scanning and a preferential choice task after fMRI. During the preferential choice task, participants performed a two-alternative forced-choice task, which is frequently used in DDM studies (Krajbich and Rangel, 2011; Lim et al., 2011; Lopez-Persem et al., 2016). First, we applied the DDM to the behavioral data from the preferential choice task and sought to identify parameters that reflect the own-age effect. We then investigated whether the vmPFC showed patterns associated with the own-age bias. Since previous fMRI studies that used DDMs showed the contribution of the functional coupling of the vmPFC and fusiform gyrus toward value representation, we also sought to identify a functional network centered on the vmPFC, which is associated with face preference, and to investigate whether these regions also show an own-age effect.

MATERIALS AND METHODS

Participants

The data of 116 participants from three datasets were included in the present study (study 1: 52 young males, mean age 21.8 years [range, 20–27]; study 2: 16 young females and 16 males, mean age 21.3 years [range, 20–25]; study 3: 16 older females and

16 males, mean age 68.3 years [range, 61–74]). Studies 1 and 2 have been published previously, and both studies included young participants (Ito et al., 2016; Murakami et al., 2018), whereas study 3 was an unpublished study that employed older participants. The older participants underwent mini-mental state examination, in which the most common cutoff scores are 23 and 24 (Mori et al., 1985; Tsoi et al., 2015). The results showed that they had normal cognitive function (min = 25, max = 30; mean score, 28.3 ± 2.0). Four other older participants (three older females and one older male) were excluded from the analysis because they had a cough, experienced technical problems during the experiment, or had asymptomatic infarction. No pathological findings were identified in the brains of the other participants. All participants had normal or corrected-to-normal vision. After the participants received a detailed description of the study, they provided written informed consent in accordance with the tenets of the Declaration of Helsinki. The protocol of study 1 was approved by the Ethical Committee of Hokkaido University, whereas those of studies 2 and 3 were approved by the Ethical Committee of Tohoku University.

Stimuli and Tasks

The same stimulus set was used across the three studies, and details of stimulus preparation are shown in the original report (Ito et al., 2016). The stimulus set comprised the faces of 64 older males, 64 older females, 64 young males, and 64 young female volunteers. A separate group of 13 young volunteers who did not participate in the fMRI study rated 256 facial photographs using a 10-point scale for pleasantness. The mean pleasantness score was ranked within the four stimulus groups (i.e., older males, older females, young males, and young females). Within each stimulus group, the photographs ranked "n" (n = 1–32) were paired with the photographs ranked "n + 32," which resulted in 32 pairs of photographs per group.

The experiment consisted of two tasks: a pleasantness-rating task during fMRI scanning and a preference-choice task after the fMRI scanning. For the pleasantness-rating task during the fMRI scan, each of the 256 face photographs was presented in a random order. Each stimulus was presented for 2.5 s, and the inter-stimulus interval, during which the cross-fixation was constantly presented, ranged between 3.5 and 11.5 s to maximize the efficiency of the event-related design (Dale, 1999). The pleasantness-rating task was divided into four consecutive runs, each lasting approximately 10 minutes. The participants were asked to rate each face based on how pleasant or unpleasant it was (study 1: 8-point scale; studies 2 and 3: 5-point scale). The preference-choice task was performed outside the scanner immediately after the fMRI. The 128 pairs of photographs were displayed as two side-by-side photographs, and the participants were asked to choose the face that they preferred by pressing one of two buttons. The positions of the two photographs within each pair were counterbalanced across participants.

Neuroimaging Data Acquisition

A T2*-weighted echo planar imaging (EPI) sequence sensitive to blood oxygenation level-dependent (BOLD) contrast was used for functional imaging with the following parameters: repetition time = 2,500 ms, echo time = 30 ms, flip angle = 90°, acquisition matrix = 80 × 80, field of view = 240 mm, in-plane resolution = 3 × 3 mm, slice thickness = 3 mm, and interslice gap = 0.5 mm (number of axial slices: 42 for study 1, 43 for studies 2 and 3). An acquisition sequence tilted at 30° to the intercommissural (anterior commissure-posterior commissure) line was used to recover the magnetic susceptibility-induced signal losses due to the sinus cavities (Deichmann et al., 2003). A high-resolution (spatial resolution 1 × 1 × 1 mm) structural image was also acquired using a T1-weighted magnetization-prepared rapid-acquisition gradient echo pulse sequence. Each participant's head motion was restricted using a firm padding that surrounded the head. EPI images were acquired over four consecutive runs. The first four scans in each run were discarded to allow for equilibration effects.

Preprocessing

Preprocessing was performed using the Statistical Parametric Mapping 12 software (Wellcome Department of Imaging Neuroscience, London, United Kingdom). All volumes acquired from each participant were realigned to correct for small movements that occurred between scans. This process generated an aligned set of images and a mean image for each participant. The realigned images were subsequently corrected for different slice acquisition times. Each participant's T1-weighted structural MRI was co-registered to the mean of the realigned EPI images and segmented to separate the gray matter, which was normalized to the gray matter in a template image based on the Montreal Neurological Institute (MNI) reference brain (resampled voxel size, 2 × 2 × 2 mm). Using the parameters from this normalization process, the EPI images were subsequently normalized to the MNI template and smoothed using an 8-mm full-width, half-maximum Gaussian kernel.

Statistical Analysis of the Imaging Data

We employed four generalized linear models (GLMs) to analyze the fMRI data. All GLMs incorporated only one event per trial, which was the onset of face presentation. In GLM 1, pleasantness-rating scores were entered as a parametric regressor which revealed brain regions associated with the representation of subjective pleasantness of faces. Thus, GLM 1 included the raw value of (i.e., not mean-centered) pleasantness scores as parametric regressors. GLM 2 was used to compare brain activity between older and young faces. In GLM 2, trials were sorted based on stimulus age, with no parametric modulation. GLM 3 was applied in the psycho-physiological interaction (PPI) analysis, which identified a network centered on the vmPFC associated with the representation of the subjective value of a faces (i.e., preference). We used the coordinates of the group maximum identified in the parametric modulation analysis (study 1, $x = -2$, $y = 36$, $z = -8$; study 2, $x = 6$, $y = 38$, $z = -8$; study 3, $x = -2$, $y = 44$, $z = -6$). The coordinates served as a starting point for identifying a nearby local maximum in each participant-specific dataset. At the individual level, we identified the local activation peak within a sphere with a radius of 16 mm around the group maximum. We then extracted the first eigenvariate of the BOLD response in each participant within a sphere with a radius of 4 mm around the individual activation peak. Participants who did not demonstrate activation in the seed region at a liberal threshold of $p < 0.05$, uncorrected for multiple comparisons, were excluded from the PPI analysis (study 1, two participants; study 2, eight participants; study 3, nine participants). This GLM included the following three regressors: (1) first eigenvariate of the vmPFC; (2) a regressor specifying a psychological variable which codes preference convolved with canonical HRF; and (3) an interaction term between the two variables. The PPI analysis generated a contrast that represented the regions that exhibited stronger functional connectivity with the vmPFC for the preferred faces than for the non-preferred faces. A group-level random-effects analysis was subsequently performed by applying one-sample t-tests to the first-level t-maps. Events that involved multiple responses or no responses were modeled as events of no interest. A high-pass filter of 1/128 Hz was used to remove low-frequency noise, and an autoregressive (1) model was used to correct for temporal autocorrelations. For all whole-brain analyses, the threshold of significance was set at $p < 0.05$ (family wise error corrected for multiple comparisons at the voxel level), unless otherwise specified. The peak voxels of clusters that exhibited reliable effects were reported in the MNI coordinates. GLM 4, which employed regressors specifying the onset of each face, was used to extract the percent signal change for each stimulus face, and was subsequently applied in the analysis using the DDM as detailed in the next section.

Drift-Diffusion Model

To examine the own-age bias in face preference, we first applied the DDM to our behavioral data. The DDM assumes that the noisy evidence for a decision accumulates over time at a certain speed (drift rate: v), and that the choice is executed when this accumulating evidence crosses one of two boundaries (decision threshold: a). We performed a hierarchical Bayesian estimation of DDM parameters for each participant using the hierarchical DDM (HDDM) 0.6.0 toolbox (Wiecki et al., 2013) implemented in Python 3[1]. If a face that had been rated as more pleasant than the paired face was chosen in the preferential choice task (i.e., consistent choice), the corresponding trial was deemed to be a correct trial. On the other hand, if a face that had been rated as less pleasant than the paired face was chosen (i.e., inconsistent choice), the corresponding trial was considered an erroneous trial. Choice trials that contained two equally rated photographs were excluded from the analysis (study 1: 9.1 ± 2.9%, study 2: 11.2 ± 3.7%, study 3: 15.2 ± 6.2%). To ensure the independence of the estimated parameter, each participant's data were fit separately. Our model had three free parameters: non-decision time (t), decision threshold (a), and drift rate (v). The model was fit to accuracy-coded data (i.e., the upper boundary indicates a correct (consistent) choice, which is predicted by the rating data, while the lower boundary indicates an incorrect (inconsistent) choice, which is not predicted by the rating data), and the starting point was fixed at a/2. The HDDM used Markov chain Monte Carlo sampling to approximate the posterior distribution over parameter estimates. For parameter estimation, three chains were run each with 2,000 samples, and the first 500 samples in each run were discarded as burns in. In addition, Gelman and Rubin's $\widehat{R}$ for each parameter was calculated to assess the convergence. This value is supposed to be close to 1 and not exceed 1.1 if convergence is successful. The mean posterior estimate parameters of each participant were extracted for subsequent statistical tests.

We also applied the HDDM using neuroimaging data to examine whether the areas related to face preference were associated with own-age bias in the preferential choice task. We used MarsBaR software[2] to extract the activity in the regions of interest (ROIs), which was then normalized on a within-participant basis. Percent signal changes were extracted from the vmPFC and visual areas (fusiform and occipital gyri), which showed greater functional coupling with the vmPFC when participants were presented with preferred faces. To avoid the double dipping problem (Kriegeskorte et al., 2009), the percent signal change of the vmPFC was extracted from a spherical mask with a radius of 6 mm centered on the coordinates $x = -10$, $y = 44$, and $z = -8$, which were obtained from a previous study that showed that the vmPFC plays a critical role in the evaluation of faces (Lebreton et al., 2009). Percent signal changes in the visual areas were extracted from the masks of the fusiform face area (FFA) ($x = -37$, $y = -57$, $z = -18$ for the left hemisphere and $x = 40$, $y = -48$, $z = -22$ for the right hemisphere) and occipital face area (OFA) ($x = -36$, $y = -75$, $z = -12$ for the left hemisphere and $x = 40$, $y = -73$, $z = -14$ for the right hemisphere), which have been identified in a previous study (Julian et al., 2012). We applied the HDDMRegressor (Wiecki et al., 2013) to the preference choice task to examine whether the trial-by-trial brain activity in each ROI could modulate parameters that showed significant differences between older and young faces. First, we calculated the difference in percent signal change between chosen and unchosen faces by subtracting the percent signal change of the unchosen face from that of the chosen face. This calculation was performed for each face pair used in the preference choice task. Then, the difference in percent signal change between chosen and unchosen faces in each trial face pair was entered as an explanatory variable for each trial, and HDDMRegressors were estimated. In this analysis, participant IDs were included in the Bayesian hierarchical model. Therefore, the distributions of the DDM parameters, such as the intercept of the drift rate, allowed for individual differences within the group in their posterior distributions. For the regression, three chains were run each with 3,000 samples, and the first 1,000 samples in each run were discarded to improve convergence. The intercept and slope in each ROI for estimating the DDM parameters were calculated for each simulation, and the average values of 6,000 samples were reported. Convergence was assessed using Gelman and Rubin's $\widehat{R}$ for each estimated parameter, including the intercept and slope. Furthermore, to formally test whether the vmPFC is differentially involved in evidence accumulation of the preferential choice for own-age faces, we performed a supplementary HDDMRegressor analysis that modeled trial-by-trial effects as follows: drift rate $\sim$ vmPFC activity + face

[1]https://www.python.org/

[2]http://marsbar.sourceforge.net/

age + vmPFC activity × face age. Here, both own-age and other-age face trials were included in a single model for each study (face age was dummy coded: 1 for own-age and −1 for other-age face), and the own-age bias (i.e., a greater slope for own-age faces relative to other-age faces) was operationalized as a positive regression coefficient for the interaction term. All other analytical settings were identical to those of the aforementioned HDDMRegressor analysis. In addition, we aggregated the results across studies using the *meta* package in the R environment (Schwarzer et al., 2015). We used the mean values of the estimated regression coefficients indicating the own-age bias and standard deviations of the three studies (i.e., the outputs of the HDDMRegressor analysis) to compute the pooled mean by considering the between-study variance (similar to a random-effect meta-analysis). This analysis was performed *post hoc* after observing the statistical results of single studies (i.e., the results should be viewed as a basis for future research rather than for drawing conclusions from the present study). Also, the data used in the three studies are all available as datasets collected by us with regard to testing the neural own-age bias (i.e., there was no selection bias or intentional stopping during data collection).

RESULTS

Behavioral Data

We first investigated whether the rating scores for own-age faces were higher than those for other-age faces using paired t-tests (**Figure 1**). In studies 1 and 2, the rating scores for own-age faces were significantly higher than those for other-age faces (study 1, $t(51) = -2.35$, $p = 0.02$, $d = -0.33$, 95% confidence interval [CI] [−0.52, −0.04]; study 2, $t(31) = -2.37$, $p = 0.02$, $d = -0.42$, 95% CI [−0.43, −0.03]). On the other hand, the rating scores for other-age faces were significantly higher than those for own-age faces in study 3 ($t(31) = -4.11$, $p < 0.001$, $d = -0.73$, 95% CI [−0.41, −0.14]). These results suggest that young faces are perceived as more pleasant than older face.

Next, for the results of the DDM, we investigated whether non-decision time (t), decision threshold (a), and drift rate (v) for own-age faces were significantly higher than those for other-age faces using paired t-tests (**Figure 2**). We found that the drift rates for own-age faces were significantly higher than those for other-age faces in studies 1 and 2 (study 1, $t(51) = -8.38$, $p < 0.001$, $d = -1.16$, 95% CI [−0.37, −0.23]; study 2, $t(31) = -8.02$, $p < 0.001$, $d = -1.42$, 95% CI [−0.53, −0.32]), but study 3 showed that the drift rates for other-age faces were significantly higher than those for own-age faces ($t(31) = -4.76$, $p < 0.001$, $d = -0.84$, 95% CI [−0.49, −0.20]) (**Figure 2C**). These results suggest that the drift rates for young faces were higher than those for older faces, regardless of the participants' groups. We found no significant differences in the non-decision time and decision threshold (**Figures 2A,B**, all p-Values > 0.1). The ranges of the $\widehat{R}$ value in all parameter estimates indicated satisfactory convergence (study 1: 0.999–1.007, study 2: 0.999–1.007, and study 3: 0.999–1.019).

Imaging Data

First, we performed a parametric modulation analysis (**Figure 3**). The results of studies 1 and 2 have been reported previously (Ito et al., 2016; Murakami et al., 2018) and have been displayed here only for display purposes. Consistent with the results of studies 1 and 2, the results of study 3 showed a significant positive correlation between the pleasantness-rating score and the activity in the vmPFC ($x = -2$, $y = 44$, $z = -6$, z-value = 5.02, $k = 9$; $x = -8$, $y = 30$, $z = -10$, z-value = 4.75, $k = 1$) (**Figure 3**). Direct comparison between studies 2 and 3 showed no significant difference, indicating no age-related decline with regard to the representation of the subjective pleasantness of faces (**Supplementary Results**).

In the PPI analysis, several brain regions, including structures within the ventral visual area, such as the fusiform and occipital gyri, showed greater functional connectivity with the vmPFC (**Figure 4**, **Table 1**, and **Supplementary Table 2**). Across the three studies, we identified clusters that overlapped with the FFA and OFA, as shown in previous study by Julian et al. (2012) and Neurosynth (**Supplementary Figure 1**). Direct comparison between studies 2 and 3 showed no significant difference, suggesting little evidence of age-related change of functional coupling between the vmPFC and ventral visual area with respect to value representation (**Supplementary Results**).

We then performed a subtraction analysis to identify areas that showed greater activity for either own-age or other-age faces (**Table 2** and **Supplementary Table 3**). Both young and older participants showed greater activation for own-age faces. For example, young participants showed greater activity in reward-related regions such as the vmPFC and ventral striatum, and older participants showed greater activity in the superior occipital gyrus.

Hierarchical Drift-Diffusion Model Regression

Since the drift rate for young faces was significantly higher than that for older faces among the three studies, we applied an HDDMRegressor analysis to examine whether the trial-by-trial activity of the ROI could explain the drift rate for each dataset. For each ROI, the difference in percent signal change between chosen and unchosen faces (percent signal change for the chosen face − percentage signal change for the unchosen faces) in each trial was entered as the explanatory variable. A significant slope value for the explanatory variable implied that the signal of the ROI could explain the drift rate. The results are summarized in **Table 3**. Across the three studies, the slope value of the vmPFC was significantly positive for own-age faces. In addition, the results in study 1 showed a significantly positive slope value for the occipital gyrus among other-age faces. In study 2, the results also showed a significantly positive slope value for the fusiform and occipital gyrus among own-age faces and that for the vmPFC among other-age faces. In study 3, the results also showed a significantly positive slope value for the fusiform gyrus among other-age faces. The ranges of $\widehat{R}$ values for all parameter estimates indicated satisfactory convergence (study 1: 0.9996–1.0027, study 2: 0.9997–1.0004, and study 3: 0.9997–1.0006). To

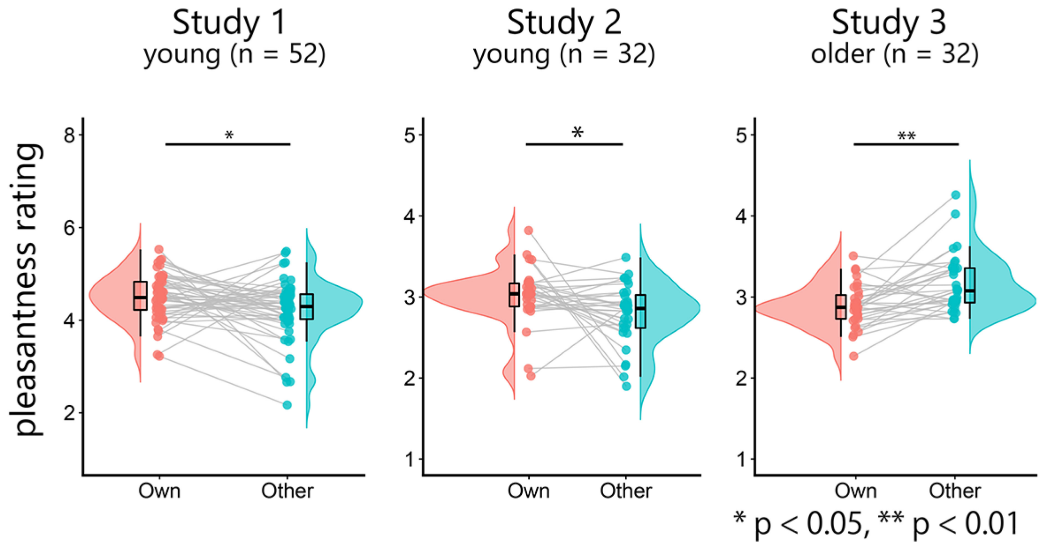


FIGURE 1 | Raincloud plots of the mean pleasantness-rating scores for own-age and other-age faces. Note that study 1 used a different scale (1–8) compared to the other two studies (1–5).

formally assess whether the vmPFC showed own-age bias (i.e., a greater slope for own-age faces relative to other-age faces), we performed a follow-up HDDMRegressor analysis that modeled an interaction term of vmPFC activity × face age. As expected, the regression coefficients for the interaction term were consistently positive across all three studies (study 1, $\beta = 0.0129$, $p = 0.0665$, 95% CI [–0.004, 0.029]; study 2, $\beta = 0.006$, $p = 0.290$, 95% CI [–0.016, 0.029]; study 3, $\beta = 0.029$, $p = 0.023$, 95% CI [0.0004, 0.057]). These results suggested trends toward the own-age bias in the trial-by-trial effect of vmPFC activity on the drift rate. A *post hoc* aggregation of results across the three studies indicated that the pooled mean of the regression coefficient was positive (mean = 0.0158, 95% CI [0.006, 0.026], between-study variance $\tau^2 = 7.81 \times 10^{-5}$). These results suggested that the vmPFC has greater role in the preferential choice of own-age faces. It should be noted that we did not find significant effects in other regions (left OFA, mean = 0.0017, 95% CI [–0.007, 0.01], between-study variance $\tau^2 = 5.36 \times 10^{-5}$; right OFA, mean = –0.003, 95% CI [–0.0224, 0.016], between-study variance $\tau^2 = 2.88 \times 10^{-4}$; left FFA, mean = –0.0038, 95% CI [–0.02, 0.013], between-study variance $\tau^2 = 2.14 \times 10^{-4}$; right FFA, mean = 0.001, 95% CI [–0.017, 0.019], between-study variance $\tau^2 = 2.55 \times 10^{-4}$). The supplementary analysis combining data from studies 2 and 3 also supports this notion (see **Supplementary Results** for details).

DISCUSSION

Regardless of the age of the participants, the behavioral results demonstrated higher rating scores and higher drift rates for young faces than for older faces. These behavioral results showed no evidence of own-age bias in face preference, suggesting the important role of youth in face preference. Although neuroimaging results from the three datasets suggest a possibility that the vmPFC was associated with evidence of accumulation of own-age faces, results from study 1 did not reach significance and no robust evidence was provided. Importantly, we found no age-related decline in vmPFC responsiveness to subjective pleasantness of faces. The results of the present study suggest that preferential choice of face is less susceptible to own-age bias across the life span of individuals.

Across the three studies, the behavioral results consistently showed a higher drift rate for young faces, suggesting the role of youth rather than own-age bias. The supplemental analysis showed that the difference in rating scores of the young face pairs, as calculated by subtracting the score for face B from that for face A, was significantly larger than that of the older face pairs (**Supplementary Figure 2**). Thus, the larger difference in the subjective value between items in the young pairs which is originated from the nature that the young face is more pleasant than older face was related to this young-age effect. Another possibility is that the difference of salience in young face pairs was larger than that of the older face pairs, which might make it easier to make preferential choices among young face pairs. Consistent with this idea, differences in salience between young and older faces have been implicated in previous psychological studies (Ebner, 2008), and this idea is in line with the previously established evidence that salience benefits value-based choice (Towal et al., 2013; Chen et al., 2021). It is plausible that features closely linked to young faces, such as facial fullness and smooth and clear skin, are associated with these behavioral patterns (Buss and Barnes, 1986; Coleman and Grover, 2006; Popenko et al., 2017; Sakano et al., 2021).

Although the current findings are inconclusive and future studies are required to corroborate them, the overall pattern of the HDDM regression analysis suggests that the vmPFC exhibits an own-age bias in evidence accumulation. Since the ROI of the vmPFC was obtained from a previous study that

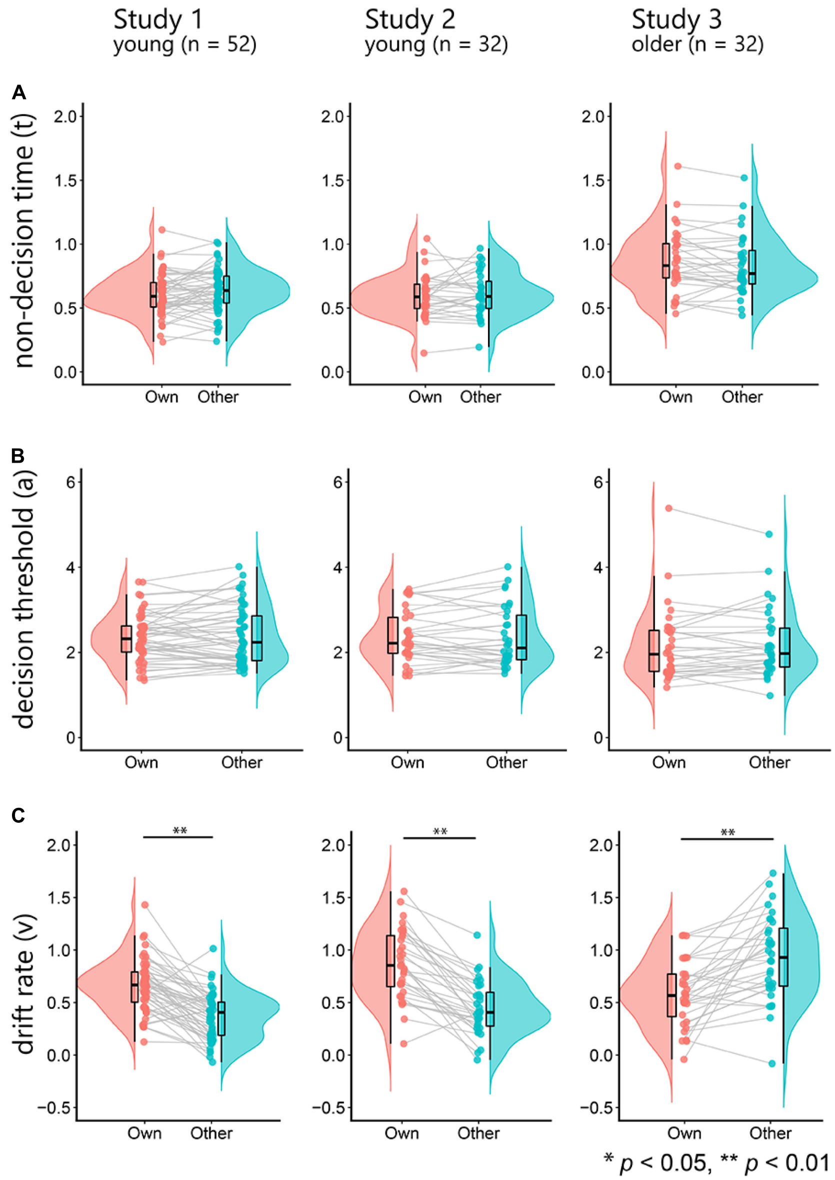


FIGURE 2 | The top row depicts raincloud plots of the non-decision time (*t*) for own-age and other-age faces **(A)**. The left panel shows the results of study 1, the middle panel shows the results of study 2, and the right panel shows the results of study 3. The middle row depicts the decision threshold (*a*) for own-age and other-age faces **(B)**. The bottom row depicts the drift rate (*v*) for own-age and other-age faces **(C)**.

showed that the vmPFC played a role in preference-related value representation (Lebreton et al., 2009), a straightforward interpretation is that the preferential choice of own-age faces mainly relies on the comparison of the subjective values, whereas the preferential choice of other-age faces also relies on other factors. This view is supported by previous literature that argued that various factors such as familiarity, preference, salience, or more available prototypes may underlie the differences in the processing of in-group and out-group faces (Ebner et al., 2011b, 2013).

Value-based choice consists of two stages: valuation and selection (Rangel et al., 2008; Kable and Glimcher, 2009). Previous research focusing on the neural substrates of the two-stage processes demonstrated that the initial valuation process is supported by the vmPFC, and that the subsequent selection processes (i.e., integration and read-out) are supported by the dorsolateral prefrontal cortex and posterior parietal cortex, respectively (Domenech et al., 2017). Thus, it is also possible that the perceptual expertise of own-age faces enables the vmPFC to work solely on the valuation process, whereas participants showed less perceptual expertise for other-age faces, which required other regions such as the fusiform and occipital gyri to support the vmPFC. It should be noted that the vmPFC activity applied in the HDDM regression analysis was assessed during

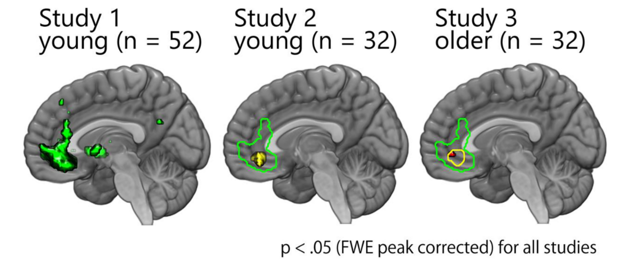


FIGURE 3 | Activity of the overlapping cluster within the vmPFC was significantly correlated with the pleasantness-rating scores across the three studies. The results of studies 1 and 2 have been reported previously (Ito et al., 2016; Murakami et al., 2018) and are presented here only for display purposes. Consistent with these results, there was a significant correlation in the vmPFC among older participants.

the pleasantness-rating task but not during the choice task. Thus, the differential contributions of the vmPFC may be due to the different processes in the valuation stage. Future studies assessing vmPFC activity during both the rating task and choice task are warranted to formally investigate the role of the vmPFC in valuation and subsequent selection processes.

Importantly, regardless of the age of the participants, we found that the vmPFC parametrically represented the subjective value of faces, and showed stronger functional connectivity with visual areas when participants were presented with faces that they preferred. It should be noted that direct group comparisons between the young participants (study 2) and older participants (study 3) showed no significant difference (**Supplementary Results**). These results suggest that the function of computing subjective values of faces can be maintained in older participants, and that the own-age effect identified in the vmPFC was not attributable to age-related functional decline. Recent reports employing a mixed-lottery choice task that required participants to calculate the expected value of the options have indicated the influence of aging on value-based decision making (Su et al., 2018; Chen et al., 2021), which is considered to be followed by a decline in dopaminergic modulation and fronto-striatal network functioning (Samanez-Larkin and Knutson, 2015). Compared to such processes, the preference-related value representation of faces appears to be less susceptible to aging.

The functional network of the vmPFC and ventral visual stream comprises a part of an extended face processing system in the human brain (Haxby et al., 2000; Elbich et al., 2019). Previous studies have revealed the contribution of the ventral visual stream to facial attractiveness (Kranz and Ishai, 2006; Winston et al., 2007; Chatterjee et al., 2009; Pegors et al., 2015). More directly, another fMRI study demonstrated a functional coupling between the vmPFC and fusiform gyrus in representing the subjective value of a T-shirt (Lim et al., 2013). Taken together with the function of the vmPFC in modality- and category-independent value representation (Chib et al., 2009; Lebreton et al., 2009; Ishizu and Zeki, 2011; McNamee et al., 2013), the representation of the subjective value of a face may be supported by two valuation systems: the BVS which represents subjective value in a domain-general manner subserved by the vmPFC (Lebreton et al., 2009) and the extended BVS which represents subjective value as a domain-specific manner subserved by the functional coupling between the BVS and sensory area.

The function of the BVS is to reflect both the value, that is explicitly revealed by participants (e.g., rating), and preference, which is revealed in binary choices. The BVS often includes reward-related regions, such as the vmPFC and ventral striatum (Lebreton et al., 2009; Ito et al., 2016; Murakami et al., 2018). On the other hand, the role of the extended BVS is to serve as a functional network between the BVS and sensory areas, which are involved in the representation of subjective values. Thus, we argue that the functional network among the vmPFC, fusiform gyrus, and occipital gyrus identified in this study meets this requirement and can be considered as the extended BVS for face preference. Importantly, based on recent evidence showing functional coupling of the vmPFC and the fusiform gyrus in value computation of brand names (Zhang et al., 2021), the role of the extended BVS might not be limited to the face. In fact, a previous study that employed music as a stimulus identified an extended BVS for auditory preference (Salimpoor et al., 2013). The researchers found a functional coupling between the ventral striatum and the auditory cortex, which represented subjective values of musical excerpts. A recent understanding of

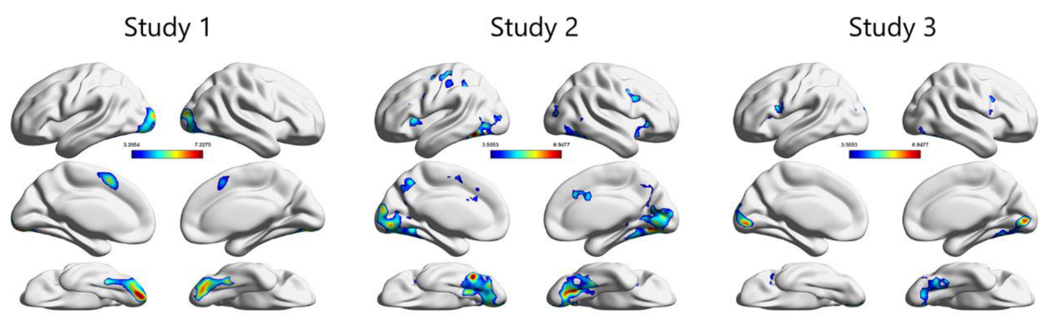


FIGURE 4 | Across the three studies, the psycho-physiological interaction (PPI) analysis revealed that structures in the ventral visual areas, including those in the fusiform and occipital areas, demonstrated greater functional connectivity with the vmPFC for preferred faces compared with non-preferred faces. Here, the threshold of significance was set at $p < 0.05$ at the cluster level (FWE corrected for multiple comparisons) for display purposes (see **Supplementary Table 2** for more information). The original statistical threshold was set at $p < 0.05$ at the peak level (FWE corrected for multiple comparisons) and the results are shown in **Table 1**.

TABLE 1 | Brain regions showing greater functional connectivity (preferred vs. non-preferred faces).

Region	Coordinates			Z-value	Cluster size
	x	y	z		
Study 1					
Left inferior occipital gyrus	−22	−86	−8	5.93	564
Right inferior occipital gyrus	18	−88	−4	5.64	415
Left fusiform gyrus	−36	−58	−14	4.96	33
Right fusiform gyrus	38	−48	−16	5.27	38
left SMA	−6	12	48	4.88	29
Left middle frontal gyrus	−44	2	30	4.58	5
Study 2					
Bilateral occipital gyrus	0	−96	6	5.27	8
Right fusiform gyrus	26	−74	−8	5.26	31
Left fusiform gyrus	−42	−58	−14	5.23	30
Right orbitofrontal cortex	28	30	−12	4.85	1
Left fusiform gyrus	−34	−80	−12	4.83	1
Study 3					
Right occipital gyrus	10	−86	2	5.00	23
Left occipital gyrus	−8	−88	−2	4.81	3
Left occipital gyrus	−14	−98	18	4.81	2

The threshold of significance was set at $p < 0.05$ at the peak level (FWE corrected).

TABLE 2 | Brain regions showing significant activation for own-age faces and other-age faces.

Region (Brodmann's Area)	Coordinates			Z-value	Cluster size
	x	y	z		
Study 1					
Own-age face vs. other-age face					
Left occipital gyrus	−26	−92	0	5.99	305
Left orbitofrontal cortex (extending to operculum)	−28	26	−14	5.67	49
Left brain stem (extending to the right side)	−4	−26	−10	5.57	71
Left thalamus (extending to right hemisphere)	−2	−18	8	5.44	101
Left inferior frontal gyrus	−40	18	22	5.18	40
Left fusiform gyrus	−44	−60	−12	5.04	59
Left anterior insula	−32	18	0	4.9	9
Left ventral striatum	−6	2	−6	4.77	2
Left inferior occipital gyrus	−42	−68	−4	4.64	1
Right fusiform gyrus	44	−52	−14	6.8	828
Right superior occipital gyrus	30	−74	28	5.38	110
Right orbitofrontal cortex (extending to operculum)	36	28	−2	5.11	99
Right middle cingulate gyrus	4	6	28	4.86	13
Right fusiform gyrus	48	−38	−12	4.83	1
Right ventral striatum	12	4	0	4.81	4
Right fusiform gyrus	42	−36	−14	4.78	2
Right cerebellum	8	−80	−28	4.63	1
Other-age face vs. own-age face					
no suprathreshold activation					
Study 2					
Own-age face vs. other-age face					
Left superior frontal gyrus	−16	40	50	4.79	5
Right orbitofrontal cortex	32	28	−16	5.13	6
Right orbitofrontal cortex	38	32	−12	4.7	2
Other-age face vs. own-age face					
no suprathreshold activation					
Study 3					
Own-age face vs. other-age face					
Right cuneus	12	−90	14	5.39	39
Right calcarine sulcus	14	−70	18	5.04	7
Right superior occipital gyrus	20	−92	22	4.85	6
Other-age face vs. own-age face					
no suprathreshold activation					

The threshold of significance was set at $p < 0.05$ at the peak level (FWE corrected).

how subjective value was formed through bottom-up processing (Iigaya et al., 2020; Pham et al., 2021) indicates that the extended BVS might serve as a gateway for domain-specific value computation or attribute value computation (Lim et al., 2013).

A natural interpretation of the higher pleasantness-rating scores for young faces found in both age groups is that young faces are more rewarding than older faces, regardless of participants' age. Previous research which proposed the three-dimensional model of social inference from faces showed that the 'youthful-attractiveness' factor consistently emerged from an unconstrained set of face stimuli (Sutherland et al., 2013). The role of youthfulness in impression formation and social interaction has long been advocated in previous seminal reviews (Buss and Schmitt, 1993; Rhodes, 2006). Supporting these findings, evidence from the field of esthetic surgery revealed age-related facial structural changes, such as progressive bone resorption, decreased tissue elasticity, and loss of facial fullness (Coleman and Grover, 2006), and indicated that the fullness of the face is directly linked to attractiveness (Popenko et al., 2017). Facial pleasantness appears to be independent of perceptual expertise or familiarity with one's own age group.

The results of the whole-brain analysis showed that own-age faces elicited stronger activation than other-age faces, but not vice versa. One possibility is that the stronger activation for own-age faces might reflect a differential gaze pattern to own-age and other-age faces. This view is supported by a previous eye-tracking study that revealed that people see own-age faces longer than other-age faces (Ebner et al., 2011c; He et al., 2011). Perhaps these differential gaze patterns might reflect perceptual expertise or familiarity with own-age faces acquired through recent experiences with one's own age group (Chiroro and Valentine, 1995; Anastasi and Rhodes, 2005; Rhodes and Anastasi, 2012). However, although we found greater activity of the visual area for own-age faces in studies 1 and 3, this finding was not replicated in study 2 and further investigations are required in this regard. Furthermore, we did not collect eye-tracking data in the present study. Future studies that combine neuroimaging techniques with eye tracking could provide direct evidence that disentangles causes related to own-age-specific brain activity. Alternatively, subjective similarity to one's own-age faces might be related to differential activity. Ebner et al. (2011a) found the contribution of the vmPFC to own-age face processing (Ebner et al., 2011a),

TABLE 3 | Results of the hierarchical drift-diffusion model analysis.

Region	Own-age faces				Other-age faces			
	slope	*p-Value*	*credible interval*	*DIC*	*slope*	*p-Value*	*credible interval*	*DIC*
Study 1								
vmPFC	0.028	0.02*	0.003 0.052	7531.28	0.000	0.49	–0.024 0.024	8174.66
L fusiform	0.007	0.29	–0.018 0.032	7535.64	0.013	0.16	–0.014 0.036	8174.39
R fusiform	0.013	0.15	–0.012 0.04	7534.04	0.009	0.24	–0.016 0.034	8176.85
L occipital	0.004	0.38	–0.022 0.029	7536.52	0.011	0.19	–0.012 0.039	8176.17
R occipital	–0.004	0.39	–0.029 0.022	7535.30	0.025	0.02*	0.001 0.051	8173.16
Study 2								
vmPFC	0.051	0.001**	0.019 0.086	3824.80	0.037	0.01*	0.006 0.068	4436.40
L fusiform	0.027	0.05	–0.007 0.06	3830.81	0.000	0.50	–0.032 0.033	4442.82
R fusiform	0.034	0.03*	0.001 0.069	3829.01	–0.005	0.40	–0.038 0.029	4441.92
L occipital	0.023	0.09	–0.011 0.057	3831.20	0.005	0.39	–0.028 0.037	4440.96
R occipital	0.040	0.01*	0.006 0.076	3826.66	0.011	0.26	–0.021 0.042	4440.60
Study 3								
vmPFC	0.055	0.004**	0.014 0.095	2876.76	–0.007	0.37	–0.049 0.036	2585.06
L fusiform	–0.004	0.42	–0.046 0.037	2881.78	0.028	0.09	–0.012 0.069	2580.97
R fusiform	–0.013	0.25	–0.049 0.024	2883.79	0.021	0.17	–0.021 0.064	2583.41
L occipital	0.012	0.27	–0.026 0.049	2881.22	0.016	0.22	–0.024 0.058	2585.03
R occipital	0.022	0.12	–0.016 0.058	2880.26	0.035	0.04*	0.006 0.077	2582.20

$^*p < 0.05$ and $^{**}p < 0.01$; L, left; R, right.

supporting the previous evidence that the vmPFC shows higher activation when thinking about similar than dissimilar others (Amodio and Frith, 2006; Mitchell et al., 2006; Van Overwalle, 2009). Given its contribution to various aspects of social cognition, such as emotional empathy, social reputation, moral judgment, and value-based decision making (Saxe, 2006; Forbes and Grafman, 2010; Ito et al., 2011, 2015; Delgado et al., 2016; Kawasaki et al., 2016; Hiser and Koenigs, 2018; Yoon et al., 2018; Suzuki and O'Doherty, 2020; Pessiglione and Daunizeau, 2021) as well as its multifaceted anatomical connections (Rolls and Grabenhorst, 2008; Grabenhorst and Rolls, 2011; Riga et al., 2014), dissociable neural populations would contribute to own-age face recognition and facial preference.

DATA AVAILABILITY STATEMENT

The 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 Ethical Committee of Hokkaido University and the Ethical Committee of Tohoku University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AI, KY, and RA contributed to conception, design of the study, and performed the statistical analysis. AI, TF, IK, AH, AU, SS, SM, and ST collected the data. AI organized the database. AI wrote the first draft of the manuscript. KY and RA wrote sections of the manuscript. EM supervised the study. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was partly supported by KOSE Cosmetology Research Foundation.

ACKNOWLEDGMENTS

We are grateful to Tatsuo Nagasaka, Kazuomi Yamanaka, Hironobu Sasaki, Yohei Inaba, Tomoyoshi Kimura, Hisanobu Fukuda, Hitoshi Nemoto, Satoshi Sakuraba, and Ayumi Seo for their assistance in data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.822234/full#supplementary-material

Supplementary Figure 1 | Overlapping areas (shown in pink) between significant clusters identified in the PPI analysis of each study (studies 1–3, from the top to the bottom, shown in red), clusters of the FFA and OFA reported by Julian et al. (2012) (left panels), and clusters of the FFA and OFA identified in the "ffa" Z-maps obtained from Neurosynth (right panels) (shown in blue). Study 1 **(a,b)** and study 2 **(c,d)** revealed an overlap in both hemispheres, and study 3 **(e,f)** showed overlap in the right hemisphere. Here, the threshold of significance was set at $p < 0.05$ (cluster-level FWE corrected).

Supplementary Figure 2 | Differences in rating scores among own-age face pairs, as calculated by subtracting scores for face B from those of face A and those of the other-age face pairs.

Supplementary Figure 3 | From top to bottom: study 1 **(a)**, study 2 **(b)**, and study 3 **(c)**. From left to right: non-decision time (*t*), decision threshold **(a)**, and drift rate (v). In study 1, parameters were calculated based on stimulus gender (female and male) and stimulus age (own and other). In studies 2 and 3, parameters were calculated based on participant gender (F and M), stimulus gender (female and male), and stimulus age (own and other). F, female participants; M, male participants.

REFERENCES

Amodio, D. M., and Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277. doi: 10.1038/nrn1884

Anastasi, J. S., and Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychon. B Rev.* 12, 1043–1047. doi: 10.3758/Bf03206441

Armel, K. C., Beaumel, A., and Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judg. Decision Mak.* 3:8.

Basten, U., Biele, G., Heekeren, H. R., and Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21767–21772. doi: 10.1073/pnas.0908104107

Berggren, N., Jordahl, H., and Poutvaara, P. (2010). The looks of a winner: beauty and electoral success. *J. Public Econ.* 94, 8–15. doi: 10.1016/j.jpubeco.2009.11.002

Bernstein, M. J., Young, S. G., and Hugenberg, K. (2007). The cross-category effect: mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychol. Sci.* 18, 706–712. doi: 10.1111/j.1467-9280.2007.01964.x

Brunton, B. W., Botvinick, M. M., and Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science* 340, 95–98. doi: 10.1126/science.1233912

Buss, D. M. (1985). Human mate selection. *Am. Sci.* 73, 47–51.

Buss, D. M., and Barnes, M. (1986). Preferences in human mate selection. *J. Pers. Soc. Psychol.* 50, 559–570. doi: 10.1037//0022-3514.50.3.559

Buss, D. M., and Schmitt, D. P. (1993). Sexual strategies theory - an evolutionary perspective on human mating. *Psychol. Rev.* 100, 204–232. doi: 10.1037/0033-295x.100.2.204

Chatterjee, A., Thomas, A., Smith, S. E., and Aguirre, G. K. (2009). The neural response to facial attractiveness. *Neuropsychology* 23, 135–143. doi: 10.1037/a0014430

Chen, H.-Y., Dix, A., Goh, J. O. S., Smolka, M. N., Thurm, F., and Li, S.-C. (2021). Effects and mechanisms of information saliency in enhancing value-based decision-making in younger and older adults. *Neurobiol. Aging* 99, 86–98. doi: 10.1016/j.neurobiolaging.2020.11.018

Chib, V. S., Rangel, A., Shimojo, S., and O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J. Neurosci.* 29, 12315–12320. doi: 10.1523/JNEUROSCI.2575-09.2009

Chiroro, P., and Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *Q. J. Exp. Psychol. A* 48, 879–894. doi: 10.1080/14640749508401421

Coleman, S., and Grover, R. (2006). The anatomy of the aging face: volume loss and changes in 3-dimensional topography. *Aesth. Surg. J.* 26, S4–S9. doi: 10.1016/j.asj.2005.09.012

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Hum. Brain Mapp.* 8, 109–114. doi: 10.1002/(SICI)1097-0193(1999)8:2/3<109::AID-HBM7>3.0.CO;2-W

Deichmann, R., Gottfried, J. A., Hutton, C., and Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. *NeuroImage* 19, 430–441.

Delgado, M. R., Beer, J. S., Fellows, L. K., Huettel, S. A., Platt, M. L., Quirk, G. J., et al. (2016). Viewpoints: dialogues on the functional role of the ventromedial prefrontal cortex. *Nat. Neurosci.* 19, 1545–1552. doi: 10.1038/nn.4438

Domenech, P., Redouté, J., Koechlin, E., and Dreher, J.-C. (2017). The neuro-computational architecture of value-based selection in the human brain. *Cereb. Cortex* 28, 585–601. doi: 10.1093/cercor/bhw396

Ebner, N. C. (2008). Age of face matters: age-group differences in ratings of young and old faces. *Behav. Res.* 40, 130–136. doi: 10.3758/BRM.40.1.130

Ebner, N. C., He, Y., and Johnson, M. K. (2011c). Age and emotion affect how we look at a face: visual scan patterns differ for own-age versus other-age emotional faces. *Cogn. Emot.* 25, 983–997. doi: 10.1080/02699931.2010.540817

Ebner, N. C., He, Y., Fichtenholtz, H. M., McCarthy, G., and Johnson, M. K. (2011b). Electrophysiological correlates of processing faces of younger and older individuals. *Soc. Cogn. Affect. Neurosci.* 6, 526–535.

Ebner, N. C., Gluth, S., Johnson, M. R., Raye, C. L., Mitchell, K. J., and Johnson, M. K. (2011a). Medial prefrontal cortex activity when thinking about others depends on their age. *Neurocase* 17, 260–269. doi: 10.1080/13554794.2010.536953

Ebner, N. C., and Johnson, M. K. (2009). Young and older emotional faces: are there age-group differences in expression identification and memory? *Emotion* 9, 329–339. doi: 10.1037/a0015179

Ebner, N. C., Johnson, M. R., Rieckmann, A., Durbin, K. A., Johnson, M. K., and Fischer, H. (2013). Processing own-age vs. other-age faces: neuro-behavioral correlates and effects of emotion. *Neuroimage* 78, 363–371. doi: 10.1016/j.neuroimage.2013.04.029

Elbich, D. B., Molenaar, P. C. M., and Scherf, K. S. (2019). Evaluating the organizational structure and specificity of network topology within the face processing system. *Hum. Brain Mapp.* 40, 2581–2595. doi: 10.1002/hbm.24546

Forbes, C. E., and Grafman, J. (2010). The role of the human prefrontal cortex in social cognition and moral judgment. *Annu. Rev. Neurosci.* 33, 299–324. doi: 10.1146/annurev-neuro-060909-153230

Gherman, S., and Philiastides, M. G. (2018). Human VMPFC encodes early signatures of confidence in perceptual decisions. *eLife* 7:e38293. doi: 10.7554/eLife.38293

Grabenhorst, F., and Rolls, E. T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends Cogn. Sci.* 15, 56–67. doi: 10.1016/j.tics.2010.12.004

Hare, T. A., Camerer, C. F., Knoepfle, D. T., and Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J. Neurosci.* 30, 583–590. doi: 10.1523/JNEUROSCI.4089-09.2010

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

He, Y., Ebner, N. C., and Johnson, M. K. (2011). What predicts the own-age bias in face recognition memory? *Soc. Cogn.* 29, 97–109. doi: 10.1521/soco.2011.29.1.97

Hiser, J., and Koenigs, M. (2018). The multifaceted role of the ventromedial prefrontal cortex in emotion, decision making, social cognition, and psychopathology. *Biol. Psychiatry* 83, 638–647. doi: 10.1016/j.biopsych.2017.10.030

Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F., and Behrens, T. E. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nat. Neurosci.* 15, 470–S3. doi: 10.1038/nn.3017

Iigaya, K., Yi, S., Wahle, I. A., Tanwisuth, K., and O'Doherty, J. P. (2020). Aesthetic preference for art emerges from a weighted integration over hierarchically structured visual features in the brain. *Neuroscience* [Epub ahead of print]. doi: 10.1101/2020.02.09.940353

Ishizu, T., and Zeki, S. (2011). Toward a brain-based theory of beauty. *PLoS One* 6:e21852. doi: 10.1371/journal.pone.0021852

Ito, A., Abe, N., Kawachi, Y., Kawasaki, I., Ueno, A., Yoshida, K., et al. (2015). Distinct neural correlates of the preference-related valuation of supraliminally and subliminally presented faces. *Hum. Brain Mapp.* 36, 2865–2877. doi: 10.1002/hbm.22813

Ito, A., Fujii, T., Abe, N., Kawasaki, I., Hayashi, A., Ueno, A., et al. (2016). Gender differences in ventromedial prefrontal cortex activity associated with valuation of faces. *Neuroscience* 328, 194–200. doi: 10.1016/j.neuroscience.2016.04.047

Ito, A., Fujii, T., Ueno, A., Koseki, Y., Tashiro, M., and Mori, E. (2011). Neural basis of pleasant and unpleasant emotions induced by social reputation. *Neuroreport* 22, 679–683. doi: 10.1097/WNR.0b013e32834a20b8

Julian, J. B., Fedorenko, E., Webster, J., and Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage* 60, 2357–2364. doi: 10.1016/j.neuroimage.2012.02.055

Kable, J. W., and Glimcher, P. W. (2009). The neurobiology of decision: consensus and controversy. *Neuron* 63, 733–745. doi: 10.1016/j.neuron.2009.09.003

Kawasaki, I., Ito, A., Fujii, T., Ueno, A., Yoshida, K., Sakai, S., et al. (2016). Differential activation of the ventromedial prefrontal cortex between male and female givers of social reputation. *Neurosci. Res.* 103, 27–33. doi: 10.1016/j.neures.2015.07.010

Korthase, K. M., and Trenholme, I. (1982). Perceived age and perceived physical attractiveness. *Percept. Motor Skills* 54, 1251–1258. doi: 10.2466/pms.1982.54.3c.1251

Krajbich, I., and Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13852–13857. doi: 10.1073/pnas.1101328108

Kranz, F., and Ishai, A. (2006). Face perception is modulated by sexual preference. *Curr. Biol.* 16, 63–68. doi: 10.1016/j.cub.2005.10.070

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., and Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* 12, 535–540. doi: 10.1038/nn.2303

Lebreton, M., Jorge, S., Michel, V., Thirion, B., and Pessiglione, M. (2009). An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron* 64, 431–439. doi: 10.1016/j.neuron.2009.09.040

Levy, D. J., and Glimcher, P. W. (2011). Comparing apples and oranges: using reward-specific and reward-general subjective value representation in the brain. *J. Neurosci.* 31, 14693–14707. doi: 10.1523/JNEUROSCI.2218-11.2011

Lim, S. L., O'Doherty, J. P., and Rangel, A. (2011). The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. *J. Neurosci.* 31, 13214–13223. doi: 10.1523/JNEUROSCI.1246-11.2011

Lim, S.-L., O'Doherty, J. P., and Rangel, A. (2013). Stimulus value signals in ventromedial PFC reflect the integration of attribute value signals computed in fusiform gyrus and posterior superior temporal gyrus. *J. Neurosci.* 33, 8729–8741. doi: 10.1523/JNEUROSCI.4809-12.2013

Lopez-Persem, A., Domenech, P., and Pessiglione, M. (2016). How prior preferences determine decision-making frames and biases in the human brain. *eLife* 5:e20317. doi: 10.7554/eLife.20317

Macchi Cassia, V., Pisacane, A., and Gava, L. (2012). No own-age bias in 3-year-old children: more evidence for the role of early experience in building face-processing biases. *J. Exp. Child Psychol.* 113, 372–382. doi: 10.1016/j.jecp.2012.06.014

McLellan, B., and Mckelvie, S. (1993). Effects of age and gender on perceived facial attractiveness. *Can. J. Behav. Sci.* 25, 135–142. doi: 10.1037/h0078790

McNamee, D., Rangel, A., and O'Doherty, J. P. (2013). Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nat. Neurosci.* 16, 479–485. doi: 10.1038/nn.3337

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., and Rangel, A. (2010). The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judg. Decis. Making* 5, 437–449.

Mitchell, J. P., Macrae, C. N., and Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 50, 655–663. doi: 10.1016/j.neuron.2006.03.040

Mori, E., Mitani, Y., and Yamadori, A. (1985). Usefulness of a Japanese version of the Mini-Mental State Test in neurological patients. *Jpn. J. Neuropsychol.* 1, 82–90. doi: 10.1016/j.psychres.2011.04.025

Murakami, Y., Sakai, S., Takeda, K., Sawamura, D., Yoshida, K., Hirose, T., et al. (2018). Autistic traits modulate the activity of the ventromedial prefrontal cortex in response to female faces. *Neurosci. Res.* 133, 28–37. doi: 10.1016/j.neures.2017.11.003

Pegors, T. K., Kable, J. W., Chatterjee, A., and Epstein, R. A. (2015). Common and unique representations in pFC for face and place attractiveness. *J. Cogn. Neurosci.* 27, 959–973. doi: 10.1162/jocn_a_00777

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., et al. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature* 394, 884–887. doi: 10.1038/29772

Pessiglione, M., and Daunizeau, J. (2021). Bridging across functional models: the OFC as a value-making neural network. *Behav. Neurosci.* 135, 277–290. doi: 10.1037/bne0000464

Pham, T. Q., Yoshimoto, T., Niwa, H., Takahashi, H. K., Uchiyama, R., Matsui, T., et al. (2021). Vision-to-value transformations in artificial network and human brains. *Biorxiv* [Preprint]. doi: 10.1101/2021.03.18.435929

Polanía, R., Krajbich, I., Grueschow, M., and Ruff, C. C. (2014). Neural oscillations and synchronization differentially support evidence accumulation in perceptual and value-based decision making. *Neuron* 82, 709–720. doi: 10.1016/j.neuron.2014.03.014

Popenko, N. A., Tripathi, P. B., Devcic, Z., Karimi, K., Osann, K., and Wong, B. J. F. (2017). A quantitative approach to determining the ideal female lip aesthetic and its effect on facial attractiveness. *JAMA Facial Plast. Surg.* 19, 261–267. doi: 10.1001/jamafacial.2016.2049

Rangel, A., Camerer, C., and Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9, 545–556. doi: 10.1038/nrn2357

Ratcliff, R., and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* 20, 873–922. doi: 10.1162/neco.2008.12-06-420

Ratcliff, R., Philiastides, M. G., and Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *PNAS* 106, 6539–6544. doi: 10.1073/pnas.0812589106

Ratcliff, R., Smith, P. L., Brown, S. D., and McKoon, G. (2016). Diffusion decision model: current issues and history. *Trends Cogn. Sci.* 20, 260–281. doi: 10.1016/j.tics.2016.01.007

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annu. Rev. Psychol.* 57, 199–226. doi: 10.1146/annurev.psych.57.102904.190208

Rhodes, M. G., and Anastasi, J. S. (2012). The own-age bias in face recognition: a meta-analytic and theoretical review. *Psychol. Bull.* 138, 146–174. doi: 10.1037/a0025750

Riga, D., Matos, M. R., Glas, A., Smit, A. B., Spijker, S., and Van den Oever, M. C. (2014). Optogenetic dissection of medial prefrontal cortex circuitry. *Front. Syst. Neurosci.* 8:230. doi: 10.3389/fnsys.2014.00230

Rolls, E. T., and Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: from affect to decision-making. *Prog. Neurobiol.* 86, 216–244. doi: 10.1016/j.pneurobio.2008.09.001

Sakano, Y., Wada, A., Ikeda, H., Saheki, Y., Tagai, K., and Ando, H. (2021). Human brain activity reflecting facial attractiveness from skin reflection. *Sci. Rep.* 11:3412. doi: 10.1038/s41598-021-82601-w

Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., and Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science* 340, 216–219. doi: 10.1126/science.1231059

Samanez-Larkin, G. R., and Knutson, B. (2015). Decision making in the ageing brain: changes in affective and motivational circuits. *Nat. Rev. Neurosci.* 16, 278–289. doi: 10.1038/nrn3917

Saxe, R. (2006). Uniquely human social cognition. *Curr. Opin. Neurobiol.* 16, 235–239. doi: 10.1016/j.conb.2006.03.001

Schwarzer, G., Carpenter, J., and Rücker, G. (2015). *Meta-Analysis with R*. Berlin: Springer.

Shimojo, S., Simion, C., Shimojo, E., and Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nat. Neurosci.* 6, 1317–1322. doi: 10.1038/Nn1150

Strickland-Hughes, C. M., Dillon, K. E., West, R. L., and Ebner, N. C. (2020). Own-age bias in face-name associations: evidence from memory and visual attention in younger and older adults. *Cognition* 200:104253. doi: 10.1016/j.cognition.2020.104253

Su, Y.-S., Chen, J.-T., Tang, Y.-J., Yuan, S.-Y., McCarrey, A. C., and Goh, J. O. S. (2018). Age-related differences in striatal, medial temporal, and frontal involvement during value-based decision processing. *Neurobiol. Aging* 69, 185–198. doi: 10.1016/j.neurobiolaging.2018.05.019

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., and Young, A. W. (2013). Social inferences from faces: ambient images generate a three-dimensional model. *Cognition* 127, 105–118. doi: 10.1016/j.cognition.2012.12.001

Suzuki, S., Cross, L., and O'Doherty, J. P. (2017). Elucidating the underlying components of food valuation in the human orbitofrontal cortex. *Nat. Neurosci.* 20, 1780–1786. doi: 10.1038/s41593-017-0008-x

Suzuki, S., and O'Doherty, J. P. (2020). Breaking human social decision making into multiple components and then putting them together again. *Cortex* 127, 221–230. doi: 10.1016/j.cortex.2020.02.014

Todd, P. M., Penke, L., Fasolo, B., and Lenton, A. P. (2007). Different cognitive processes underlie human mate choices and mate preferences. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15011–15016. doi: 10.1073/pnas.0705290104

Todorov, A., Mandisodza, A. N., Goren, A., and Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science* 308, 1623–1626. doi: 10.1126/science.1110589

Towal, R. B., Mormann, M., and Koch, C. (2013). Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *PNAS* 110, E3858–E3867. doi: 10.1073/pnas.1304429110

Tsoi, K. K. F., Chan, J. Y. C., Hirai, H. W., Wong, S. Y. S., and Kwok, T. C. Y. (2015). Cognitive tests to detect dementia: a systematic review and meta-analysis. *JAMA Intern. Med.* 175, 1450–1458. doi: 10.1001/jamainternmed.2015.2152

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* 30, 829–858. doi: 10.1002/hbm.20547

Wiecki, T. V., Sofer, I., and Frank, M. J. (2013). HDDM: hierarchical bayesian estimation of the drift-diffusion model in python. *Front. Neuroinform.* 7:14. doi: 10.3389/fninf.2013.00014

Winston, J. S., O'Doherty, J., Kilner, J. M., Perrett, D. I., and Dolan, R. J. (2007). Brain systems for assessing facial attractiveness. *Neuropsychologia* 45, 195–206. doi: 10.1016/j.neuropsychologia.2006.05.009

Wright, D. B., and Stroud, J. N. (2002). Age differences in lineup identification accuracy: people are better with their own age. *Law Hum. Behav.* 26, 641–654. doi: 10.1023/A:1020981501383

Yoon, L., Somerville, L. H., and Kim, H. (2018). Development of MPFC function mediates shifts in self-protective behavior provoked by social feedback. *Nat. Commun.* 9:3086. doi: 10.1038/s41467-018-05553-2

Zebrowitz, L. A., Fellous, J.-M., Mignault, A., and Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: evidence from connectionist modeling. *Pers. Soc. Psychol. Rev.* 7, 194–215. doi: 10.1207/S15327957PSPR0703_01

Zhang, Z., Wang, S., Good, M., Hristova, S., Kayser, A. S., and Hsu, M. (2021). Retrieval-constrained valuation: toward prediction of open-ended decisions. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2022685118. doi: 10.1073/pnas.2022685118

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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 17 May 2022
doi: 10.3389/fpsyg.2022.864936

The Attractiveness of Masked Faces Is Influenced by Race and Mask Attitudes

Veronica Dudarev[1], Miki Kamatani[2], Yuki Miyazaki[3], James T. Enns[1] and Jun I. Kawahara[2]*

[1]*Department of Psychology, University of British Columbia, Vancouver, BC, Canada,* [2]*Faculty of Humanities and Human Sciences, Hokkaido University, Sapporo, Japan,* [3]*Department of Psychology, Fukuyama University, Hiroshima, Japan*

OPEN ACCESS

Edited by:
Ayahito Ito,
Kochi University of Technology, Japan

Reviewed by:
Giulio Gabrieli,
Italian Institute of Technology (IIT), Italy
Ryuhei Ueda,
Kyoto University, Japan

***Correspondence:**
Veronica Dudarev
vdudarev@mail.ubc.ca

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

Received: 29 January 2022
Accepted: 20 April 2022
Published: 17 May 2022

Citation:
Dudarev V, Kamatani M, Miyazaki Y, Enns JT and Kawahara JI (2022) The Attractiveness of Masked Faces Is Influenced by Race and Mask Attitudes. Front. Psychol. 13:864936. doi: 10.3389/fpsyg.2022.864936

This study tests the influence of wearing a protective face mask on the perceived attractiveness of the wearer. Participants who identified as White, and who varied in their ideological stance toward mask wearing, rated the attractiveness of facial photographs. The photos varied in baseline attractiveness (low, medium, and high), race (White and Asian), and whether or not the face was wearing a protective mask. Attitudes regarding protective masks were measured after the rating task using a survey to identify participants as either pro- or anti-mask. The results showed that masked individuals of the same race were generally rated as more attractive than unmasked individuals, but that masked individuals of another race were rated as less attractive than unmasked individuals. Moreover, pro-mask participants rated masked individuals as generally more attractive than unmasked individuals, whereas anti-maskers rated masked individuals as less attractive. A control experiment, replicating the procedure but replacing the protective masks with a partially occluding notebook, showed that these effects were mask-specific. These results demonstrate that perceived attractiveness is affected by characteristics of the viewer (attitudes toward protective masks), their relationship to the target (same or different race), and by circumstances external to both (pandemic).

Keywords: sanitary mask, protective mask, COVID-19, facial attractiveness, microvalence, affective appreciation, affective devaluation

INTRODUCTION

The theory of microvalences proposes that every-day objects are assigned emotional values that influence our perception and action (Lebrecht et al., 2011, 2012; Todd and Manaligod, 2018). The idea is that even the most mundane objects accrue a valence through our experience with them that is more than just random variance around a neutral mean (Lebrecht et al., 2012). This is supported by growing evidence that, just as highly-valenced objects such as guns and roses can attract our attention and guide our choices, the microvalences of common objects can do the same (Lebrecht et al., 2012; Lebrecht and Tarr, 2012; Roos et al., 2013). One untested prediction of the theory, as far as we are aware, is that the individual microvalences of objects encountered in the same context can affect one another (Lebrecht et al., 2012). If this is occurring implicitly, then person perception (the evaluation of one object) may be influenced by the microvalence of an object associated with that person (the evaluation of a second object), even if the presence of that object does not express the person's *choice*. That is, seeing someone with a cake or a credit card in hand may increase the emotional attitude toward the person, while seeing one with a knife or a bag of garbage may decrease liking of the person. In the present study, we test this prediction by exploring how the microvalence of a newly familiar object can influence the emotional appraisal of a person using it.

One great challenge to testing this theory is that the microvalence of an object is shaped by one's personal history of interacting with it (Lebrecht et al., 2012), which means that most microvalences are personally idiosyncratic and therefore hard to control in research. The recent worldwide COVID-19 pandemic, however, introduced a new object into the lives of many people around the world—the protective mask (Institute for Health Metrics and Evaluation, 2021)—providing an opportunity to study the microvalence associated with it in people from different populations and circumstances.

Although cloth and medical masks are effective in controlling the spread of COVID-19 (Gandhi et al., 2020; Lyu and Wehby, 2020; Howard et al., 2021; Rader et al., 2021), many people oppose this practice because of personal beliefs or ideological affiliations (Aratani, 2020; Beer, 2020; Karalis, 2020; Knotek et al., 2020; Miller and Brueck, 2020). The efficiency of the masks in reducing the spread of the virus and the resistance of some people to wearing them make investigation into the emotional appraisal of masks not only theoretically interesting, but practically relevant. In a recent study, we explored the emotional appraisal of protective masks and showed that the frequency of their use during the pandemic predicted their positive emotional appraisal in the context of a study of consumer products (Dudarev et al., 2021). This *usage-liking* relationship held up even after controlling for individual differences in beliefs about the dangers of COVID-19 and for environmental exposure to COVID-19 (Dudarev et al., 2021). In the present study, we go a step further to study the effects of a protective mask on the perception of the attractiveness of its wearer.

This question has been asked before the COVID-19 pandemic. Miyazaki and Kawahara (2016) took advantage of the fact that many Japanese citizens took to wearing masks in public following outbreaks of Spanish flu. Since then, face masks are often used in Japan to prevent the spread of colds and flu. Miyazaki and Kawahara (2016) compared the attractiveness ratings given by Japanese viewers of Japanese faces covered with a protective mask, with the same faces when they were not wearing a mask. They found that masked faces were generally rated as less attractive than unmasked faces. This cost of wearing a protective mask interacted with the baseline attractiveness, such that there was a greater cost to the ratings when the faces were otherwise rated as high in attractiveness.

Miyazaki and Kawahara (2016) identified two factors at work in the attractiveness costs of masked faces. The first factor was occlusion: A mask covering the lower part of the face hides advantageous features for attractive faces (e.g., smooth skin and symmetry), at the same time, that it hides disadvantageous features for less attractive faces (e.g., skin blemishes, scars, or asymmetry). This has the effect of truncating the range of attractiveness ratings for masked compared to unmasked faces. However, this single factor did not fully explain the results. The second factor was an appreciation that, in keeping with the social norms in Japan at the time, wearing a mask was associated with potential vulnerability of the wearer to pollen allergies, a sore throat, or more serious illness. This "unhealthiness priming" effect resulted in overall decrease in attractiveness ratings for masked faces that was independent of their baseline attractiveness. This interpretation was supported by a control experiment, where masks were replaced by health-neutral cards and notebooks, and the results showed only a truncation of the range of ratings. And as further support, Miyazaki and Kawahara's (2016) showed that when participants were asked specifically to rate the health of mask wearers (rather than their attractiveness), the results were consistent with the effects they had attributed to "unhealthiness priming."

The same research group replicated the procedures of the 2016 study in 2020, during the height of the COVID-19 pandemic in Japan, where virtually the entire population had taken to regularly wearing mask (Kamatani et al., 2021). They hypothesized that this normalization of mask wearing would reduce the "unhealthiness priming" effect, since masks no longer held a signal regarding any individual's personal health status and they had become a symbol representing conformity to societal norms (Nakayachi et al., 2020). Indeed, the only effect remaining in the attractiveness ratings for masked faces in 2020 was the truncated range of ratings attributable to visible occlusion of the face.

Here, we follow the paradigm and logic developed by Miyazaki and Kawahara (2016) and Kamatani et al. (2021) to study how a protective mask affects the attractiveness of the mask wearer during the height of the COVID-19 pandemic in North America. Our study design allowed us to test two hypotheses. First, we tested whether the attractiveness of a wearer of this object depends on the perceived similarity of the wearer with the viewer. We manipulated this similarity by presenting participants with facial photographs of people of either the same race as the viewer (White) or a different race (Asian). If personal similarity to the viewer influences rated attractiveness, then protective masks should increase the rated attractiveness of own-race (White) faces relative to the rated attractiveness of other-race (Asian) faces. This own-race bias can stem from any number of factors, including the improved perceptual fluency when assessing own-race faces (Alter and Oppenheimer, 2006), greater empathy for faces that are seen as more similar to the viewer (Chiao and Mathur, 2010), and the tendency to process own-race faces with a greater level of detail than other-races faces (Levin, 2000). Here, we simply used own- versus same-races face as a convenient way to manipulate the perceived closeness of rated faces to the participant.

Our second hypothesis was that the effects of the mask on the rated attractiveness of the wearer will depend on the viewer's attitude toward the mask itself. We measured attitudes toward masks and COVID-19 protective measures in a survey administered after the face rating task, in order to index the

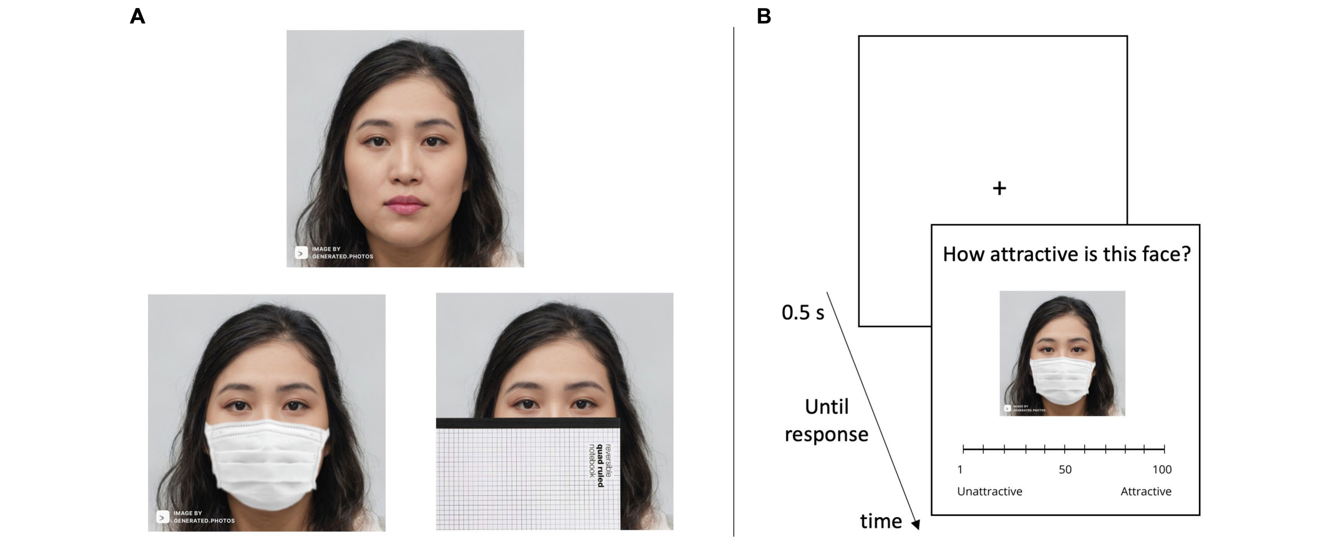


FIGURE 1 | (A) Examples of an unmasked Asian face shown together with the same face when masked (Experiment 1) or when partially occluded by a notebook (Experiment 2). **(B)** The trial event sequence in the study. The image depicted in the figure was computer generated for the purposes of illustration (https://generated.photos/) and was not one of the faces used in the study.

microvalence associated with masking for each participant. We hypothesized that more positive attitudes toward masks would be associated with more positive attractiveness ratings for people wearing masks. Conversely, anti-mask sentiment would be associated with lower perceived attractiveness of mask wearers.

The study was planned in April 2021 and the data were collected between 25 April and 5 May 2021. It is important to note that at this time, mask mandates had been implemented almost everywhere in the US for at least several months. Just after we collected our data, on 13 May, the Centers for Disease Control lifted mask mandate for fully vaccinated people, yet soon afterward this decision was reversed. Thus, seeing a person with a facial mask, in April–May 2021, could hardly be interpreted as expression of uniquely pro-social behavior. Rather, we interpreted it as minimally complying with the new rules and norms.

EXPERIMENT 1

We randomly assigned participants to one of two versions of a face attractiveness experiment: faces of the same race as the participants (White) or faces of another race (Asian). As shown in **Figure 1**, half of the faces in each condition, presented randomly, were wearing a white protective face mask and the other half were wearing no mask. We hypothesized that attractiveness ratings for masked faces would be relatively more positive for faces of the same race as the participant than for faces of a different race. Such an own-race advantage for masked faces could come about either because participants are better able to infer missing details from familiar faces (Carragher and Hancock, 2020; Freud et al., 2020; Molnar-Szakacs et al., 2021) or because masks on own-race faces are seen as more similar to the participant and thus deserving of empathy (Dudarev et al., 2021).

Method

Participants

Based on the previous studies (Miyazaki and Kawahara, 2016; Kamatani et al., 2021), we aimed to collect data from at least 32 participants in each condition. We estimated the statistical power of our experimental design to detect the previously documented interaction between attractiveness and masking (Miyazaki and Kawahara, 2016; Kamatani et al., 2021), based on a sample size of 32 and an effect size of η_p^2 =0.23, was approaching 1. Our estimate of the power to detect an interaction between race and masking, based on two independent groups of 32 participants for each race of faces, was 0.82 for a large effect (η_p^2 = 0.13) and 0.42 for a smaller effect (η_p^2 = 0.06; Lakens and Caldwell, 2021).

Anticipating potential data loss due to participants' failure to complete all the trials, inattention to the task, or failure to follow instructions, we set our recruitment sample size to be 36 participants per condition. We also automated the random assignment of participants to conditions (race of faces) through MTurk, which meant that there was no guarantee the final sample size would be exactly equal.

Participants self-identifying as White and aged 18 or older were recruited on Amazon's MTurk platform.[1] Participants' race was determined *via* an eligibility survey that included demographic questions about their ethnicity, family income, age, and gender. Another requirement for

[1]https://www.mturk.com

participation was that the experiment was conducted on a desktop or laptop computer.

Materials

The photographic images used in the study consisted of female faces. The Asian faces were of Japanese individuals used in a previous study of face attractiveness (Miyazaki and Kawahara, 2016). This set of images consists of 48 photographs of young adult women with either neutral or slightly smiling expressions. The images were rated for attractiveness by a group of Japanese participants, as described in Miyazaki and Kawahara (2016), and were subdivided into three highly reliable categories of attractiveness: low, medium, and high (see **Supplementary Material** for details). The White female faces used in the study were taken from the Chicago Face Database (Ma and Wittenbrink, 2015).[2] Only versions with neutral expressions were used. The images in this database already have normed attractiveness ratings, which allowed us to select 48 images, evenly divided to represent low, medium, and high levels of attractiveness. Descriptive statistics on the images selected for the present study from norming studies (Ma and Wittenbrink, 2015; Miyazaki and Kawahara, 2016) and the present study are available in the **Supplementary Material**. To obtain 96 corresponding masked faces, we digitally edited each of the photos in order to cover the lower portion of the face with a white protective mask. We pasted the same mask image into each face after adjusting it for size and smoothing artifactual edges using a Gaussian blur tool (Adobe Photoshop CS6).

Procedure

Figure 1 shows examples of the face images in the study and illustrates the trial event sequence. During the attractiveness rating task, participants were presented with a face on each trial that was either wearing a protective mask or not. Participants were instructed to rate the attractive of this face on a 100-point scale (1 = most unattractive and 100 = most attractive). A line depicting the scale from 1 to 100 was shown below each face and participants responded by clicking the line in the location corresponding to their rating. Participants were not given any time limit for their responses.

In several practice trials, prior to testing, participants were shown a very attractive face from the set that was not used in the study, as well as an extremely unattractive face. This was done in order to help them calibrate their responses to the full range of the scale and to reduce idiosyncratic individual differences.

Images of faces covered with a mask and faces wearing no mask were presented in random order. Each participant only saw a given individual's face one time, either masked or not masked, and these selections were made randomly by the computer program so that each participant rated a total of 48 individuals, evenly distributed over the three attractiveness levels.

[2]https://www.chicagofaces.org/

Participants were randomly assigned to view either White or Asian faces by the computer program. We chose a between-participant design for two reasons. First, we did not want participants to knowingly compare their own ratings for faces of one race to ratings of another race. Rather we wanted only their spontaneous responses to faces that varied in baseline attractiveness within a single race, without possible contamination from thoughts of how they were handling cross-race ratings. Second, we wanted the study to be a brief online study, so that considerations of tediousness or fatigue would not be a factor.

After completing the rating task, participants filled out survey questions about their demographic and sociocultural status, and their attitudes, practices, and experiences with the COVID-19 virus and protective masks (Dudarev et al., 2021). The entire testing session took no longer than 30 min, and participants were paid 2$. The study was approved by the Behavioral Research and Ethics Board at the University of British Columbia (approval number H21-00305). The data were collected between 25 April and 5 May 2021.

Results

The Effects of Race on Perceived Attractiveness of Masked Faces

Seventy-nine participants completed all 48 trials of the face attractiveness ratings, with 30.4% self-identified as female, and a mean reported age of 39.9 (ranging = 22–72 years). Forty-two participants were randomly assigned to rate White faces (31% female); 37 were assigned to rate Asian faces (29.7% female). Two participants (one from each group) responded with the same rating on all trials, but exclusion of their data did not change the results.

Figure 2 shows mean attractiveness ratings as a function of the within-participant factors of attractiveness level (low, medium, and high), and mask (off and on), and the between-participant factor of race of face (White, Asian). The three levels of baseline attractiveness were established using normed data from previous studies (Ma and Wittenbrink, 2015; Miyazaki and Kawahara, 2016). A mixed model ANOVA of these ratings showed an expectedly strong effect of attractiveness, $F(2,154) = 102.58$, $p < 0.001$, $\eta_p^2 = 0.571$, replicating previous findings for these faces (White faces: Ma and Wittenbrink, 2015; Asian faces: Miyazaki and Kawahara, 2016). The main effect of attractiveness was qualified by an interaction with mask, $F(2,154) = 21.15$, $p < 0.001$, $\eta_p^2 = 0.215$, reflecting that the addition of a protective mask reduced ratings for highly attractive faces and increased ratings for unattractive faces. The main effect of mask was not significant, $F(1,77) = 0.18$, $p = 0.67$, $\eta_p^2 = 0.002$.

Race also yielded a main effect, $F(1,77) = 5.86$, $p = 0.018$, $\eta_p^2 = 0.071$, reflecting that Asian faces were rated overall as more attractive than White faces. But this main effect was qualified by an interaction with Mask, $F(1,77) = 6.57$, $p = 0.012$, $\eta_p^2 = 0.079$, reflecting that masked Asian faces were rated lower than unmasked Asian faces, while masked White faces were rated higher than unmasked White faces. No other effects reached significance, $Fs < 2.5$, $ps > 0.09$.

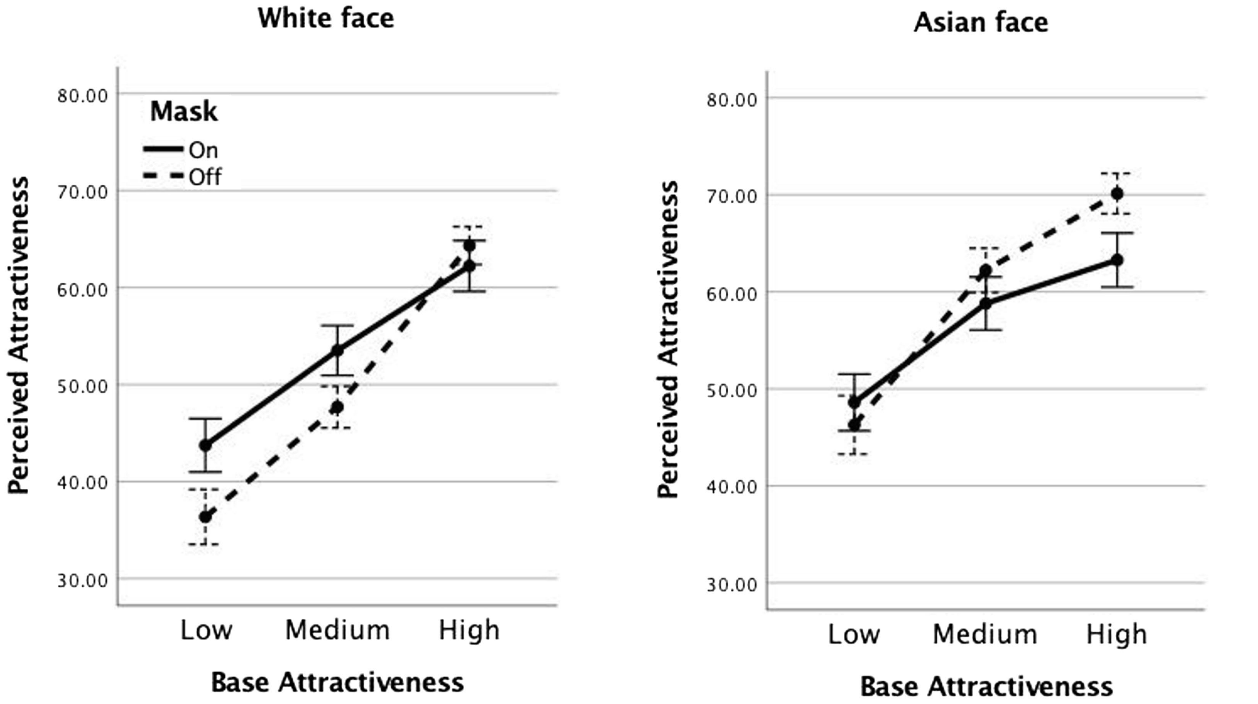


FIGURE 2 | Mean attractiveness ratings for White and Asian faces of low, medium, and high base attractiveness, when covered with a protective mask (solid line) or not (dashed line). Error bars represent +/– 1 standard error.

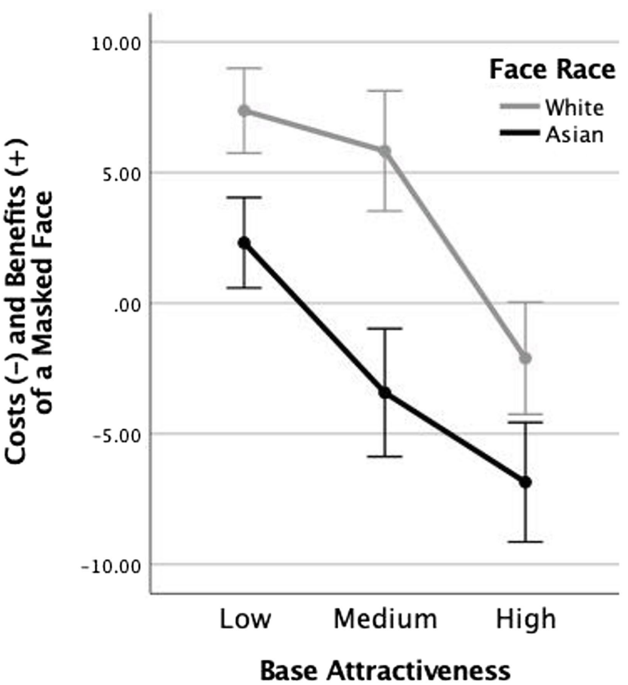


FIGURE 3 | Mean difference scores show the difference between ratings for masked and unmasked faces, across the three levels of baseline attractiveness and two races. The difference scores were calculated for each participant as the difference between ratings for masked and unmasked faces. Error bars represent +/– 1 standard error.

To examine whether individual differences in the way participants used the attractiveness scale affected the results, we have computed a centered score by subtracting each participant's mean rating from each rating on a single trial. The centered scores revealed the same pattern of results as the raw ratings. Note that participants' use of the scale was calibrated during the practice trials.

To help visualize the nature of the race × mask interaction, we summarized the costs and benefits of wearing a protective mask by subtracting each participant's attractiveness ratings for faces with no mask from faces with a mask. The mean difference scores are shown in **Figure 3** and were submitted to a mixed model ANOVA examining the within-participant factor of attractiveness level (low, medium, and high) and the between-participant factor of race (White and Asian). We note that this analysis is redundant with the previous one, the only difference is that once the mask factor is expressed as a difference score, the interaction between attractiveness and mask becomes a main effect of attractiveness, and the interaction between race and mask becomes a main effect of race. The three-way interaction between attractiveness × race × mask is now accordingly a two-way interaction between attractiveness × race. This approach to examining repeated-measures effects as simple contrasts, based on theoretical considerations, is recommended by statisticians both for its increased clarity of presentation and for its efficient handling of any concerns regarding assumptions of sphericity (Wickens and Keppel, 2004). Here, the theoretical consideration is that we are primarily interested in what wearing a mask does to face attractiveness scores.

The analysis of the difference scores showed two main effects. The main effect of attractiveness, $F(1,77) = 21.15$, $p < 0.001$, $\eta_p^2 = 0.215$, shows that wearing a mask benefits attractiveness ratings when a face is already seen as highly attractive and reduces the ratings when it is seen as lower in attractiveness. The main effect of race, $F(1,77) = 6.57$, $p = 0.012$, $\eta_p^2 = 0.079$, shows that masked Asian faces were generally seen as lower in attractiveness than unmasked Asian faces, while for the White faces, the opposite was true. The absence of any interaction between these two factors, $F(2,154) = 1.52$, $p > 0.2$, $\eta_p^2 = 0.019$, highlights that the race and attractiveness effects on the ratings are independent of one another.

Discussion

The results of Experiment 1 first of all replicate an important finding reported by Kamatani et al. (2021) concerning the consequences of wearing a mask on face attractiveness ratings. Faces that are reliably rated low in attractiveness when viewed in full are rated as more attractive when masked; highly attractive faces are rated as less attractive when partially covered with a mask. Along with Kamatani et al. (2021), we interpret the truncated range of attractiveness ratings for masked faces to the occlusion of visible evidence that places these faces at each end of the attractiveness scale. Facial blemishes and flaws that contribute to low ratings are less visible, as are the features supporting high ratings of attractiveness at the other end of the scale.

The novel finding from Experiment 1 is that attractiveness rates of faces are influenced differently by the addition of a protective mask, depending on the similarity of the race of the face to the viewer's own race. Adding a mask to the White faces improved their rated attractiveness, while adding the same mask to Asian faces decreased their attractiveness ratings. It is worth considering two possible explanations for this effect. One possibility is that people are better at distinguishing identities and emotional expressions of faces of their own race than faces of a different race (Elfenbein and Ambady, 2002; Brielmann et al., 2014), possibly because scanning patterns are different for faces of different races (Brielmann et al., 2014). On this account, the occlusion of the face that accompanies the mask may have a differential influence on own-race versus other-race faces because viewers are better able to infer missing details from familiar faces (Carragher and Hancock, 2020; Freud et al., 2020; Molnar-Szakacs et al., 2021). Own-race faces that are more familiar will benefit in their rated attractiveness from the addition of a mask, whereas other-race faces that are less familiar will receive reduced ratings.

A second possibility is that protective masks carry an additional emotional value for participants—a microvalence that is positive for some viewers and negative for others (Dudarev et al., 2021). A microvalence attached to a protective mask may differentially influence ratings of faces that are seen to be more relevant to the participant (own race) than faces that are less personally relevant (other race). In Experiment 2, we devised a test to distinguish between the possibilities by occluding the lower portion of the faces with a notebook rather than a protective mask. Our assumption was that a notebook would be seen as more emotionally neutral than a protective mask by viewers during the pandemic. If notebooks produced similar results for those obtained in Experiment 1 for protective masks, it would support the interpretation that occlusion of the lower portion of the face has a differential influence on own- versus other-race faces. However, if the results using notebook occlusion eliminated the other-race effect seen in Experiment 1, then it would support the microvalence account. This account proposes that the emotional value associated with protective masks affects own- and other-race faces differently because of an interaction between the emotional appraisal of the mask and the personal relevance of the face wearing the mask.

EXPERIMENT 2

Method

Experiment 2 followed the procedures of Experiment 1 in their entirety, with a single change. Half of the faces now had their lower portions covered with a notebook (as shown in **Figure 1**) rather than a protective mask.

Participants

Seventy-five participants completed the experiment (29.3% female, mean age = 40.5, ranging from 22 to 72 years), with 37 of them randomly assigned to rate White faces (32.4% female) and 38 to rate Asian faces (26.3% female). None of the participants in Experiment 2 had participated in Experiment 1. Two participants (one from each group) responded with the same rating on all trials, but exclusion of their data does not change the results.

Results

These data were analyzed in the same way as in Experiment 1. The mixed model ANOVA on the ratings showed only a main effect of attractiveness, $F(2,146) = 84.87$, $p < 0.001$, $\eta_p^2 = 0.538$, and an interaction between notebook occlusion and attractiveness, $F(2, 146) = 29.04$, $p < 0.001$, $\eta_p^2 = 0.285$. Notebook occlusion had neither a main effect nor was involved in any other interactions, $Fs < 1$, $ps > 0.7$. These results are similar to the effects produced by mask occlusion in both magnitude and pattern (see **Supplementary Figure S1**). However, a crucial difference in these data involving notebook occlusion was that there was not a significant interaction between race and occlusion, $F(1,73) = 1.88$, $p = 0.17$, $\eta_p^2 = 0.025$.

Figure 4 shows the differences scores summarizing the benefits and costs of being rated for attractiveness when the lower portion of the face is occluded by a notebook. Like the masked faces in Experiment 1, notebook occlusion differentially affected faces that varied in baseline attractiveness, $F(2, 146) = 29.04$, $p < 0.001$, $\eta_p^2 = 0.285$, specifically increasing ratings of low-attractive faces and reducing ratings of high-attractive faces. However, there was no effect of race in the notebook occlusion differences scores, $F(1,73) = 1.88$, $p = 0.174$, $\eta_p^2 = 0.025$. The interaction was not significant either, $F(2, 146) = 1.28$, $p = 0.28$, $\eta_p^2 = 0.017$.

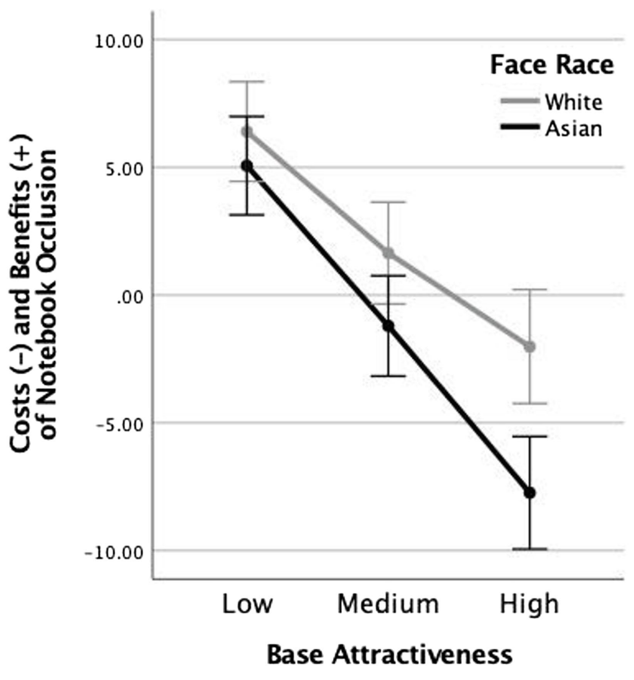


FIGURE 4 | Notebook occlusion cost and benefit score, computed as the difference between attractiveness ratings for faces occluded with a notebook and those not occluded by anything, for White and Asian faces. Error bars represent +/− 1 standard error.

Could it be that the different pattern of results in Experiments 1 and 2 reflected group differences in the way attractiveness was being evaluated? To test for this possibility, we focused on the parts of Experiments 1 and 2 that were most similar, namely, the ratings of full faces viewed without any occlusion. These data are shown in **Figure 5** and they confirm that baseline attractiveness ratings were comparable in the two experiments. A 3 (base attractiveness) × 2 (race) × 2 (Experiment: 1 vs. 2) ANOVA revealed that the factor of Experiment had no main effect, $F(1,150) = 0.306$, $p = 0.58$, $\eta_p^2 = 0.002$, and was not involved in any interactions, $Fs < 1.2$, $ps > 0.2$. Thus, the pattern of results we observed when faces were partially covered with a protective mask versus a notebook cannot be attributed to baseline differences in the way these different groups of participants rated the attractiveness of faces in general.

Race, on the other hand, had a main effect, $F(1,150) = 12.97$, $p < 0.001$, $\eta_p^2 = 0.08$, which was qualified by an interaction with base attractiveness, $F(2,300) = 4.004$, $p = 0.019$, $\eta_p^2 = 0.026$. The interaction arose from participants assigning a smaller range of ratings to other-race than same-race faces.

Discussion

The results of Experiment 2 fully supported the microvalence account of the results of Experiment 1. Partially occluding faces with a notebook did not result in the own-race effect that occurred when faces were partially occluded with protective masks. We interpret this as supporting the view that both

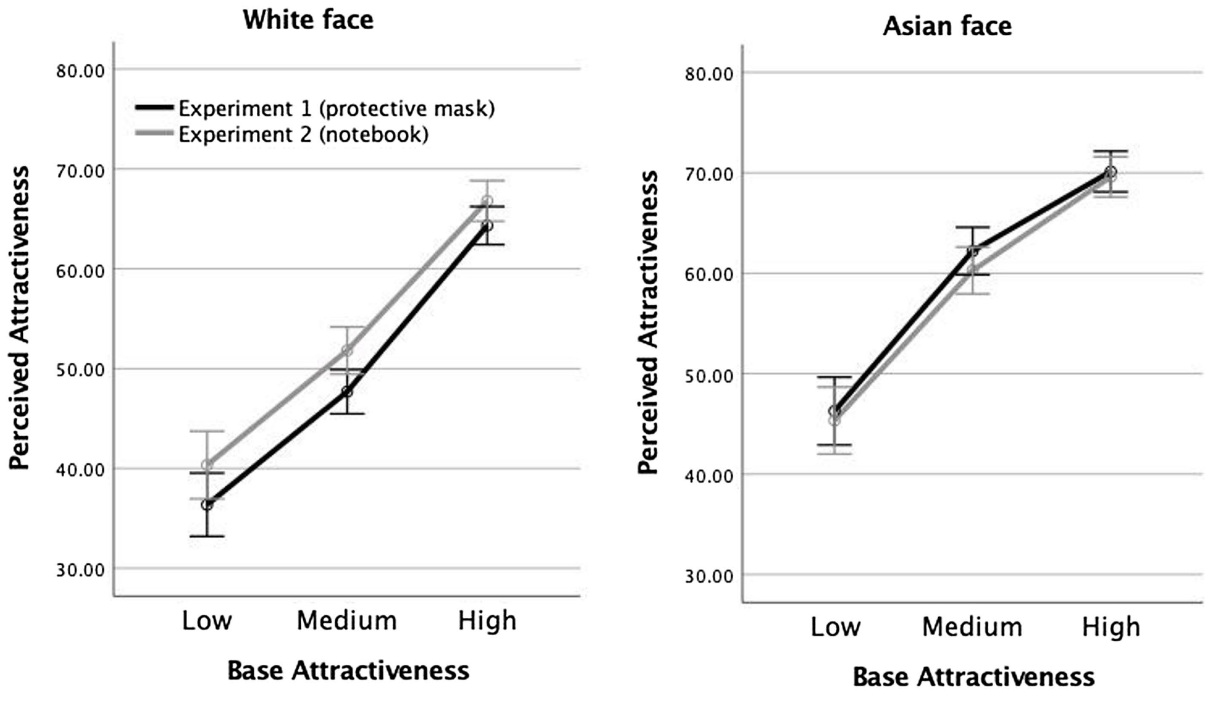


FIGURE 5 | Perceived attractiveness for faces without any occlusion in Experiment 1 (black line) and Experiment 2 (gray line). Error bars represent +/− 1 standard error.

protective masks and other human faces are implicitly associated with emotional evaluations that influence our perceptions and actions. We will return to this interpretation below; in our analysis of the results from the attitude survey, we conducted with all participants.

But first, it is worth considering the consequences of partial face occlusion on attractiveness ratings. This refers to those results that were held in common between Experiments 1 and 2. As proposed by Miyazaki and Kawahara (2016) and Kamatani et al. (2021), partially occluding a face in any way has the consequence of truncating the range of attractiveness ratings, resulting effectively in a regression to the mean. This means it improves the ratings of low-attractive faces and reduces the ratings of high-attractive faces. And when the occlusion was attributable to an emotionally neutral object, its moderating effect was similar for own- and other-race faces.

At the same time, we acknowledge that occlusion by notebook versus mask differs not only in emotionally relevant ways, but also in a number of lower-level stimulus features. For example, the notebook in Experiment 2 occluded a slightly larger portion of the face than the mask in Experiment 1, possibly leading to greater truncation of ratings across the three levels of attractiveness. Notebook occlusion was also less symmetrical, a factor which might decrease overall ratings of attractiveness. These issues have been addressed to some extent in previous research. Miyazaki and Kawahara (2016) compared attractiveness ratings made to faces occluded with a notebook (Experiment 3a, 3b) with faces occluded with a paper card that was more similar in its size and its symmetry to a protective face mask (Experiment 4). The data were clear in showing that notebooks and cards produced very similar effects (Experiments 3 and 4) and that both types of occlusion resulted in a general improvement in attractiveness ratings, compared to faces occluded by protective masks. Thus, previous data collected in a pre-pandemic setting suggested that the size and symmetry of the occlusion did not play a measurable role in attractiveness ratings. And when it came to a comparison between these accidental types of occlusion and the wearing of a protective mask, facial occlusion by masks generally improved attractiveness ratings.

We were also able to address these concerns in our own data, collected during the same period of time and from the same participant population as the protective masking data. The full statistical analysis of the notebook occlusion data along with the graph of these data that correspond to **Figure 2** for the mask data is reported in the **Supplementary Material**. The main take-away from this full analysis is that occlusion by notebook did not result in any discernible enhancement nor decrease in attractiveness ratings when compared to occlusion by mask. Thus, both previously reported data and the data presented here suggest that the stimulus differences that remain between notebooks and protective masks do not affect face attractiveness. Neither do they account for the different patterns in the attractiveness ratings for faces of different races.

THE INFLUENCE OF MASK ATTITUDES ON PERCEIVED ATTRACTIVENESS OF FACES WEARING PROTECTIVE MASKS

Microvalence account posits that emotional appraisal of objects is determined, among other things, by participants' personality and individual history. According to this interpretation, participants' ideological attitudes toward COVID-19 protective measures should be a mediating factor in determining the rated attractiveness of people wearing masks. It is now well known that political and medical advice to wear protective masks during the pandemic sharply divided people in the United States. Our intent was therefore to elicit participants' attitudes regarding society's response to the pandemic and regarding various proposed measures of protection, which we did in the post-test survey to avoid contaminating our dependent variable (attractiveness ratings).

In a previous study using the same survey items (Dudarev et al., 2021), the results showed that attitudes expressed by participants about COVID-19 measures were better predictors of their emotional appraisal of consumer products associated with the pandemic (e.g., protective masks, hand hygiene products, and protective gloves) than of similar items that were not related to COVID-19 (e.g., ski goggles, shampoo, and winter gloves). In the present study, we therefore hypothesized that participants who identified as pro-mask would show a greater increase in attractiveness ratings for mask-wearers than participants who identified as anti-mask. Our secondary question was whether these mask-related biases would also be expressed more strongly for faces seen as more personally relevant to the participants (own-race) than faces seen as less personally relevant (other-race). In other words, we tested whether the wearer's race and the viewer's attitudes toward masks have interactive influence on the wearer's perceived attractiveness or whether their effects are additive.

Attitude Survey Data

All participants in Experiment 1 and 2 were presented with the attitude and demographic survey after first the completing the rating task (136 of the participants completed it). The complete list of 48 questions is presented in the **Supplementary Material**. The questions included participants' age, gender, income, profession, cultural background, and religion (Whitman et al., 2018), as well as questions about participants' political leanings and interest in politics. In addition, participants were asked several questions focusing on their COVID-19 experiences ("Have you had COVID?"), vulnerability and exposure to the virus ("Do you consider yourself to be especially vulnerable to COVID?"; "How many people do you meet offline on a typical week?"—Wright et al., 2020), and their use of protective equipment, mainly protective masks ("These days, are you using a protective face mask on a regular basis?"; "How many masks do you have now?"). We aggregated this information by submitting participants' answers to a principal component analysis (PCA), with the aim of discovering latent constructs that might summarize a participant's emotional appraisal of protective masks and other COVID-19-related measures.

TABLE 1 | Twenty-eight questions for PCA and their loading on the first component.

Question	Loading
Politically, would you describe yourself as left or right wing?	**−0.717**
I put on a mask when going into a store/supermarket/pharmacy	**0.713**
Do you think the danger of COVID-19 has been underestimated?	**−0.708**
Do you think the usefulness of protective face masks against COVID-19 infection has been underestimated?	**−0.707**
These days, are you using a protective face mask on a regular basis?	**0.671**
During a typical week, on how many days do you meet with people offline?	**−0.649**
On how many days in the past week have you met other people (offline)?	**−0.647**
I put on a mask every time I leave the house	**0.620**
When I'm outside, I have a mask with me	**0.592**
Politically, would you describe your close friends and family as left or right wing?	**−0.578**
Have you had COVID-19(coronavirus)?	**0.516**
These days, are you using hand sanitizer on a regular basis?	**0.458**
How often do you use a face mask?	**0.454**
Are you working from home?	**0.343**
How often are you checking the news on COVID-19?	−0.219
How often are you checking news unrelated to COVID-19?	0.106
Do you consider yourself to be especially vulnerable to COVID-19?	−0.258
I put on a mask even when I exercise	−0.012
Are you interested in politics?	−0.128
How often were you checking the news before the COVID-19pandemic?	0.046
Do you live in close contact with a person who is especially vulnerable to COVID-19?	−0.248
Do you enjoy debating political issues?	−0.233
Is a mask mandatory for your work?	0.429
I estimate the situation and only put the mask on when I think it is necessary	0.026
I put on a mask only when I am in a crowded place	0.220
Do you tend to be an anxious person?	0.250
How many masks do you have now?	0.077
Do you consider yourself an independent thinker?	0.058

Bold font denotes items that loaded on the first of the two components.

Table 1 shows the 28 questions that were included in the PCA, including questions on political views and interests as well as focused on COVID-19 (experiences, vulnerability, exposure, and protective practices). The overall dataset fits the requirement of PCA reasonably well, with Bartlett's test of sphericity, $\chi^2(378) = 1819$, $p < 0.001$, and a Kaiser-Meyer-Olkin measure of sampling adequacy equal to 0.75.

The first component accounted for 20.5% of the total variance and was readily interpretable as one's general attitude toward COVID-19, differentiating between participants who believed that the danger of COVID-19 was overestimated and those who thought it was underestimated. Participants' overall self-reported political orientation (right vs. left) loaded highly on this component, as did the question "Do you think the usefulness of protective face masks against COVID has been overestimated or underestimated?" It thus seems reasonable, for the purposes of the present study, to treat this latent variable as a scale ranging from strong anti-mask at one end to strong pro-mask at the other end.

We next computed each participant's mask attitude score using the regression method. These scores were symmetrically distributed around a value near zero with positive scores representing pro-mask attitudes, and negative scores representing anti-mask attitudes. For ease of presentation, we performed a median split on this score (Median = −0.046) and used it as a factor in ANOVAs to examine the mask attitudes of the viewers alongside the factors of the race of the face (White and Asian) and the type of partial occlusion (protective mask and notebook). We note that regression analyses, in which these scores, were treated as a continuous variable and the factors of race and occlusion were coded as dummy predictors, led to the same conclusions.

Analysis of the Attitude Survey Data

The Influence of Mask Attitudes on Ratings of Faces Occluded by a Mask

We first investigated the combined effects of mask attitudes and the race of the faces on participants who rated the attractiveness of faces shown with and without protective masks (Experiment 1). Because base attractiveness did not interact with race on mean rating differences scores (**Figures 3**, **5**), we averaged each participant's difference scores (the difference between attractiveness ratings for faces occluded with mask/notebook and fully visible faces) across attractiveness levels and examined it with an ANOVA in which mask attitude and race were the two factors. These scores are shown in **Figure 6A**.

The ANOVA showed a marginal main effect of race, $F(1,67) = 3.84$, $p = 0.054$, $\eta_p^2 = 0.054$, with own-race faces generally receiving higher scores than other-race faces. There was also a main effect of mask attitude, $F(1,67) = 10.84$, $p = 0.002$, $\eta_p^2 = 0.139$, with pro-mask participants giving higher scores than anti-mask participants. The interaction of these two factors was not significant, $F(1,67) = 0.43$, $p = 0.43$, $\eta_p^2 = 0.009$.

Single-sample t-tests were used to test for costs and benefits (i.e., differences from zero, p values and *CI*s are Bonferroni

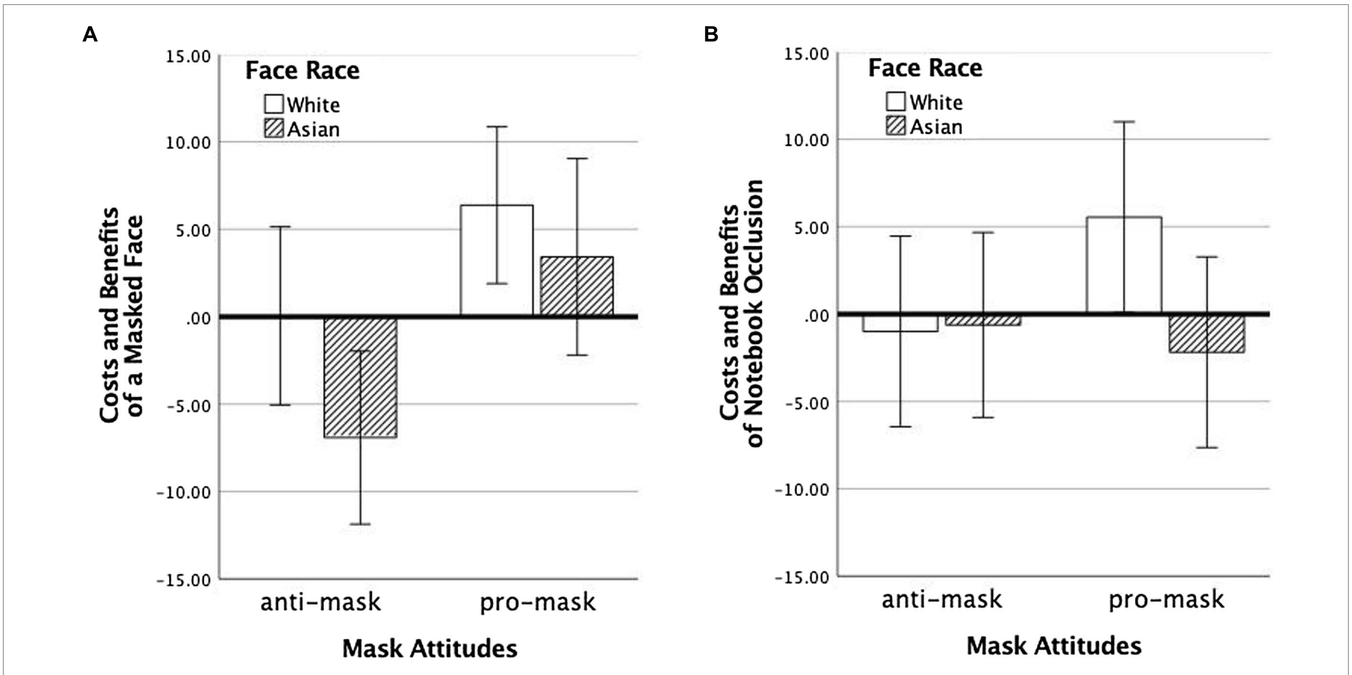


FIGURE 6 | Costs and benefits of an occluded face (computed as the difference between averaged ratings for occluded and full faces, across the three levels of baseline attractiveness) as a function of the viewer's attitudes to masks (horizontal axis) and the race of the target face (color of the bars). Panel **A** shows results for faces covered with masks; Panel **B** shows results for faces occluded by a notebook. Error bars represent +/− 1 standard error.

corrected) of the occlusion in these scores for each group of participants. Pro-mask participants showed positive biases in their scores for both own-race (White), 95% *CI* [1.83 10.92], $t(21) = 3.83$, $p = 0.004$, *Cohen's* $d = 0.817$, but not other-race (Asian) faces, 95% *CI* [−0.92 7.76], $t(13) = 2.28$, $p = 0.16$, *Cohen's* $d = 0.610$. Anti-mask participants, on the other hand, showed a trend toward negative bias for other-race faces, 95% *CI* [−15.60 1.77], $t(17) = -2.22$, $p = 0.16$, *Cohen's* $d = -0.524$, while for own-race faces, the scores were not different from 0, 95% *CI* [−8.95 9.04], $t(16) = 0.014$, $p > 0.36$, *Cohen's* $d = 0.003$.

This pattern of results suggests that when participants held positive attitudes toward masks, they also tended to give higher attractiveness ratings to all mask wearers, regardless of their race. When participants held negative attitudes toward masks, they did the reverse, tending to give lower attractiveness ratings to all people wearing them, regardless of their race. And the results show that this influence from the personal opinions of participant was additive in its effect with the race of the person wearing the mask. The practical implication of this pattern of results is that the decrease in attractiveness that derives from an anti-masking attitude, when combined with the general effect of a bias against masked Asian faces, means that half of the participants showed a negative bias toward mask wearers of the other race.

It should be noted that participants' attitudes toward masks were closely correlated with their political affiliations. Specifically, right-leaning participants generally disliked protective masks and had doubts about their efficiency. Therefore, it could be political affiliation that drives the bias against masked Asian faces observed here. Yet previous studies suggest that attitudes toward masks and esthetic judgments of them are affected not only by political affiliation (Dudarev et al., 2021), but by frequency of wearing the masks and situational exposure to COVID-19. This is why we chose to use the aggregate of all potential variables to assess participants' overall attitudes toward masks, instead of trying to distinguish among the possible determinants. The question of whether it is the political views alone or other aspects of mask attitudes cannot be addressed in this correlational data set, as these two issues are closely intertwined.

Another important question is whether attitudes toward protective masks are stable within an individual, and how changing attitudes over time affect the perception of masked faces. A previous study conducted by some authors in our group indicated that mask attitudes in a similar North American sample of participants did not change significantly within a three-month period during the height of the pandemic, from July to October, 2020 (Dudarev et al., 2021). However, it is reasonable to expect that changing circumstances or even merely a longer period of exposure to masked faces might be associated with change in the attitudes. Tracing the trajectory of changing attitudes toward masks on the perceived attractiveness of mask wearers is thus an important question for future studies to consider.

The Influence of Mask Attitudes on Ratings of Faces Occluded by a Notebook

We next examined the combined effects of mask attitudes and race of the faces for participants who rated faces that were partially occluded by notebooks (Experiment 2). These results are shown in **Figure 6B**. This analysis revealed no main effects and no interactions, all $p > 0.13$. The only significant single-sample t-test for the four groups of participants in Experiment 2 was for pro-mask participants, who showed a positive bias when viewing own-race faces, 95% *CI* [1.54 9.55], $t(15) = 3.93$, $p = 0.004$, *Cohen's d* $= 0.981$ (all other $p > 0.4$).

GENERAL DISCUSSION

The present study tested the consequences of wearing a protective mask on facial attractiveness during the worldwide COVID-19 pandemic. Our scientific interest in this question derives from the theory of microvalences, which proposes that our actions and perceptions are subtly influenced by the automatic emotional appraisal we give to objects (Lebrecht et al., 2012; Lebrecht and Tarr, 2012; Roos et al., 2013). These authors point out that even common, every-day objects are not emotionally neutral. Rather, the automatic emotional appraisal we attribute to objects, derived from our past experiences with them, influences the actions we make toward them. For example, we more readily approach positive objects and tend to avoid negative objects (Chen and Bargh, 1999; Kramer et al., 2020).

The specific focus of the present study was on potential interactions between the microvalenced appraisals people give to protective masks and the appraisals of people who wear them. To examine this question, our study looked at two factors that might mediate the effect of protective masks on facial attractiveness. One factor was the perceived similarity of the faces being rated to the participant (same- vs. other-race). We hypothesized that people who appeared to be more similar to the participant (same race) might gain an attractiveness benefit when donning a mask, compared to people who appeared to be less similar (other-race). The second factor was the participants' stance toward protective masks themselves. Wearing a protective mask became quite a contentious personal issue for many people in North America as the pandemic unfolded, with some people strongly identifying as pro-mask and others as anti-mask. It is not within the scope of this paper to reiterate the many factors leading to this societal divide. For the purpose of the present study—testing microvalence theory in a real-world context—it was simply convenient to be able to use these pre-existing individual differences in valence for masks as a way to examine object-person interactions in ratings of facial attractiveness.

Our practical interest in testing these hypotheses concerns the societal consequences of mask wearing and person perception. Given the rise in anti-minority sentiment, and specifically anti-Asian acts of violence and aggression that accompanied the pandemic in North America (Center for the Study of Hate and Extremism, 2021; Choi, 2021), it is important to better understand some of the factors that influence person perception in a health crisis such as the COVID-19 pandemic. Previous studies have documented that Muslim head and face coverings reduce the rated attractiveness of the wearer and the participant's empathy toward them (Saroglou et al., 2009; Mahmud and Swami, 2010; Unkelbach et al., 2010). Yet it is unclear whether these effects reflect stereotyping of an out-group (Saroglou et al., 2009; Mahmud and Swami, 2010) or the emotional valence of the head-coverings themselves. In any case, it is important to remind the reader that protective masks were worn for health reasons by people worldwide during the COVID-19 pandemic. This did not allow masks to act as stereotyping signal for specific religious or ethnic groups. Despite this, the present data show that masks have a disproportionate negative influence on the perception of minority individuals, and this negative influence is compounded by pre-existing negative attitudes regarding protective masks.

It will be an important question for future studies to try and identify the specific factors that are at play in the other-race effect we see here for mask wearing. These factors may range from having less sensitivity to the face parameters of other races (Rhodes et al., 2009), to failing to treat faces from other races with the individuality afforded to same-race faces (Roos et al., 2013), to processing other-race faces less holistically (Roisson and Michel, 2009), to biases that render other-race faces less attractive (Burke et al., 2013) or less relevant to ourselves (Scott and Monesson, 2009). Regardless of the underlying factors, the present finding makes it important for both health professionals and politicians to be aware of the consequences of taking decisions such as public mask mandates on minority individuals. At a minimum, a health-focused measure such as a mask mandate should also take into account the negative consequences of masks for minority wearers. More proactive thinking dictates that additional public measures (e.g., public messaging and empathy building media portrayals) should be taken to mitigate the negative socio-political consequences of health-related decisions.

Before turning to the present findings that speak most directly to microvalence theory, as it applies to protective masks, it is important to establish two important background effects that influence attractiveness ratings when the lower portion of a face is occluded. One of these effects is not mask-specific; the other is mask-specific. A non-mask-specific effect of lower face occlusion is that it truncates the range of attractiveness ratings that are obtained for the same faces when they are seen in full. This finding has been reported by two independent research groups (Patel et al., 2020; Kamatani et al., 2021) and it was replicated again in the present Experiment 2, where we used a notebook to occlude the lower portion of the faces. The effect manifests as a simple regression to the mean of all ratings, such that highly attractive faces are rated as less attractive with a mask, and unattractive faces are rated as more attractive. Our interpretation of this effect is that occlusion limits the available evidence concerning both extremes of attractiveness; hiding the imperfections that otherwise lead to lower ratings as well as the outstanding features that would generate the highest ratings if they were visible (Kamatani et al., 2021).

In this context, it is also of interest that the range of attractiveness ratings given to other-race full faces in this study were truncated compared to the range of ratings given to same-race full faces. This can be seen by comparing the range of attractiveness rating functions for White vs. Asian faces in

Figure 5. A possible interpretation of this finding is that when faces consist of features we are less familiar with, it becomes harder to assess the extreme ends of the attractiveness scale. Similar to visible occlusion, unfamiliar facial features limit the evidence available to participants for rendering a strong judgment in favor of either end of the scale.

It could be argued that this difference stems from the fact that Asian faces in this study had either neutral or slightly smiling expressions, while all the White faces study had strictly neutral expressions. It is indeed possible that the slight smiles in the Asian faces could have contributed to the slightly higher mean ratings of attractiveness given to them versus to White faces (see **Figure 2**). However, it is less likely that those smiles contributed to the reduced range of ratings given to Asian versus White faces by the white participants in our study (see results of Experiment 2 and **Figure 5**). And it is even more difficult to sustain the argument that those slight smiles were responsible for the specific effect of masks which manifested in reducing ratings at all three levels of baseline attractiveness for Asian faces when compared to White faces (the main effect seen in **Figure 3**).

A mask-specific effect that has been previously documented also provides important background for the present findings. A study conducted in pre-COVID-19 times by Miyazaki and Kawahara (2016) first documented that, in addition to the truncation of attractiveness ratings, protective masks contribute to an "unhealthiness priming" effect that decreases the wearer's attractiveness overall. That is, in addition to the occlusive effects of a mask, there was a specific health-related stigma that reduced attractiveness ratings for all faces, but especially faces in the medium and high range of the baseline attractiveness scale. This reduction did not extend to a control experiment where masks were replaced by health-neutral cards and notebooks. In 2020, the same group showed that health-related stigma disappeared, as during COVID-19 wearing a protective mask was not perceived as a sign of healthiness of the wearer (Kamatani et al., 2021).

These two previous findings—truncated ratings for occluded faces and a health stigma for protective masks—thus provide a critical backdrop for the two main present findings. First, there was a same-race benefit along with an other-race cost for mask wearers. Second, participants who identify as pro-mask gave an attractiveness benefit in their ratings to individuals of both same and other races wearing masks. This contrasted with anti-mask participants, who assigned an attractiveness cost specifically to the ratings of individuals from another race. We next consider each of these two main findings in turn.

The same-race benefit and the other-race cost of mask wearing on attractiveness ratings cannot be readily interpreted as merely another instance of the truncation of ratings for occluded and for less familiar faces. If that were the case, we would not expect to see such a large benefit in attractiveness ratings for both low and medium attractive faces that were masked (see **Figure 3**: White faces). Clearly, something more than regression to the mean is influencing these results for White viewers of White faces. Our initial hypothesis is that this reflects the generally positive evaluation of the mask during the pandemic "leaking" into the attractiveness ratings of the wearer. Similarly, something more than regression to the mean is influencing the cost of mask wearing for Asian faces (see **Figure 3**: Asian faces). Here too, the negative effects of masks go beyond simple regression to the mean. Now protective masks take on a negative stigma, perhaps more related to personal dissimilarity and a lack of empathy than to reduction in evidence for attractiveness. These two points are both supported when we consider the notebook control data in Experiment 2 (see **Figure 4**). Note that the regression to mean effect that accompanies partial face occlusion is more similar for same- and other-race faces, supporting the truncation in attractiveness ratings that accompanies the mere absence of evidence.

This interpretation of the other-race effect for mask wearers is only strengthened when we consider the mediation of these effects by a participant's pre-existing ideological stance regarding mask wearing (see **Figure 6A**). Participants who were pro-mask offered an attractiveness benefit in their ratings to individuals of both races while those who were anti-mask participants assigned an attractiveness cost specifically to the ratings of other-race faces. And again, this pattern was much reduced when the partial occlusion of the face was attributable to a non-COVID-19-related factor: notebooks (see **Figure 6B**). These findings clearly point to protective masks not being emotionally neutral objects in the context of person perception. The attractiveness of a face is clearly influenced, in both positive and negative ways, by the microvalences that have become associated with the objects associated with a face.

In summary, the present study demonstrates once again that protective masks were microvalenced for North American participants during the pandemic, both positively and negatively (Dudarev et al., 2021). However, the present study goes a step further in documenting that these object-based microvalences have important consequences for attractiveness ratings of a person wearing a mask. This finding surely advances our understanding of microvalence theory, showing for the first time that the emotional appraisal of objects encountered in the same context (here masks and people) can influence one another (Lebrecht et al., 2012). However, it also holds important background information for public officials in health and government charged with making decisions for the population as a whole. As a first step, the data show that the consequences of health-related decisions (i.e., mask mandates) have differential effects on mask wearers from majority and minority communities.

It will be important in future research to study cultural differences in the response to worldwide adoption of protective masks. Readers should note that while personal attitudes toward mask wearing are polarized in North America at this time, the same is not true in some Asian countries. For example, altruistic intentions to act collectively against the pandemic have strongly motivated mask wearing in Japan (Nakayachi et al., 2020). As such, the two factors shown to influence masked-face attractiveness ratings in the present study—an individual's similarity to the rated face and their attitudes toward masks—may not contribute to the perceived attractiveness of mask wearers in Japan. This certainly awaits further examination.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found at: https://zenodo.org/badge/latestdoi/394767599.

ETHICS STATEMENT

All study procedures were approved by Behavioral Research and Ethics Board at the University of British Columbia (approval number H21-00305). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MK and YM processed the photographs and created stimuli. VD programmed the study and collected and processed the data. VD, JTE, and JK analyzed the data. All authors collectively designed the study and contributed to the manuscript text.

FUNDING

VD's involvement in this research was funded by Mitacs Accelerate Postdoctoral fellowship. The research costs associated with this study were funded by a Discovery Grant to JTE from the Natural Sciences and Engineering Council of Canada. JK was funded by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (20H04568).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at https://www.frontiersin.org/articles/10.3389/fpsyg.2022.864936/full#supplementary-material

REFERENCES

Alter, A. L., and Oppenheimer, D. M. (2006). Predicting short-term stock fluctuations by using processing fluency. *Proc. Natl. Acad. Sci. U. S. A.* 103, 9369–9372. doi: 10.1073/pnas.0601071103

Aratani, L. (2020). How did face masks become a political issue in America? The Guardian. Available at: https://www.theguardian.com/world/2020/jun/29/face-masks-us-politics-coronavirus (Accessed April 26, 2022).

Beer, T. (2020). Anti-mask rallies continue in U.S. amid rising coronavirus cases and deaths. Forbes. Available at: https://www.forbes.com/sites/tommybeer/2020/07/16/anti-mask-rallies-continue-in-us-amid-rising-coronavirus-cases-and-deaths/#708d50d52246 (Accessed April 26, 2022).

Brielmann, A. A., Bülthoff, I., and Armann, R. (2014). 'Looking at faces from different angles: Europeans fixate different features in Asian and Caucasian faces. *Vision Res.* 100, 105–112. doi: 10.1016/j.visres.2014.04.011

Burke, D., and Sulikowski, D. (2013). The evolution of holistic processing of faces. *Front. Psychol.* 4:11. doi: 10.3389/fpsyg.2013.00011

Carragher, D. J., and Hancock, P. J. B. (2020). Surgical face masks impair human face matching performance for familiar and unfamiliar faces. *Cognit. Res. Princi. Imp.* 5:59. doi: 10.1186/s41235-020-00258-x

Center for the Study of Hate and Extremism (2021). *Report to the nation: Anti-Asian prejudice & hate crime*. Available at: https://www.csusb.edu/sites/default/files/Report%20to%20the%20Nation%20-%20Anti-Asian%20Hate%202020%20Final%20Draft%20-%20As%20of%20Apr%2030%202021%206%20PM%20corrected.pdf

Chen, M., and Bargh, J. A. (1999). Consequences of automatic evaluation: immediate behavioral predispositions to approach or avoid the stimulus. *Personal. Soc. Psychol. Bull.* 25, 215–224. doi: 10.1177/0146167299025002007

Chiao, Y. J., and Mathur, V. A. (2010). Intergroup empathy: how does race affect empathic neural responses?. *Curr. Biol.* 20, 478–480.

Choi, S. (2021). "People look at me like I AM the virus": fear, stigma, and discrimination during the COVID-19pandemic. *Qual. Soc. Work.* 20, 233–239. doi: 10.1177/1473325020973333

Dudarev, V., Manaligod, M. G., Enns, J. T., and Todd, R. M. (2021). In the hands of the beholder: wearing a COVID-19mask is associated with its attractiveness. *Q. J. Exp. Psychol.* 75, 598–615. doi: 10.1177/17470218211037128

Elfenbein, H. A., and Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychol. Bull.* 128, 203–235. doi: 10.1037/0033-2909.128.2.203

Freud, E., Stajduhar, A., Rosenbaum, R. S., Avidan, G., and Ganel, T. (2020). The COVID-19pandemic masks the way people perceive faces. *Sci. Rep. Nat.* 10, 22344–22348. doi: 10.1038/s41598-020-78986-9

Gandhi, M., Beyrer, C., and Goosby, E. (2020). Masks do more than protect others during COVID-19: reducing the inoculum of SARS-Cov-2 to protect the wearer. *J. Gen. Int. Med.* 35, 3063–3066. doi: 10.1007/s11606-020-06067-8

Howard, J., Huang, A., Li, Z., Tufekci, Z., Zdimal, V., der Westhuizen, V., et al. (2021). An evidence review of face masks against COVID-19. *Proceedings of the National Academy of Sciences of the United States of America.* 118:e2014564118. doi: 10.1073/pnas.2014564118

Institute for Health Metrics and Evaluation (2021, May 27). *COVID-19 Maps of Mask Use. Institute for Health Metrics and Evaluation*. Available at: https://www.healthdata.org/acting-data/covid-19-maps-mask-use

Kamatani, M., Ito, M., Miyazaki, Y., and Kawahara, J. I. (2021). Effects of masks worn to protect Against COVID-19on the perception of facial attractiveness, i-perception 12:20416695211027920. doi: 10.1177/20416695211027920

Karalis, K. (2020). No more masks: Hundreds attend anti-mask mandate rally in St. George. *ABC4*. Retrieved from: https://www.abc4.com/news/top-stories/no-more-masks-hundreds-attend-anti-mask-mandate-rally-in-st-george/ (Accessed April 26, 2022).

Knotek, E. S., Schoenle, R. S., Dietrich, A. M., Müller, G. J., Myrseth, K. O. R., and Weber, M. (2020). Consumers and COVID-19: Survey results on mask-wearing behaviors and beliefs. *Economic Commentary*, 1–7.

Kramer, R. S. S., Mulgrew, J., Anderson, N. C., Vasilyev, D., Kingstone, A., Reynolds, M. G., et al. (2020). Physically attractive faces attract us physically. *Cognition* 198:104193. doi: 10.1016/j.cognition.2020.104193

Lakens, D., and Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Adv. Meth. Pract. Psychol. Sci.* 4:2515245920951503.

Lebrecht, S., Bar, M., Barrett, L. F., and Tarr, M. J. (2012). Micro-valences: perceiving affective valence in everyday objects. *Front. Psychol.* 3:107. doi: 10.3389/fpsyg.2012.00107

Lebrecht, S., Bar, M., Sheinberg, D. L., and Tarr, M. J. (2011). Micro-valence: nominally neutral visual objects have affective valence. *J. Vis.* 11:856. doi: 10.1167/11.11.856

Lebrecht, S., and Tarr, M. J. (2012). Can neural signals for visual preferences predict real-world choices? *Bio Sci.* 3, 937–938. doi: 10.3389/fpsyg.2012.00107

Levin, D. T. (2000). Race as a visual feature: using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *J. Exp. Psychol. Gen.* 129:559.

Lyu, W., and Wehby, G. L. (2020). Community use of face masks and COVID-19: evidence from a natural experiment of state mandates in the US. *Health Aff.* 39, 1419–1425. doi: 10.1377/hlthaff.2020.00818

Ma, C., and Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behav. Res. Methods* 47, 1122–1135. doi: 10.3758/s13428-014-0532-5

Mahmud, Y., and Swami, V. (2010). The influence of the hijab (Islamic head-cover) on perceptions of women's attractiveness and intelligence. *Body Image* 7, 90–93. doi: 10.1016/j.bodyim.2009.09.003

Miller, A. M., and Brueck, H. (2020). Anti-maskers are the new ati-vaxxers. *Business Insider*. Retrieved from: https://www.businessinsider.com/anti-maskers-are-new-anti-vaxxers-threatening-public-health-coronavirus-2020-7 (Accessed April 26, 2022).

Miyazaki, Y., and Kawahara, J. I. (2016). The sanitary-mask effect on perceived facial attractiveness. *Jpn. Psychol. Res.* 58, 261–272. doi: 10.1111/jpr.12116

Molnar-Szakacs, I., Uddin, L. Q., and Heffernan, M. B. (2021). The face behind the mask: The future of interpersonal interaction. *Neuron* 109, 1918–1920. doi: 10.1016/j.neuron.2021.05.030

Nakayachi, K., Ozaki, T., Shibata, Y., and Yokoi, R. (2020). Why do Japanese people use masks against COVID-19, even though masks are unlikely to offer protection from infection? *Front. Psychol.* 11:1918. doi: 10.3389/fpsyg.2020.01918

Patel, V., Mazzaferro, D. M., Sarwer, D. B., and Bartlett, S. P. (2020). Beauty and the Mask. *Plastic Reconstr. Surgery Global Open* 8, 1–4. doi: 10.1097/GOX.0000000000003048

Rader, B., White, L. F., Burns, M. R., Chen, J., Brilliant, J., Cohen, J., et al. (2021). Mask-wearing and control of SARS-CoV-2 transmission in the USA: A cross-sectional study. *The Lancet. Digital Health* 3, e148–e157. doi: 10.1016/S2589-7500(20)30293-4

Rhodes, G., Locke, V., Ewing, L., and Evangelista, E. (2009). Race coding and the other-race effect in face recognition. *Perception* 38, 232–241. doi: 10.1068/p6110

Roos, L. E., Lebrecht, S., Tanaka, J. W., and Tarr, M. J. (2013). Can singular examples change implicit attitudes in the real-world? *Front. Psychol.* 4:594. doi: 10.3389/fpsyg.2013.00594

Rossion, B., and Michel, C. (2011). "An experience-based holistic account of the other-race face effect" in *The Oxford Handbook of Face Perception*. eds. A. Calderet al. (Oxford: Oxford University Press), 215–244.

Saroglou, V., Lamkaddem, B., van Pachterbeke, M., and Buxant, C. (2009). Host society's dislike of the Islamic veil: The role of subtle prejudice, values, and religion. *Int. J. Intercult. Relat.* 33, 419–428. doi: 10.1016/j.ijintrel.2009.02.005

Scott, L. S., and Monesson, A. (2009). The origin of biases in face perception. *Psychol. Sci.* 20, 676–680. doi: 10.1111/j.1467-9280.2009.02348.x

Tanaka, J. W., Heptonstall, B., and Hagen, S. (2013). Perceptual expertise and the plasticity of other-race face recognition. *Vis. Cog.* 21, 1183–1201.

Todd, R. M., and Manaligod, M. G. M. (2018). Implicit guidance of attention: The priority state space framework. *Cortex* 102, 121–138. doi: 10.1016/j.cortex.2017.08.001

Unkelbach, C., Schneider, H., Gode, K., and Senft, M. (2010). A turban effect, too: selection biases Against women wearing Muslim headscarves. *Soc. Psychol. Personal. Sci.* 1, 378–383. doi: 10.1177/1948550610378381

Whitman, J. C., Zhao, J., Roberts, K. H., and Todd, R. M. (2018). Political orientation and climate concern shape visual attention to climate change. *Clim. Change* 147, 383–394. doi: 10.1007/s10584-018-2147-9

Wickens, T. D., and Keppel, G. (2004). Design and analysis: A researcher's handbook. Upper Saddle River, NJ: Pearson Prentice-Hall.

Wright, L., Steptoe, A., and Fancourt, D. (2020). What predicts adherence to COVID-19government guidelines? Longitudinal analyses of 51,000 UK adults. medRxiv [Preprint]. 1–19. doi: 10.1101/2020.10.19.20215376

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: 26 May 2022
doi: 10.3389/fpsyg.2022.880548

A PCA-Based Active Appearance Model for Characterising Modes of Spatiotemporal Variation in Dynamic Facial Behaviours

David M. Watson[1,2]* *and Alan Johnston*[1]

[1] *School of Psychology, University of Nottingham, Nottingham, United Kingdom,* [2] *Department of Psychology, University of York, York, United Kingdom*

OPEN ACCESS

Edited by:
Koyo Nakamura,
University of Vienna, Austria

Reviewed by:
Masashi Komori,
Osaka Electro-Communication University, Japan
Jiayu Zhan,
University of Glasgow, United Kingdom

****Correspondence:***
David M. Watson
david.watson@york.ac.uk

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

Received: *21 February 2022*
Accepted: *22 April 2022*
Published: *26 May 2022*

Citation:
Watson DM and Johnston A (2022) A PCA-Based Active Appearance Model for Characterising Modes of Spatiotemporal Variation in Dynamic Facial Behaviours. Front. Psychol. 13:880548. doi: 10.3389/fpsyg.2022.880548

Faces carry key personal information about individuals, including cues to their identity, social traits, and emotional state. Much research to date has employed static images of faces taken under tightly controlled conditions yet faces in the real world are dynamic and experienced under ambient conditions. A common approach to studying key dimensions of facial variation is the use of facial caricatures. However, such techniques have again typically relied on static images, and the few examples of dynamic caricatures have relied on animating graphical head models. Here, we present a principal component analysis (PCA)-based active appearance model for capturing patterns of spatiotemporal variation in videos of natural dynamic facial behaviours. We demonstrate how this technique can be applied to generate dynamic anti-caricatures of biological motion patterns in facial behaviours. This technique could be extended to caricaturing other facial dimensions, or to more general analyses of spatiotemporal variations in dynamic faces.

Keywords: dynamic faces, facial caricaturing, ambient faces, computational neuroscience, face perception

INTRODUCTION

Faces provide a wealth of information about people including their identity (Ellis, 1975), social traits (Oosterhof and Todorov, 2008), and emotional state (Bruce and Young, 1986; Calder and Young, 2005). Faces encountered in the real world are often dynamic and highly variable. Despite this, much research to date has employed static images of faces taken under tightly controlled conditions. Such images risk controlling away potentially important sources of variation, such as within-person variability (Jenkins et al., 2011). Furthermore, both behavioural (Lander and Butcher, 2015) and neurological (O'Toole et al., 2002; O'Toole and Roark, 2010; Bernstein and Yovel, 2015) evidence supports a processing advantage for dynamic over static faces, indicating that dynamic faces convey information that static faces do not.

Facial caricatures present a common method for studying key dimensions of facial variation (Benson and Perrett, 1991b), by either increasing (caricaturing) or decreasing (anti-caricaturing) differences in facial features between an exemplar and reference face, where the reference is typically a neutral or average face. Such methods have been applied to static images to study facial features underlying the perception of identity (Benson and Perrett, 1991a; Blanz et al., 2000; Leopold et al., 2001; Jiang et al., 2006), age (Burt and Perrett, 1995), and emotional expressions (Calder et al., 1997, 2000; Juricevic and Webster, 2012). Caricaturing dynamic faces poses a further challenge as the process must account for both spatial and temporal patterns of variation. Previous approaches have manipulated the magnitude of motion in facial landmarks during simple facial behaviours, using the resulting motion vectors to drive virtual head models (Hill et al., 2005; Furl et al., 2020, 2022). However, such artificial head models lack many of the features present in real faces. A method for dynamically caricaturing facial behaviours in natural videos is therefore still lacking.

In a recent study (Watson et al., 2020), we developed a paradigm for eliciting dynamic and natural facial behaviours by video recording subjects while they delivered short sentences conveying either good or bad news. Speech patterns such as these provide a key source of non-rigid motion in the face. Using a principal component analysis (PCA)-based active appearance model, along with machine learning, we were able to reconstruct a behaviourally interpretable dimension of emotional valance from the facial behaviours. This technique considered motion information in the sense that it included changes in shape and texture over time. However, it did not consider the temporal structure of this information as each frame was simply represented as an independent sample within the model, and the order of the frames was ignored. It thus remains unclear whether such a model is able to capture more nuanced patterns of temporal variation.

Here we present an alternative application of our previous methods (Watson et al., 2020) that aims to model patterns of both spatial and temporal variation within natural facial behaviours. We demonstrate how this technique can be used to create dynamic anti-caricatures of biological motion patterns by morphing between an exemplar and an average timeseries of facial behaviours evoked during speech movements. As before, we initially use a PCA-based active appearance model to capture modes of spatial variation in the face over time. We then use an established dynamic time warping algorithm to align the PCA timeseries over clips. Finally, we present a novel method for capturing patterns of temporal variation by submitting the PCA timeseries to a further second-order PCA. This second-order PCA space represents deviations between exemplar clips and the average first-order PCA timeseries. Weighting and then back-projecting samples from this space yields anti-caricatured videos that vary in terms of their spatial and temporal deviations from the average timeseries.

METHODS

The datasets and some of the methods presented here have previously been described in Watson et al. (2020).

Recordings

Three subjects (two females, one male, and age range 26–42) were video recorded. The study was approved by the Ethics Committee of the School of Psychology at the University of Nottingham (Ethics approval number: 717) and conducted in accordance with the guidelines and regulations of this Committee and the Declaration of Helsinki. All subjects provided informed written consent to take part in the study and for their likeness to be used in publication.

Subjects were recorded against a uniform visual background in an anechoic chamber. Recordings were made with a Sony HXR-NX5U NXCAM camera connected to an Atomos Ninja-2 recorder that recorded videos in Apple ProRes RAW format. Videos were acquired at a resolution of 1,920 × 1,080 pixels and at 25 fps with a 6.67 ms exposure. Audio was recorded at a 48 kHz sampling rate. Videos were then encoded using MPEG-4 lossless compression prior to further processing.

Each subject delivered multiple repeats of 20 unique phrases, each conveying either good or bad news (10 unique phrases within each type). A list of the phrases is provided in **Supplementary Table 1**. Subjects 1 and 2 performed 15 repeats of each phrase (300 total), and Subject 3 performed 16 repeats (320 total). Subjects were not told to pose any specific expressions or behaviours; instead, they were instructed to simply deliver the phrases in whatever manner felt most natural to them. While delivering the phrases, subjects viewed silent videos of putative recipients presented on a teleprompter directly in front of the camera. Recipient videos showed video-conference style calls obtained from YouTube and helped give subjects the impression of having a person listen to them while they delivered their phrases.

Video Pre-processing

Each phrase repeat was then clipped to just the common prefix portion of each phrase ("Good news . . ." or "I'm sorry to say . . ."), excluding the later variable suffix portions (e.g., ". . . the operation went well!" or ". . . we're going to have to let you go"). The Google Cloud Speech-to-Text algorithm[1] was used to generate timestamps for each word in each phrase, which were then used to define onsets and offsets for each prefix portion. Onsets were adjusted 200 ms before the first word onset so as to include facial movements commencing immediately prior to the vocalisation. Manual corrections were applied where necessary. Clips varied in duration because the length of each vocalisation could differ over repeats.

Each clip was then cropped to a square region around the face. A Haar cascade face-detection algorithm implemented in OpenCV[2] extracted the position of the face within the scene on each frame. A square bounding box was then defined around the

[1] https://cloud.google.com/speech-to-text

[2] https://opencv.org/

average face position, allowing for a small border around the face. This ensured the face was placed approximately centrally within the scene. Each clip was then down-sampled to a resolution of 128×128 pixels via an anti-aliasing filter.

Multi-Channel Gradient Model and First-Order Principal Component Analysis

An overview of the remaining processing pipeline is illustrated in **Figure 1**. We employed a two-frame version of the Multi-channel Gradient Model (McGM; Johnston et al., 1992, 1999; Cowe, 2003) to capture shape and texture changes in the face over frames. This model has previously been shown to capture key dimensions of facial variation including gender (Griffin et al., 2011), speech movements (Scholes et al., 2020), and emotional valence (Watson et al., 2020), and can identify critical facial features for image reconstruction (Berisha et al., 2010). For each frame within the cropped and down-sampled clips, a warp vector field was calculated to register the frame to a standard reference image. The reference image was initially defined by an individual frame extracted from one of the recordings but was then replaced with the average of all textures after warping. This process was iterated three times, recalculating the warps and replacing the reference image with the average warped textures each time: this provided a more standardised final reference image. For each frame, the McGM yields a 5-channel image comprising the x- and y-direction warp components needed to warp the original image to the final average reference image, plus a "shape-free" version of the RGB textures after warping to the reference. These images were then vectorised and stacked over frames and clips, such that each frame in each clip is represented as an independent sample within an 81,920-dimensional ($128 \times 128 \times 5$) feature space defined by the pixels of the McGM images. Each clip is therefore represented by a high-dimensional multivariate timeseries within the McGM space. Note that the use of the McGM here for image registration (Cowe, 2003) differs from some previous applications that instead used the same model to measure local image velocities (Johnston et al., 1992, 1999).

Samples within the McGM space were then split between the phrase-types ("Good news" and "I'm sorry to say"), and each phrase-type was processed separately thereafter. The dimensionality of the McGM space was reduced via principal component analysis (PCA). All available components were retained (one fewer than the total number of frames over all clips within the given phrase-type). Because the number of samples is less than the number of McGM features, this allows a lossless PCA where 100% of variance remains explained while still reducing the dimensionality. Whereas a lossy PCA can reduce the dimensionality further, the lossless PCA was chosen to best preserve the fidelity of individual clips. This produced the first-order PCA (PCA_1) space, in which samples are given by the frames over all the clips (within the given phrase-type), and dimensions are given by the first-order principal components. Each clip is thus represented by a multivariate timeseries within the PCA_1 space. The first-order principal components encode modes of common spatial variation amongst the McGM image pixels, but do not consider the temporal order of this information.

Temporal Alignment

Next, we temporally aligned the PCA_1 timeseries over the clips within each phrase-type independently. Although all clips were cut to the initial prefix portion of each phrase, the onset of each vocalisation won't necessarily occur at the same time point within each clip, and the duration of each vocalisation may vary over repeats. We based the temporal alignment on the audio streams as they provide a precise measure of the temporal evolution of each vocalisation and show a good correspondence over repeats.

Audio streams were averaged over stereophonic channels to produce a single monophonic audio signal for each clip. As we are primarily interested in aligning modulations in the audio amplitudes, we applied a Hilbert transform to extract the amplitude envelopes from the audio signals. The original cuttings of each clip include a brief period prior to the onset of the first word to capture any facial movements commencing immediately prior to the start of the vocalisation. However, these periods are largely silent and so lack any consistent amplitude modulations on which to base a temporal alignment. We therefore re-cut the onset of each audio stream to lie closer to the onset of the actual vocalisation. This was done by identifying the timepoint of the initial rise in audio amplitude in the amplitude envelope. This timepoint was then adjusted to 80 ms (two video frames) prior to this to allow a small margin of error prior to the audio onset, and then rounded to the timepoint of the nearest video frame onset (i.e., to the nearest 40 ms). Manual corrections were applied where necessary.

The re-cut audio streams were then temporally aligned using a dynamic time warping (DTW) algorithm implemented using the *dtw-python* package (Giorgino, 2009; Tormene et al., 2009)[3]. For purposes of computational tractability, the audio amplitude envelopes were down-sampled by a factor of 10 (yielding an effective sampling rate of 4.8 kHz) using scipy's *decimate* function. The down-sampled envelopes for each clip were then aligned to a common reference envelope (**Figure 2A**). The reference was initially selected as the individual clip closest matching the median duration over all clips. However, to provide a more standardised reference, the reference envelope was then replaced with the average envelope over all clips after temporal alignment, and the DTW was recomputed for the new reference. This process was iterated three times, updating the average reference envelope each time, to allow the reference to stabilise. To minimise extraneous effects of global amplitude differences irrelevant to the temporal alignment, all envelopes (including the reference) were rescaled to have an L^2-norm equal to one on every iteration. The DTW was computed using an asymmetric step pattern and allowed open beginnings and ends so that neither the first nor last sample need be matched exactly.

[3] https://dynamictimewarping.github.io/

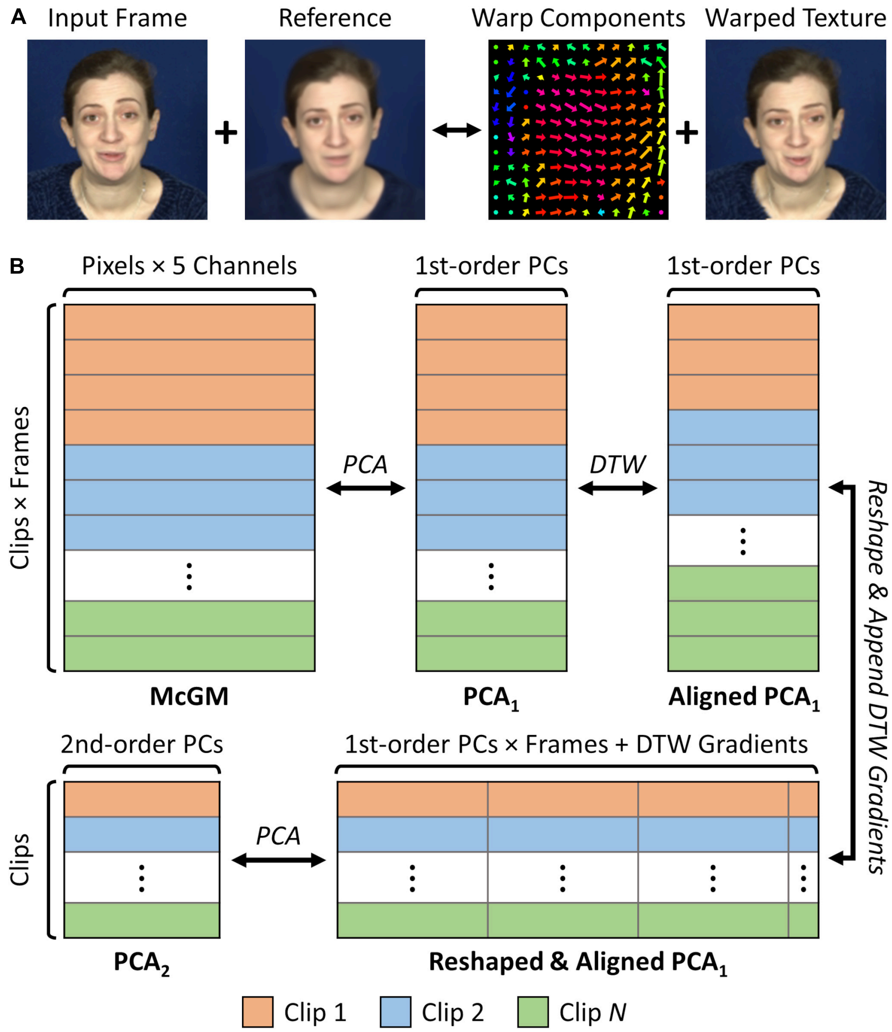


FIGURE 1 | Schematic illustration of processing pipeline. **(A)** Multi-channel Gradient Model (McGM) registration process. Each input frame is warped to an average reference image, producing a 5-channel image comprising the *x*- and *y*-warp components plus the RGB warped textures. **(B)** Principal Component Analysis (PCA) pipeline. McGM outputs are vectorised such that each frame is represented as a sample in a high dimensional space. This space is reduced via lossless PCA, yielding the PCA_1 space. Clips are temporally aligned by interpolating PCA_1 scores between frames based on dynamic time warping (DTW) of the audio streams. Aligned PCA_1 scores are vectorised within each clip and combined with the dynamic time warping gradients, such that each clip is represented as a sample in a new high dimensional space. This space is then reduced by a second lossless PCA, yielding the PCA_2 space. The pipeline is fully reversible, such that samples at any stage can be back-projected to the image space.

The resulting warping curves can be used to align the audio streams. To align the video frames, the audio warping curves were down-sampled to the resolution of the video sampling rate (25 fps) via linear interpolation (**Figure 2B**). The audio alignment excludes the period before the vocalisation onset; however, some facial movements may still occur during this period. We therefore extended the video warping curves at a 45° angle (i.e., matching clip and reference frames one-for-one) for a further four video frames (160 ms) prior to the vocalisation onset. Although this may only achieve a moderately accurate temporal alignment, it is nevertheless preferable to include the visual information from this initial period than to exclude it. The final video warping curves were then used to apply a linear interpolation between frames of the PCA_1 timeseries for each clip. Following this, each clip is still represented by a multivariate timeseries in the PCA_1 space, however, all timeseries will now be the same length and should be aligned in time (within each phrase-type).

Second-Order Principal Component Analysis and Caricaturing

The temporally aligned PCA_1 scores were vectorised within each clip. In addition, the gradients of the video time warping curve were appended to the end of each vector: this allows incorporating information about the temporal scale of the behaviours and permits the DTW curve to be reconstructed from the second-order PCA space. This generated a new high-dimensional feature space in which each clip is represented as

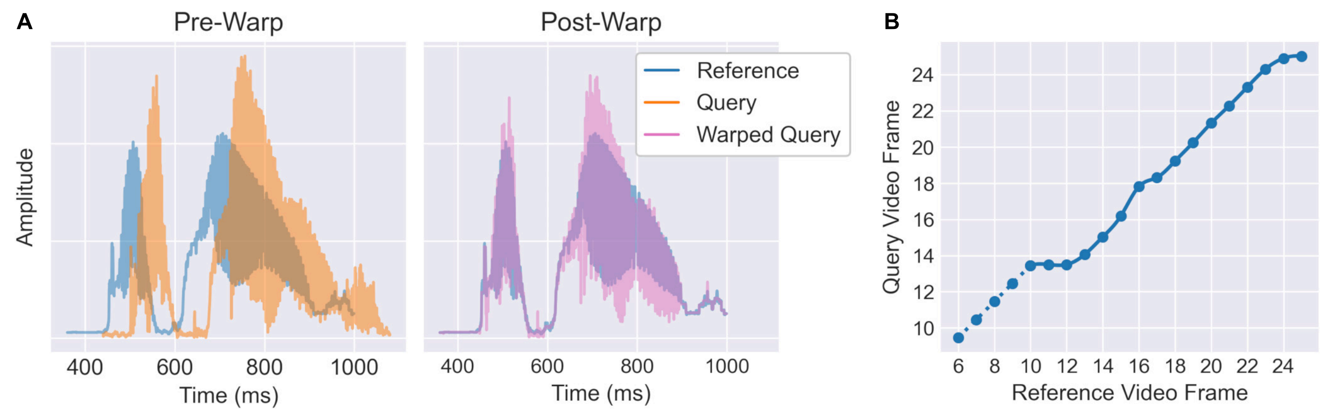


FIGURE 2 | Illustration of temporal alignment procedure for an example "Good news" clip. **(A)** Audio amplitude envelopes for each clip are aligned to an average reference envelope via dynamic time warping. **(B)** The resulting warping curve is down-sampled to the video sampling rate such that it can be applied to the PCA_1 timeseries. The interpolated curve is extended at a 45° angle for 4 video frames prior to the audio onset (dotted line) so that facial movements commencing shortly before the vocalisation are also included.

an independent sample, and where the dimensions comprise the concatenation of first-order principal components over video frames plus the DTW gradients. The dimensionality of this space was then reduced by a further lossless PCA. As before, all available components were retained (one fewer than the number of clips), such that 100% of variance remained explained after the dimensionality reduction. This produced the second-order PCA (PCA_2) space, in which each clip is a sample and the dimensions are given by the second-order principal components. Components within this space can reflect interactions between first-order principal components and timepoints and can therefore encode both spatial and temporal modes of facial variation. Whereas points within the PCA_1 space represent individual images, points within the PCA_2 space represent full temporal sequences. The origin of the PCA_2 space represents the PCA_1 timeseries averaged over clips, and individual samples are represented in terms of their spatiotemporal deviations from this average timeseries.

To produce dynamic anti-caricatures, we weighted the loading of a given sample/clip within the PCA_2 space. A weighting of zero will reduce the representation to the origin, thereby reproducing the average PCA_1 timeseries. A weighting of one will return a representation of the original clip. Intermediate weightings will yield varying levels of anti-caricature, weightings greater than one will yield caricatures, and negative weightings will yield anti-faces. For a given weighting, the resulting PCA_2 sample was back-projected to the reshaped PCA_1 space. This produced a reconstruction of the vectorised PCA_1 scores plus the DTW gradients. The reconstructed vectorised PCA_1 scores were then "un-vectorised" to return the sample to the temporally aligned PCA_1 space (with frames as samples and first-order principal components as dimensions). Meanwhile, the reconstructed DTW gradients were used to generate a time warping curve, which was in turn then used to apply a linear interpolation to the "un-vectorised" reconstructed PCA_1 scores that returned them to the original timescale. From here, the reconstructed PCA_1 scores were back-transformed to the McGM space. Finally, the reconstructed McGM warp components were inverted to spatially unwarp the reconstructed textures back to the image space. To aid visualisation, the visual contrast of the images was enhanced via unsharp masking. The complete back-projection of a given point from within the PCA_2 space therefore yields a full video animation within the image space.

Perceptual Ratings

We conducted a behavioural experiment to quantify the effect of the caricaturing on human perception of biological motion in the videos. Ten participants took part in the experiment (three male, seven female, and age range 23–36). The study was approved by the Ethics Committee of the School of Psychology at the University of Nottingham (Ethics approval number: F1249) and conducted in accordance with the guidelines and regulations of this Committee and the Declaration of Helsinki. Participants provided informed consent via an electronic form prior to participation. The experiment was run online using PsychoPy and Pavlovia (Peirce et al., 2019)[4]. To avoid confusion, for the analysis of the behavioural data we refer to the participants in this experiment as "raters" and the participants in the original video recordings as "recording subjects."

Raters were shown video clips across 5 levels of anti-caricaturing (0, 0.25, 0.5, 0.75, and 1) for the "I'm sorry to say" phrases. We avoided using weightings outside the zero to one range (caricatures and anti-faces) as these are more prone to image distortions that could confound the task. We also omitted the "Good news" phrases as these typically have relatively short durations which would make the task unduly challenging. Each rater was shown a 10% subset of clips across all three recording subjects (15 unique clips for S1 and S2, 16 unique clips for S3), such that ratings were provided for all clips across the 10 raters. Across the 5 caricaturing levels, each rater therefore completed 230 trials. The trials were split into three blocks with each recording subject presented continuously throughout

[4]https://pavlovia.org/

a block; this was done to aid raters in generating an internal standard of each recording subject's range of facial behaviours. The order of blocks was randomised for each rater. An additional shorter practice block was included at the start of the experiment, comprising the 5 anti-caricaturing levels for an example clip from each recording subject (15 trials total). To avoid priming responses, practice clips were selected from a different subset of clips to the main trials.

Raters were informed that they would view a series of silent videos showing people saying a short phrase, and that the people might appear livelier and more dynamic in some videos, and less so in others. A more precise definition of dynamicity was deliberately omitted so as to encourage raters to form their own interpretation. On each trial, raters viewed the video clip and were then asked to rate it for how "dynamic" the person appeared to be. Raters made their responses on a 5-point Likert-scale with the labels: "Not at all," "Not much," "A bit," "Fairly," and "Very." The responses were entered into a mixed effects ordinal logistic regression implemented using the *ordinal* package in R (Christensen, 2019)[5]. The caricaturing level (0, 0.25, 0.5, 0.75, and 1) was entered as the predictor variable, while the dummy coded ratings (1–5) were entered as the outcome variable. Variable intercepts were allowed over raters (R1–R10) and recording subjects (S1–S3); a more complicated model allowing variable slopes failed to converge. The slope parameter (β_1) represents the log odds of giving a higher versus lower dynamicity rating given a one unit increase in the caricaturing level. If the slope is significantly greater than zero, this would indicate that increasing the caricature level leads to a significant increase in the likelihood of raters providing a higher rather than lower dynamicity rating. We applied an alpha criterion of 0.05 for determining statistical significance.

RESULTS

Three subjects were video recorded while delivering a series of phrases conveying either positive ("Good news") or negative ("I'm sorry to say") news, eliciting dynamic and natural facial behaviours in the form of speech patterns. A two-frame version of the Multi-channel Gradient Model (McGM; Johnston et al., 1992, 1999; Cowe, 2003) was used to register the facial textures in each frame to a common average reference frame (**Figure 1A**). Each frame is then represented by a 5-channel image comprising the *x*- and *y*-direction warp components plus a "shape-free" version of the RGB textures. The remaining processing pipeline is illustrated in **Figure 1B**. This pipeline is applied within each phrase-type ("Good news" and "I'm sorry to say") and for each subject independently. First, the McGM outputs were vectorised such that each frame is represented as a sample within a high-dimensional feature space. We then reduced the dimensionality of this space using a lossless Principal Components Analysis (PCA), retaining all available components so that 100% of the variance remains explained. The resulting feature space is hereafter referred to as the first-order PCA (PCA_1) space. Components within this space encode principal modes of spatial variation in facial shape and texture over frames (Turk and Pentland, 1991), but do not consider the temporal order of such changes. Visualisations of the features encoded by the early principal components are shown for an example dataset in **Figure 3**, and animations for all datasets are shown in **Supplementary Video 1**. The first principal component typically encodes global changes over clips such as seating position or a change in clothing; exaggerations of this component often induce distortions in the image. Later components encode more salient modes of facial variation, including both rigid changes in the head position and non-rigid changes in the internal facial features such as in the shape of the mouth or opening of the eyes.

At this point, each clip is represented by a multivariate timeseries within the PCA_1 space. However, these timeseries are not temporally aligned over clips and may be different durations. To capture common patterns of temporal variation, it is therefore necessary to temporally align the PCA_1 timeseries. We based the alignment on the audio streams within each clip as these provide a precise index for the temporal evolution of each vocalisation. We used a dynamic time warping (DTW) algorithm to temporally align the audio envelopes, then down-sampled the resulting warping curves to the video sampling rate and used these to interpolate the PCA_1 timeseries between frames in each clip (**Figure 2**). Following time warping, all PCA_1 timeseries are identical in duration and temporally aligned over clips. **Figure 4** illustrates cross-sections through the first ten aligned PCA_1 timeseries averaged over clips. Clear modulations are present in each component timeseries, indicating both that common patterns of temporal variation are present in the PCA_1 scores and that the time warping procedure was successful in aligning these. This can be further illustrated by back-projecting the average PCA_1 timeseries to the image space. The resulting animations (**Supplementary Video 2**) maintain clear depictions of the phrases ("Good news" or "I'm sorry to say") being spoken.

We next extracted common patterns of temporal variation from the PCA_1 timeseries. The temporally aligned PCA_1 scores were vectorised within each clip independently and concatenated together with the DTW gradients for each clip. Including the DTW gradients allows incorporating information about the temporal scale of the behaviours and ensures that back-projections through this space include the necessary information to "unwarp" the corresponding PCA_1 timeseries back to their original timescale. This produces a new high-dimensional feature space, in which each clip is represented as a sample, and where the dimensions are defined by the combination of PCA_1 components and timepoints plus the DTW gradients. This space was then reduced via a further lossless PCA, again retaining all available components such that 100% of variance remains explained: this yields the second-order PCA (PCA_2) space. Components in the PCA_1 space capture principal modes of spatial variation in the faces, but without regard to the temporal order of those changes. By contrast, components in the PCA_2 space capture patterns of temporal variation amongst the first-order principal components. Whereas each point within the PCA_1 space represents an individual image, each point within the PCA_2 space represents a full temporal trajectory and can be visualised as a video if

[5]https://github.com/runehaubo/ordinal

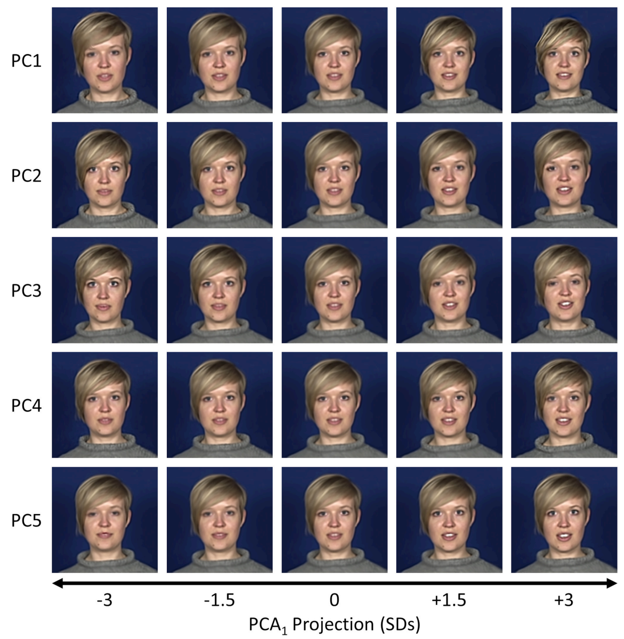


FIGURE 3 | Visualisations of the first five components from the PCA_1 space for an example dataset (S3, "Good news"). Samples were selected at positions between ± 3 standard deviations along a given principal component axis and back-projected to the image space.

back-projected to the image space. The origin point of the PCA_2 space represents the average PCA_1 timeseries (**Figure 4** and **Supplementary Video 2**), and individual clips are represented within the PCA_2 space in terms of their spatiotemporal deviations from this average timeseries.

To visualise this more clearly, we back-projected samples at varying positions along the first five components of the PCA_2 space. Still images from the final frame in each sequence for an example dataset (Subject 3, "I'm sorry to say") are shown in **Figure 5**, and animations of the full sequences are shown in **Supplementary Videos 3–8**. Similar to the PCA_1 space, early components encode global shape changes and exaggerating them often causes distortions in the image. Patterns of facial variation encoded in later components tend to be more subtle than those observed in the PCA_1 space (cf. **Figure 3**) but are still evident. For example, in Subject 3's "I'm sorry to say" dataset, PCs 1 through 4 include differences in the head tilt, while PC 5 modulates the vertical head and jaw position.

Dynamic anti-caricatures and caricatures (Leopold et al., 2001) of individual clips can be produced by multiplying a given clip's loadings within the PCA_2 space and back-projecting to the image space. A weighting of zero will reduce to the origin of the PCA_2 space and hence reproduce the average PCA_1 timeseries. A weighting of one will reproduce the timeseries of the original clip. Intermediate weightings will produce varying levels of anti-caricature between the individual and average timeseries. Weightings greater than one will produce caricatures that exaggerate the deviations between the individual and average timeseries (Benson and Perrett, 1991b), while negative weightings will produce anti-faces that invert the deviations (Leopold et al., 2001). Back-projecting a weighted sample to the image space produces a dynamic (anti-)caricatured video. **Figure 6** shows still images from anti-caricatured and caricatured sequences for an example clip from Subject 3's "I'm sorry to

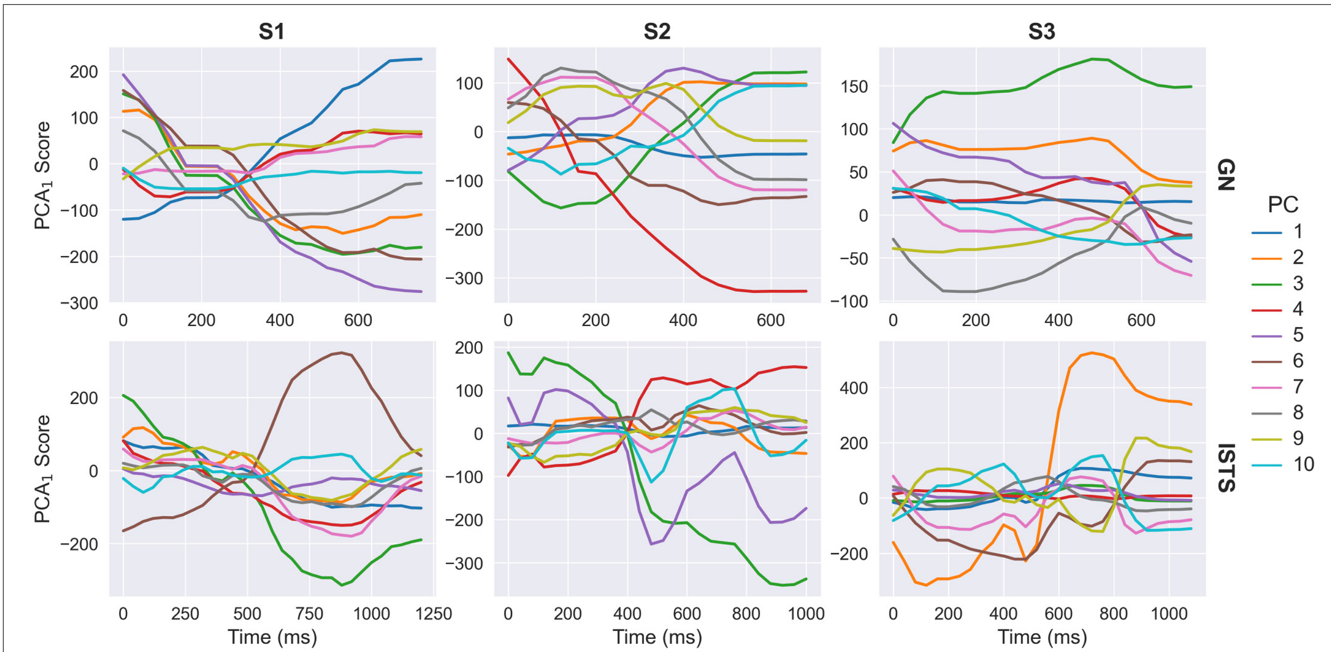


FIGURE 4 | PCA_1 timeseries averaged over clips following temporal alignment. Plots illustrate cross-sections through the first ten principal components for each subject and phrase-type (GN, "Good news"; ISTS, "I'm sorry to say").

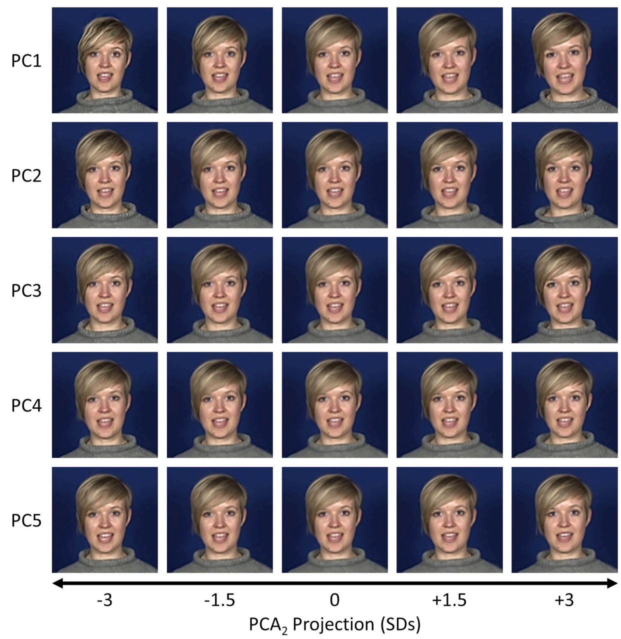


FIGURE 5 | Visualisations of the final frame in the sequence for the first five components from the PCA_2 space for an example dataset (S3, "I'm sorry to say"). Samples were selected at positions between ± 3 standard deviations along a given principal component axis and back-projected to the image space.

say" data. Animations for example clips from other datasets are shown in **Supplementary Videos 9–14**, including anti-faces, anti-caricatures, and caricatures. The caricaturing process can modulate the intensity of multiple idiosyncratic behaviours within each clip, such as the head orientation, head movements, blinks, and mouth movements. Because the original behaviours all occurred across relatively similar timescales, the caricaturing effects here are most salient for spatial features. Nevertheless, modulations of temporal features are also evident; for instance, the caricaturing also alters the duration of sequences that are shorter or longer than the average sequence. Weightings outside of the zero to one range (caricatures and anti-faces) are prone to introducing distortions into the image, especially in the case of the anti-faces. This is because the McGM features represent the facial motion in terms of changes in shape and texture over time; modulating these features outside the normal range will therefore exaggerate shape as well as temporal deviations, leading to shape distortions. Consequently, this technique may be best suited to producing anti-caricatures, using weightings within the zero to one range.

To quantify the relationship between the caricaturing process and the extent of biological motion, we conducted two further analyses. First, we obtained an objective measure of the degree of motion in each video sequence by calculating the magnitude of the vectors in the *x*- and *y*-warp components of the McGM feature space. These represent the magnitude of deviation between each frame and the original reference image (**Figure 1A**), such that larger values indicate a greater degree of movement over frames. Distributions of motion magnitudes over frames are illustrated in **Figure 7A** for varying anti-caricature and caricature levels. As the level of caricaturing increases the distributions become increasingly broad and biased toward larger values, indicating greater magnitudes of motion in the clips. Secondly, we obtained perceptual ratings from 10 naive observers for the dynamicity of each clip across anti-caricature levels of the "I'm sorry to say" phrases. Summaries of the ratings are illustrated in **Figure 7B**: as the level of caricaturing increased so too did the dynamicity ratings. This was confirmed with a mixed-effects ordinal logistic regression, which revealed that increasing the caricature level significantly increased the likelihood of providing a higher dynamicity rating [$\beta_1 = 3.62$, $\exp(\beta_1) = 37.20$, $z = 27.09$, $p < 0.001$]. Thus, both objective and perceptual measures indicated that the caricaturing process successfully modulated the degree of biological motion.

DISCUSSION

In this study, we present a novel method for capturing spatiotemporal patterns of biological motion in dynamic facial behaviours and demonstrate how this can be used to create dynamic anti-caricatures of those behaviours. This technique extends existing spatial caricaturing methods by allowing manipulation of both spatial and temporal features. A PCA-based active appearance model is first used to capture dimensions of spatial variation. Following temporal alignment of the PCA timeseries, the scores are then submitted to a second-order PCA that further encodes spatiotemporal variations amongst facial behaviours. Each of these PCA spaces embodied all gestured behaviours, including both rigid and non-rigid modes of facial variation. Weighting a given sample within the second-order PCA space yields dynamic (anti-)caricatures of that sequence relative to the average first-order PCA timeseries. Both objective and behavioural measurements confirmed this technique modulated the degree of biological motion in the facial behaviours.

Facial caricatures offer an important tool for studying face perception by allowing parametric manipulation of key facial dimensions that would be difficult or impossible for a person to pose naturally. Caricaturing manipulations predict behavioural ratings of corresponding facial features (Benson and Perrett, 1991a; Burt and Perrett, 1995; Calder et al., 1997, 2000; Blanz et al., 2000; Furl et al., 2022), produce perceptual adaptation effects (Leopold et al., 2001; Jiang et al., 2006; Juricevic and Webster, 2012), and predict neural responses in face-selective regions (Furl et al., 2020). Our approach extends existing spatial caricaturing techniques by providing a method for generating dynamic anti-caricatures from natural facial behaviours, thereby allowing investigation of both spatial and temporal features underlying face perception. The dimensions manipulated by the caricaturing will depend on the choice of reference, determined by the average of the clips in the first-order PCA space. In our demonstration, this was an average of all clips within a given phrase type ("Good news" or "I'm sorry to say"), and thus the caricaturing manipulated idiosyncratic behaviours and biological

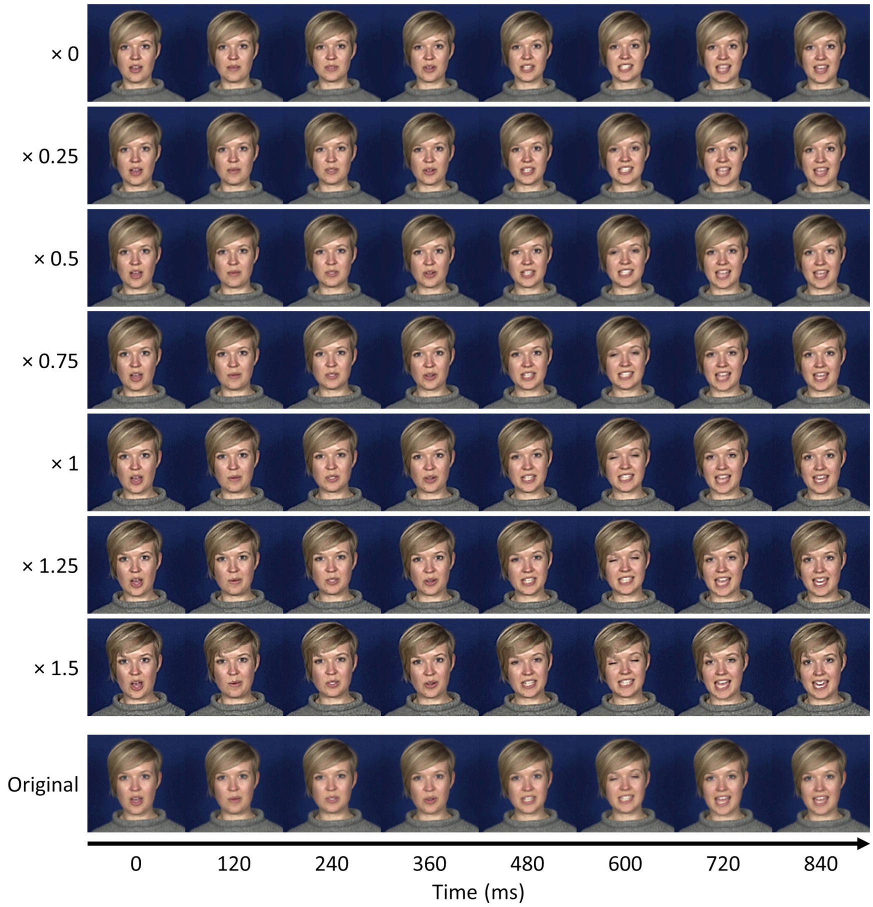


FIGURE 6 | Still frames from dynamic anti-caricatures and caricatures of an example clip (S3, "I'm sorry to say"). Weighting by zero yields the average video sequence, and weighting by one reconstructs the original clip. Intermediate weightings yield varying levels of anti-caricature, while weightings greater than one yield caricatures. The bottom row illustrates frames from the original clip matched to approximately the same timepoints.

motion patterns in each clip relative to this average. We derived dynamic facial behaviours from speech patterns: these provide a good target case as they represent a key source of non-rigid motion in the face, and also provide a degree of regularity that is helpful for forming a temporal average. Nevertheless, the technique could be applied to any pattern of facial movements provided that some temporal average (or other reference) can be formed for such movements. Such behaviours could, for instance, include poses of emotional expressions or head turns. Dynamic caricatures of facial behaviours could be applied to study recognition of those behaviours (Furl et al., 2020, 2022), similar to how static caricatures have been used to characterise recognition of facial identity (Leopold et al., 2001). Indeed, our own behavioural results confirmed that observers' perception of facial dynamicity increased with increasing levels of caricaturing. Furthermore, this technique could be extended to manipulating and characterising other facial characteristics in dynamic videos, such as emotion (Calder et al., 1997), identity (Benson and Perrett, 1991a), or age (Burt and Perrett, 1995).

A key component of this process is that the first-order PCA timeseries should be temporally aligned before conducting the second-order PCA. We used a dynamic time warping process based on the audio streams as this offered a precise measure of the vocalisation timings, but similar methods could be applied to other time-varying metrics such as the position of key facial landmarks (Hill et al., 2005; Furl et al., 2020). We included the time warping gradients along with the vectorised PCA_1 timeseries as inputs to the second-order PCA - this served a theoretical purpose by including information about the temporal scale of the original facial behaviours, as well as a practical purpose by ensuring that back-projections from the PCA_2 space would reconstruct the gradients needed to "unwarp" the corresponding reconstructed PCA_1 timeseries back to its original timescale. It is important to note that a poor temporal alignment amongst clips will likely cause modulations in the first-order PCA timeseries to cancel and average out over clips, resulting in a reference timeseries that corresponds to a largely static face. Dynamic caricatures generated relative to a static reference may include undesirable properties, such as a modulation of global motion rather than of idiosyncratic facial behaviours. For instance, Hill et al. (2005) note that caricaturing a dynamic negative expression relative to a static neutral reference would have the effect of

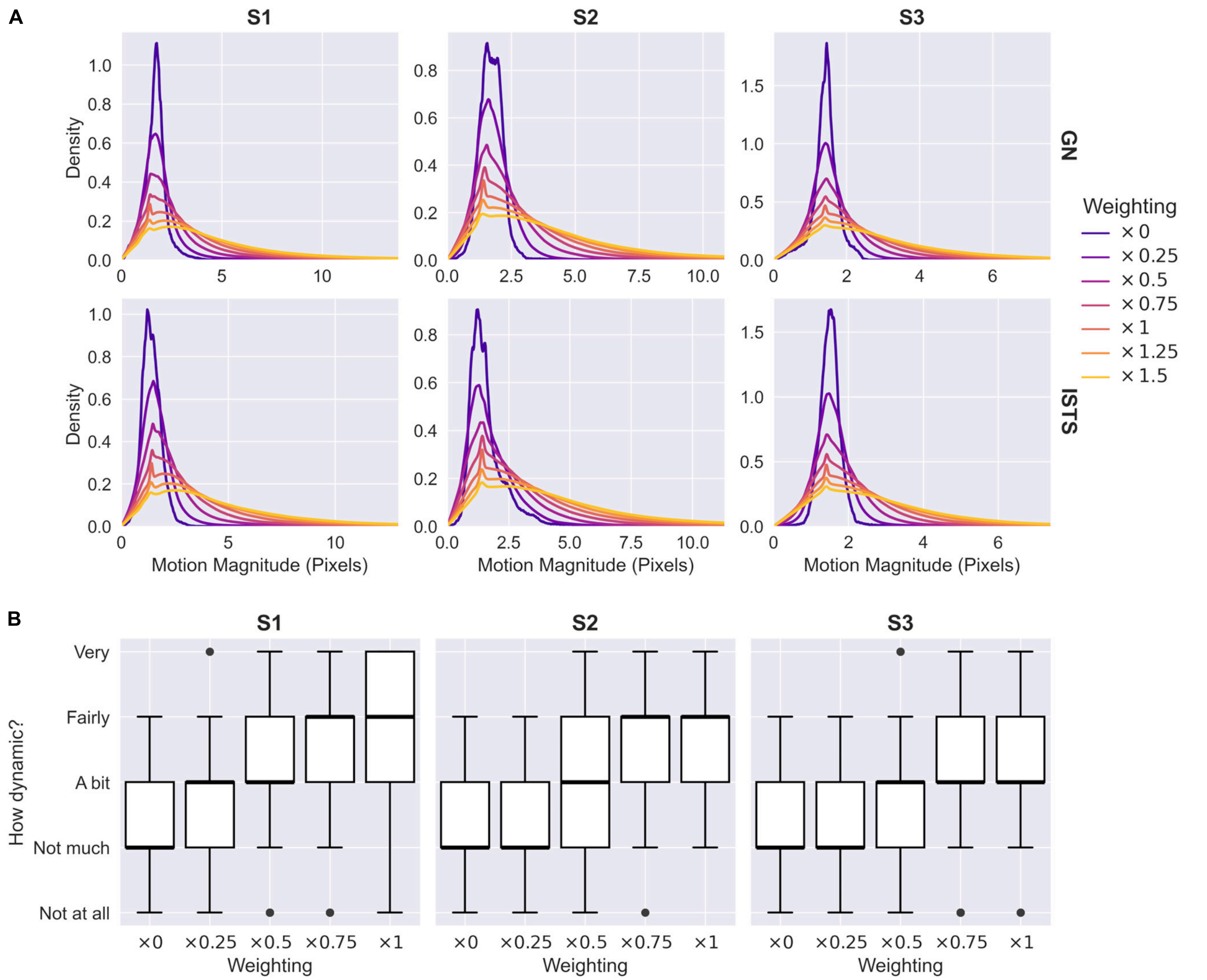


FIGURE 7 | Objective and perceptual measures of caricaturing. **(A)** Magnitudes of visual motion across varying levels of caricature and anti-caricature. Weighted samples are back-projected from the PCA_2 to the McGM space, and the magnitudes of the *x*- and *y*-warp components are extracted. Kernel density estimates illustrate distributions of magnitudes over all frames in all clips. GN, "Good news"; ISTS, "I'm sorry to say." **(B)** Perceptual ratings of dynamicity across varying levels of anti-caricature for the "I'm sorry to say" phrases. Boxplots illustrate medians, upper and lower quartiles, and data ranges excluding outliers.

increasing the overall dynamicity of the expression, yet we might expect an expression would actually become less dynamic with increasingly negative valence.

Here, we obtained recordings from a relatively small number of subjects, but with each subject performing many repeats of each phrase. This design allowed us to build PCA-based active appearance models that optimally characterised facial behaviours within each subject individually. Future applications of these techniques might additionally explore spatiotemporal variation between subjects or between different facial behaviours. For instance, PCA-based spatial caricaturing techniques have previously been used to morph between individuals varying in features such as gender (Griffin et al., 2011) and perceived political affiliation (Roberts et al., 2011). A more variable stimulus set comprising a wider variety of individuals or facial behaviours may prove more beneficial for such investigations.

Previous facial caricaturing methods have typically relied on manipulating the difference between an exemplar and reference face in terms of pre-defined facial landmarks. By contrast, the McGM employed here captures dynamic changes in shape and texture at the pixel-level (Johnston et al., 1992, 1999; Cowe, 2003). This allows our approach to advance on previous dynamic caricaturing methods by permitting manipulation of the original video textures instead of driving a virtual avatar. Consequently, our technique can represent finer and more nuanced changes in the faces, which would potentially be lost if only sampling sparse facial landmarks. Furthermore, our approach can capture changes in the texture and shape from shading that can be challenging to represent accurately in a virtual avatar. Nevertheless, our technique can be more prone to image distortions, especially outside the anti-caricature range (i.e., generating caricatures or anti-faces), which are less prevalent in virtual avatars. Manipulating the video textures may also modulate other incidental image properties; for instance, visual contrast was generally reduced for lower anti-caricaturing levels due to the averaging process. Visual contrast can influence the perception of various facial features including attractiveness (Pallak, 1983), age (Porcheron et al., 2017), emotional expression (Webb et al., 2020), and first impressions (Sato et al., 2008). The facial dynamicity ratings provided in our own behavioural experiment may potentially have been influenced by changes in visual contrast over varying levels of anti-caricature, although these changes would not be expected to influence facial motion directly, and subjective accounts of varying dynamicity are consistent with our objective measurements of visual motion magnitudes. Future applications of this technique may therefore consider whether further control or normalisation of such image features would be beneficial. Thus, the advantages and disadvantages of each approach may be best considered relative to the use case. Other methods based on temporal filtering have also been proposed for exaggerating motion in dynamic scenes (Wu et al., 2012), however, these produce a general increase of all motion within the scene, while our approach more specifically targets dynamic facial behaviours.

Weighting a sample in the second-order PCA space between zero and one allows generating varying levels of anti-caricature. Multiplication by values greater than one can create active caricatures, in which dynamic behaviours are exaggerated beyond the level present in the exemplar clip. Equally, multiplication by negative values can generate "anti-face" (anti-)caricatures, in which the encoded facial behaviours are inverted (Blanz et al., 2000; Leopold et al., 2001; Jiang et al., 2006; Juricevic and Webster, 2012; Furl et al., 2020, 2022). In our approach, however, multiplication outside the zero to one range (**Supplementary Videos 9–14**) tended to produce distortions in the image, particularly in the case of the anti-faces. The McGM features represent the motion of the face in terms of changes in shape and texture over frames. Modulation of these features outside the normal range will therefore exaggerate shape as well as temporal deviations, leading to shape distortions. At present, our technique may be best suited for generating dynamic anti-caricatures. Existing caricaturing techniques, such as those driving virtual avatars, may be more appropriate for generating more extreme caricatures or anti-faces depending on the use case.

Here we demonstrate the utility of our method for deriving dynamic anti-caricatures, however, it could be extended for many other purposes. First-order PCA face spaces have been used to classify and extract features underpinning emotional expressions (Calder et al., 2000; Watson et al., 2020) and facial identity (Kramer et al., 2017). They have also been used to generate predictive models of behavioural (Hancock et al., 1996, 1998) and neural (Chang and Tsao, 2017) representations of faces. The second-order PCA approach described here offers the opportunity to extend such investigations to include both spatial and temporal modes of facial variation. While the second-order PCA aims to identify variation along orthogonal linear components, other decomposition techniques may also be able to utilise alternative projections to extract other modes of spatiotemporal facial variation. For instance, independent components analysis would allow removing the orthogonality constraint, while manifold-learning techniques could derive a non-linear embedding. Such techniques could be used either alongside or instead of PCA.

CONCLUSION

We propose a novel PCA-based active appearance model for capturing dimensions of spatial and temporal variation in dynamic facial behaviours. A first-order PCA is used to encode modes of spatial variation in the faces. Representations within this space are then temporally aligned before being submitted to a second-order PCA. Dimensions of this space encode modes of spatiotemporal facial variation. We demonstrate how this technique can be used to produce dynamic anti-caricatures of biological motion patterns in faces, though the general method could be extended to numerous further avenues of research.

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: Open Science Framework (https://osf.io/t6crn/).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the School of Psychology at the University of Nottingham (Ethics approval numbers: 717, F1249). 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

DW performed the analysis under the supervision of AJ. Both authors conceived and developed the study, contributed to the writing of the manuscript, and approved the final version for submission.

FUNDING

This research was funded by the NIHR Nottingham Biomedical Research Centre and supported by the NIHR Nottingham Clinical Research Facility.

ACKNOWLEDGMENTS

We would like to thank Ben B. Brown for performing the original data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.880548/full#supplementary-material

Supplementary Video 1 | Visualisations of the first five components from the PCA_1 space for all datasets. Animations illustrate projections between ± 3 standard deviations along each given principal component axis. GN, "Good news"; ISTS, "I'm sorry to say."

Supplementary Video 2 | Back-projection of PCA_1 timeseries averaged over clips following temporal alignment. To aid visualisation, videos are played at half speed. GN, "Good news"; ISTS, "I'm sorry to say."

Supplementary Video 3 | Visualisations of the first five components from the PCA_2 space for S1 – "Good news" dataset. Samples were selected at positions between ± 3 standard deviations along a given principal component axis and back-projected to the image space. To aid visualisation, videos are played at half speed.

Supplementary Video 4 | As per Supplementary Video 3, for S1 – "I'm sorry to say" dataset.

Supplementary Video 5 | As per Supplementary Video 3, for S2 – "Good news" dataset.

Supplementary Video 6 | As per Supplementary Video 3, for S2 – "I'm sorry to say" dataset.

Supplementary Video 7 | As per Supplementary Video 3, for S3 – "Good news" dataset.

Supplementary Video 8 | As per Supplementary Video 3, for S3 – "I'm sorry to say" dataset.

Supplementary Video 9 | Dynamic caricatures for an example S1 - "Good news" clip. Multiplication by zero yields the average video sequence, and multiplication by one reproduces the original clip. Intermediate values yield varying levels of anti-caricature, values greater than one yield caricatures, and negative values yield anti-face (anti-)caricatures. To aid visualisation, videos are played at half speed.

Supplementary Video 10 | As per Supplementary Video 9, for an example S1 – "I'm sorry to say" clip.

Supplementary Video 11 | As per Supplementary Video 9, for an example S2 – "Good news" clip.

Supplementary Video 12 | As per Supplementary Video 9, for an example S2 – "I'm sorry to say" clip.

Supplementary Video 13 | As per Supplementary Video 9, for an example S3 – "Good news" clip.

Supplementary Video 14 | As per Supplementary Video 9, for an example S3 – "I'm sorry to say" clip.

REFERENCES

Benson, P. J., and Perrett, D. I. (1991b). Synthesising continuous-tone caricatures. *Image Vis. Comput.* 9, 123–129. doi: 10.1016/0262-8856(91)90022-H

Benson, P. J., and Perrett, D. I. (1991a). Perception and recognition of photographic quality facial caricatures: implications for the recognition of natural images. *Eur. J. Cogn. Psychol.* 3, 105–135. doi: 10.1080/09541449108406222

Berisha, F., Johnston, A., and McOwan, P. W. (2010). Identifying regions that carry the best information about global facial configurations. *J. Vis.* 10, 1–8. doi: 10.1167/10.11.27

Bernstein, M., and Yovel, G. (2015). Two neural pathways of face processing: a critical evaluation of current models. *Neurosci. Biobehav. Rev.* 55, 536–546. doi: 10.1016/j.neubiorev.2015.06.010

Blanz, V., O'Toole, A. J., Vetter, T., and Wild, H. A. (2000). On the other side of the mean: the perception of dissimilarity in human faces. *Perception* 29, 885–891. doi: 10.1068/p2851

Bruce, V., and Young, A. (1986). Understanding face recognition. *Br. J. Psychol.* 77, 305–327. doi: 10.1111/j.2044-8295.1986.tb02199.x

Burt, D. M. T., and Perrett, D. I. (1995). Perception of age in adult Caucasian male faces: computer graphic manipulation of shape and colour information. *Proc. Biol. Sci.* 259, 137–143. doi: 10.1098/rspb.1995.0021

Calder, A. J., and Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nat. Rev. Neurosci.* 6, 641–651. doi: 10.1038/nrn1724

Calder, A. J., Rowland, D., Young, A. W., Nimmo-Smith, I., Keane, J., and Perrett, D. I. (2000). Caricaturing facial expressions. *Cognition* 76, 105–146. doi: 10.1016/S0010-0277(00)00074-3

Calder, A. J., Young, A. W., Rowland, D., and Perrett, D. I. (1997). Computer-enhanced emotion in facial expressions. *Proc. Biol. Sci.* 264, 919–925. doi: 10.1098/rspb.1997.0127

Chang, L., and Tsao, D. Y. (2017). The Code for facial identity in the primate brain. *Cell* 169, 1013.e–1028.e. doi: 10.1016/j.cell.2017.05.011

Christensen, R. H. B. (2019). *Ordinal: Regression Models for Ordinal Data*. Available online at: https://github.com/runehaubo/ordinal. (accessed November 8, 2021).

Cowe, G. (2003). *Example-based Computer-generated Facial Mimicry.*Ph.D. Thesis. Ann Arbor, MI: ProQuest Dissertations Publishing.

Ellis, H. D. (1975). Recognizing faces. *Br. J. Psychol.* 66, 409–426. doi: 10.1111/j.2044-8295.1975.tb01477.x

Furl, N., Begum, F., Ferrarese, F. P., Jans, S., Woolley, C., and Sulik, J. (2022). Caricatured facial movements enhance perception of emotional facial expressions. *Perception* 51, 313–343. doi: 10.1177/03010066221086452

Furl, N., Begum, F., Sulik, J., Ferrarese, F. P., Jans, S., and Woolley, C. (2020). Face space representations of movement. *NeuroImage* 212:116676. doi: 10.1016/j.neuroimage.2020.116676

Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in R: the dtw package. *J. Stat. Softw.* 31, 1–24. doi: 10.18637/jss.v031.i07

Griffin, H. J., McOwan, P. W., and Johnston, A. (2011). Relative faces: encoding of family resemblance relative to gender means in face space. *J. Vis.* 11:8. doi: 10.1167/11.12.8

Hancock, P. J. B., Bruce, V., and Burton, M. A. (1998). A comparison of two computer-based face identification systems with human perceptions of faces. *Vision Res.* 38, 2277–2288. doi: 10.1016/S0042-6989(97)00439-2

Hancock, P. J. B., Burton, A. M., and Bruce, V. (1996). Face processing: human perception and principal components analysis. *Mem. Cognit.* 24, 26–40. doi: 10.3758/BF03197270

Hill, H. C. H., Troje, N. F., and Johnston, A. (2005). Range- and domain-specific exaggeration of facial speech. *J. Vis.* 5:4. doi: 10.1167/5.10.4

Jenkins, R., White, D., Van Montfort, X., and Mike Burton, A. (2011). Variability in photos of the same face. *Cognition* 121, 313–323. doi: 10.1016/j.cognition.2011.08.001

Jiang, F., Blanz, V., and O'Toole, A. J. (2006). Probing the visual representation of faces with adaptation. *Psychol. Sci.* 17, 493–500. doi: 10.1111/j.1467-9280.2006.01734.x

Johnston, A., McOwan, P. W., and Benton, C. P. (1999). Robust velocity computation from a biologically motivated model of motion perception. *Proc. Biol. Sci.* 266, 509–518. doi: 10.1098/rspb.1999.0666

Johnston, A., McOwan, P. W., and Buxton, H. (1992). A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proc. Biol. Sci.* 250, 297–306. doi: 10.1098/rspb.1992.0162

Juricevic, I., and Webster, M. A. (2012). Selectivity of face aftereffects for expressions and anti-expressions. *Front. Psychol.* 3:4. doi: 10.3389/fpsyg.2012.00004

Kramer, R. S. S., Young, A. W., Day, M. G., and Burton, A. M. (2017). Robust social categorization emerges from learning the identities of very few faces. *Psychol. Rev.* 124, 115–129. doi: 10.1037/rev0000048

Lander, K., and Butcher, N. (2015). Independence of face identity and expression processing: exploring the role of motion. *Front. Psychol.* 6, 1–6. doi: 10.3389/fpsyg.2015.00255

Leopold, D. A., O'Toole, A. J., Vetter, T., and Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat. Neurosci.* 4, 89–94. doi: 10.1038/82947

O'Toole, A. J., Roark, D. A., and Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends Cogn. Sci.* 6:255. doi: 10.1016/S1364-6613(02)01908-3

O'Toole, A., and Roark, D. (2010). "Memory for moving faces: the interplay of two recognition systems," in *Dynamic Faces*, (Cambridge, MA: The MIT Press). 15–30. doi: 10.7551/mitpress/9780262014533.003.0003

Oosterhof, N. N., and Todorov, A. (2008). The functional basis of face evaluation. *Proc. Natl. Acad. Sci. U.S.A* 105, 11087–11092. doi: 10.1073/pnas.0805664105

Pallak, S. R. (1983). Salience of a communicator's physical attractiveness and persuasion: a heuristic versus systematic processing interpretation. *Soc. Cogn.* 2, 158–170. doi: 10.1521/soco.1983.2.2.158

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., et al. (2019). PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* 51, 195–203. doi: 10.3758/s13428-018-01193-y

Porcheron, A., Mauger, E., Soppelsa, F., Liu, Y., Ge, L., Pascalis, O., et al. (2017). Facial contrast is a cross-cultural cue for perceiving age. *Front. Psychol.* 8:1208. doi: 10.3389/fpsyg.2017.01208

Roberts, T., Griffin, H., McOwan, P. W., and Johnston, A. (2011). Judging political affiliation from faces of UK MPs. *Perception* 40, 949–952. doi: 10.1068/p6985

Sato, S., Komori, H., Aoki, N., and Kobayashi, H. (2008). How image quality effects the impression created by facial photographs. *J. Soc. Photogr. Sci. Technol. Japan* 71, 92–98.

Scholes, C., Skipper, J. I., and Johnston, A. (2020). The interrelationship between the face and vocal tract configuration during audiovisual speech. *Proc. Natl. Acad. Sci. U.S.A* 117, 32791–32798. doi: 10.1073/pnas.2006192117

Tormene, P., Giorgino, T., Quaglini, S., and Stefanelli, M. (2009). Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation. *Artif. Intell. Med.* 45, 11–34. doi: 10.1016/j.artmed.2008.11.007

Turk, M., and Pentland, A. (1991). Eigenfaces for recognition. *J. Cogn. Neurosci.* 3, 71–86. doi: 10.1162/jocn.1991.3.1.71

Watson, D. M., Brown, B. B., and Johnston, A. (2020). A data-driven characterisation of natural facial expressions when giving good and bad news. *PLoS Comput. Biol.* 16:e1008335. doi: 10.1371/journal.pcbi.1008335

Webb, A. L. M., Hibbard, P. B., and O'Gorman, R. (2020). Contrast normalisation masks natural expression-related differences and artificially enhances the perceived salience of fear expressions. *PLoS One* 15:e0234513. doi: 10.1371/journal.pone.0234513

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., and Freeman, W. (2012). Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph.* 31, 1–8. doi: 10.1145/2185520.2185561

Author Disclaimer: The views represented are those of the authors alone and do not necessarily represent the views of the Department of Health in England, NHS, or the National Institute for Health Research.

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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Psychology

MINI REVIEW
published: 14 June 2022
doi: 10.3389/fpsyg.2022.896514

Neural Processing of Facial Attractiveness and Romantic Love: An Overview and Suggestions for Future Empirical Studies

Ryuhei Ueda[1,2]*

[1]*Institute for the Future of Human Society, Kyoto University, Kyoto, Japan,* [2]*Center for Information and Neural Networks, National Institute of Information and Communications Technology, Osaka, Japan*

OPEN ACCESS

Edited by:
Ayahito Ito,
Kochi University of Technology, Japan

Reviewed by:
Leehyun Yoon,
University of California, Davis, United States
Kazuki Yoshida,
Hokkaido University, Japan

Correspondence:
Ryuhei Ueda
ueda.ryuhei.3h@kyoto-u.ac.jp

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

Received: *15 March 2022*
Accepted: *25 May 2022*
Published: *14 June 2022*

Citation:
Ueda R (2022) Neural Processing of Facial Attractiveness and Romantic Love: An Overview and Suggestions for Future Empirical Studies. Front. Psychol. 13:896514. doi: 10.3389/fpsyg.2022.896514

Romantic love is universally observed in human communities, and the manner in which a person chooses a long-term romantic partner has been a central question in studies on close relationships. Numerous empirical psychological studies have demonstrated that facial attractiveness greatly impacts initial romantic attraction. This close link was further investigated by neuroimaging studies showing that both viewing attractive faces and having romantic thoughts recruit the reward system. However, it remains unclear how our brains integrate perceived facial attractiveness into initial romantic attraction. In addition, it remains unclear how our brains shape a persistent attraction to a particular person through interactions; this persistent attraction is hypothesized to contribute to a long-term relationship. After reviewing related studies, I introduce methodologies that could help address these questions.

Keywords: romantic love, facial attractiveness, social decision-making, social cognition, neuroimaging

INTRODUCTION

Our lives are filled with social relationships. Some close relationships, including romantic relationships, might last for years or even decades. In the initiation of a long-term relationship, deliberate thinking is likely required to infer the personal traits of a person that one encounters. Contrary to this intuition, experimental studies have demonstrated that personality inferences from facial appearance are performed very rapidly and strongly influence the perceiver's attitude toward the person (Todorov et al., 2008, 2013). This review focuses on the cognitive process in the context of the initiation of a romantic relationship; more specifically, how is the perceived facial attractiveness of a potential partner integrated into the initial romantic attraction? In this review, I employ the definition of the term "attraction" provided by Montoya and Horton (2014): "a person's immediate and positive affective and/or behavioral response to a specific individual, a response that is influenced by the person's cognitive assessments."

Close Link Between Facial Attractiveness and Romantic Attraction

Romantic love is observed in nearly all societies (Jankowiak and Fischer, 1992) and is thought to be deeply connected to human mate selection (e.g., Fisher, 2004; Walum and Young, 2018). For these reasons, extensive studies on close relationships have been devoted to understanding how people evaluate potential partners. Early experimental studies reported that facial attractiveness could predict initial romantic attraction on actual dates (e.g., Walster et al., 1966; Byrne et al., 1970). This close link has been focused on by evolutionary psychologists working on human mate selection (e.g., Buss, 1989; Buss and Schmitt, 1993). Buss (1989) conducted a cross-cultural survey and observed significantly higher desirability for partners' physical attractiveness by males than by females in almost all countries. Based on the findings, the authors hypothesized that several features shown in attractive faces (e.g., smooth skin, good muscle tone, lustrous hair, and full lips) could signal higher fertility and reproductive value or good health. A meta-analysis conducted by Feingold (1990) observed greater value placed on physical attractiveness of the female partners by males, although the sex difference was larger when examining self-reports than examining behavior. Based on these findings, Eastwick et al. (2014) predicted that sex differences in predictability of facial attractiveness of the partner on romantic evaluation could vary depending on the relationship stages. To test this prediction, the authors conducted a meta-analysis of studies that investigated romantic evaluation on an opposite-sex individual whom the participant has at least met face to face. The authors indicated that facial attractiveness of the partner strongly predicts romantic evaluation both for males and females to a similar degree, suggesting that the sex differences in ideal partner preference might emerge only in the stage of forming impressions of an ideal partner before a face-to-face interaction. This study and another study (Zsok et al., 2017) further suggested that this link might be stronger in the initiation stage than during the postformation stage of a long-term relationship.

We might think that we consider personal traits other than facial attractiveness when choosing an "ideal" partner. In line with this intuition, online dating services, which have rapidly expanded and surpassed meetings through friends in the United States this decade (Rosenfeld et al., 2019), usually require access to user profiles to facilitate searches for ideal partners. However, this intuitive view has been challenged by empirical evidence showing that stated mate preferences do not successfully predict initial romantic attraction in a live face-to-face interaction (e.g., Todd et al., 2007; Joel et al., 2017). Furthermore, a large dataset obtained from a commercial dating service indicated that while physically observable attributes, including attractiveness, strongly influenced romantic evaluation, harder-to-observe attributes had only small effects (Kurzban and Weeden, 2005). This almost exclusive influence of facial attractiveness on romantic evaluation likely reflects that people tend to pursue short-term partners. However, this view has been contradicted by a study showing that both males and females generally pursue not short-term but long-term partners in live face-to-face interactions, even at a younger age (Asendorpf et al., 2011). Based on these empirical findings, Eastwick et al. (2014) pointed out that the heavy reliance on descriptive profiles that are part of online dating services might reflect a misunderstanding of the information that people use to evaluate potential partners. In sum, facial attractiveness rapidly and strongly induces initial romantic attraction even when people pursue long-term partners, and people may tend to underestimate the impact of facial attractiveness.

Neurobiological Basis of Perceiving Facial Attractiveness

How does the human brain process attractive faces, which may inspire romantic interest? Since the 2000s, neuroimaging studies have aimed to identify the neurobiological basis for the perception of facial attractiveness. Early functional magnetic resonance imaging (fMRI) studies revealed that viewing attractive faces induces neural activity in reward-related brain regions, mainly the nucleus accumbens (NAcc) and medial orbitofrontal cortex (mOFC) (e.g., Aharon et al., 2001; O'Doherty et al., 2003; **Figure 1A**). Engagement of these regions is hypothesized to reflect the subjective reward value of faces. This prediction has been supported by subsequent studies showing that the sexual orientation of the perceiver modulates neural responses in the mOFC when viewing attractive male and female faces (Kranz and Ishai, 2006; Ishai, 2007). Other studies have provided evidence to support the model indicating that rapid and automatic engagement of the NAcc to presented faces conveys a preference signal to the mOFC underlying deliberative evaluation (Kim et al., 2007; Chatterjee et al., 2009).

Romantic Evaluation of Potential Partners Modulated by Relationship Status

Neuroimaging studies have revealed that engagement of the reward system, in which attractive faces are more valued, elicits romantic interest. After the formation of a long-term relationship, however, attractive alternative partners who may threaten the relationship may be devalued (Johnson and Rusbult, 1989; Simpson et al., 1990). Meyer et al. (2011) examined neural correlates of this derogation effect. The authors observed increased activation in the control-related prefrontal regions during successful derogation of attractive alternative partners in the absence of cognitive load, which was negatively correlated with activation in the ventral striatum. These results suggest that when cognitive resources were available, recruiting the prefrontal region could contribute to the deliberate regulation of attraction to alternative partners. In contrast, engagement of prefrontal regions during successful derogation was not observed under cognitive load, suggesting that derogation could implicitly occur without requiring cognitive resources. While early studies on relationship maintenance mainly focused on explicit and deliberate aspects of regulation (Ritter et al., 2010; Pronk et al., 2011), an increased focus has been placed on encompassing such implicit

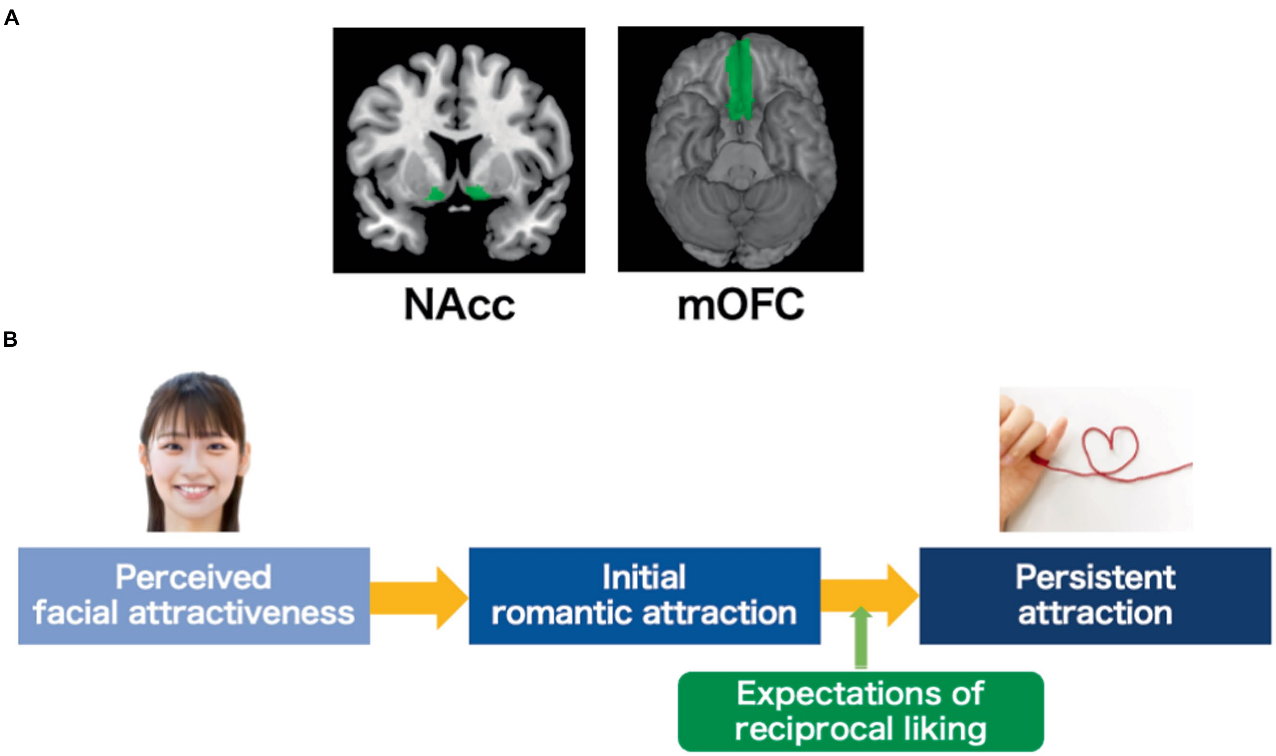


FIGURE 1 | (A) Main brain regions involved in the euphoria of viewing attractive faces: the nucleus accumbens (NAcc) and medial orbitofrontal cortex (mOFC). The images were created using the WFU_PickAtlas (Maldjian et al., 2003, 2004) and MRIcron (Rorden et al., 2007). **(B)** Schematic model of integration between perceived facial attractiveness and initial romantic attraction. A persistent attraction to a particular person must be shaped through interactions even before a long-term relationship is formed, which could be modulated by expectations of reciprocal liking. Copyright: ImageNavi, republished with permission.

and automatic aspects into the theoretical framework (Linardatos and Lydon, 2011; Finkel and Eastwick, 2015; Karremans et al., 2015; Lydon and Karremans, 2015; Pronk and Righetti, 2015; Ueda et al., 2017, 2018).

Subjective preference for attractive faces may also be modulated by the relationship status of the potential partner; that is, people tend to devalue attractive opposite-sex individuals who are in a relationship with a significant other. This kind of derogation seems to reflect the avoidance of "mate poaching" (Schmitt and Buss, 2001; Schmitt et al., 2004). Our previous work reported that individuals characterized by greater mOFC responses demonstrated a greater preference for hypothetical partners in a relationship with a significant other (Ueda et al., 2017), suggesting that this region might underpin the deliberate evaluation of potential partners (Kim et al., 2007; Chatterjee et al., 2009).

Behavioral and Neural Alterations After Building a Close Relationship

A long-term close relationship can be characterized by a persistent attraction to the partner both in human and nonhuman animals (Fisher et al., 2016; Walum and Young, 2018). For example, romantically involved individuals tend to exhibit a more positive evaluation of their partner's attractiveness (Murray et al., 1996; Murray and Holmes, 1997) and give less attention to alternative partners (Miller, 1997; Maner et al., 2008, 2009). These behavioral tendencies have been hypothesized to contribute to relationship maintenance (Linardatos and Lydon, 2011; Finkel and Eastwick, 2015; Karremans et al., 2015; Lydon and Karremans, 2015; Pronk and Righetti, 2015).

Those observations then raise the following question: how does the reward system shape a persistent attraction to a particular person? A neurobiological study on monogamous prairie voles provided evidence that revealed the relevant neural mechanisms (Aragona et al., 2006). The authors observed that the two dopamine (DA) receptor subtypes in the rostral shell of the NAcc have opposing roles in regulating pair bonding in male monogamous prairie voles. Specifically, the activation of DA D2-type receptors facilitated partner preference, whereas the activation of D1-type receptors not only did not induce partner preference but also prevented partner preference. The authors also observed increased D1-type receptors within the NAcc in pair-bonded males compared with nonpair-bonded males after 2 weeks of exposure to a female but not after only 24 h. This upregulation effect was not observed within the caudate putamen, and the other subtype, D2-type receptors, did not differ in either brain region. Given that 24 h of mating resulted in a partner preference, the authors argued that reorganization of the NAcc might not be necessary for the initial formation of the partner preference and that increased D1-type receptors in the NAcc might reflect a more fully established pair-bond, which could contribute to pair-bond maintenance.

Neural mechanisms underpinning pair-bond development in prairie voles are now considered to be precisely understood, and those findings are expected to be helpful for understanding the human mating system (Fisher et al., 2016; Walum and Young, 2018). This view has been partly supported by human neuroimaging studies, which have consistently reported engagement of the reward system, including the NAcc, when people have romantic thoughts about their committed partner (Bartels and Zeki, 2000, 2004; Aron et al., 2005; Xu et al., 2011; Acevedo et al., 2012; Takahashi et al., 2015). Our recent work using a decoding approach (see below) further demonstrated that the spatial activation patterns of the NAcc related to a romantic partner could be discriminated from those related to other opposite-sex individuals (Ueda and Abe, 2021). This finding is consistent with the hypothesis that the formation of a distinct neural representation of a long-term partner underpins a partner preference, which could contribute to relationship maintenance (Walum and Young, 2018).

While human neuroimaging studies have thus far been successful in identifying neural correlates underpinning the euphoria of viewing attractive faces and having romantic thoughts about a committed partner, it remains unclear how our brains integrate perceived facial attractiveness into initial romantic attraction. In addition, while prairie vole studies suggest that neural correlates underpinning a persistent attraction to a particular person must be shaped even before a long-term relationship is formed, we are still far from understanding the comparable neural mechanisms in humans (**Figure 1B**). In the following sections, I provide suggestions for future empirical studies to bridge this gap.

DISCUSSION: ACQUISITION OF EMPIRICAL EVIDENCE DEMONSTRATING THE NEURAL PROCESSING OF INITIAL ROMANTIC ATTRACTION

Combining Neuroimaging With an Ecologically Valid Task

To provide empirical evidence to address these issues, an ecologically valid experimental paradigm is needed first and foremost. Specifically, the assessment of romantic evaluation is more suitable in a real-life situation than in a hypothetical situation. The "speed-dating" paradigm fulfills this requirement to address questions relevant to initial attraction and relationship development in a live face-to-face interaction (Finkel et al., 2007; Eastwick and Finkel, 2008b). In the standard version of this paradigm, male and female participants who want to build a long-term relationship have a few minutes of interaction with each speed-dating partner, and at the end of each date, they are asked to evaluate their romantic interest (e.g., "Are you interested in meeting again this person?"). Male and female participants who showed mutual romantic interest in a speed-dating event are given an opportunity to meet again after completion of the study, although researchers are required to deal with several ethical issues before launching the study (e.g., potential risk of experiencing social rejection or social awkwardness, or very serious interpersonal troubles; Finkel et al., 2007).

Finkel et al. (2007) pointed out that while most studies had used ratings of hypothetical targets to assess romantic attraction (for reviews, Cooper and Sheldon, 2002; Montoya et al., 2018), some studies investigated this matter in a natural setting (e.g., Walster et al., 1966; Byrne et al., 1970; Bernstein et al., 1983; see also Feingold, 1990). Finkel et al. (2007) argued that in comparison with those real-life settings, the speed-dating paradigm has benefits other than ecological validity for evaluating initial romantic attraction. Specifically, unlike a traditional longer date with one partner, each participant simultaneously evaluates many partners and is evaluated by them in a few minutes during speed-dating events, which allows researchers to examine the degree to which features of a given interaction are derived from one partner, the other partner, or the unique dynamics between the two partners. As we see later, this feature also allows researchers to test the learning process during romantic evaluation (Cooper et al., 2014). Furthermore, Finkel et al. (2007) argued that a large array of experimental manipulations, such as comparing the effects of different date durations or disclosing personal information before starting a date on romantic evaluation, can be incorporated into the paradigm. Studies using this paradigm have mainly examined the predictability of factors related to ideal partner preference (e.g., facial attractiveness, body attractiveness, education, and income), which were originally identified through evaluation of hypothetical targets (e.g., Buss, 1989), on romantic evaluation in face-to-face interactions (e.g., Kurzban and Weeden, 2005; Fisman et al., 2006; Eastwick et al., 2007; Todd et al., 2007; Eastwick and Finkel, 2008a; Asendorpf et al., 2011; Joel et al., 2017).

Cooper et al. (2012) combined fMRI with this paradigm to test whether the neural activity that occurs while initially viewing potential partners' faces could predict romantic decisions in subsequent speed dating events (**Figure 2**). The authors found that neural activity in the paracingulate cortex was predictive of both subsequent romantic decisions and perceived facial attractiveness, suggesting that this region might be involved in the formation of an initial rapid romantic evaluation. The authors also observed that perceived facial attractiveness ratings were represented in the reward system, including in the striatum and ventromedial prefrontal cortex; however, more importantly, neural activation in the reward system was not predictive of subsequent decisions, implying distinctive neural processes during romantic evaluation.

Recent neuroimaging studies using the speed-dating paradigm have further provided intriguing findings, including synchronization of brain activities between the two people who are interacting (Yuan and Liu, 2022), the automatic formation of a "reflected impression" on the speed-dating partner (i.e., the degree to which participants expect the partner to positively view them; Ito et al., 2020), and similarities of functional

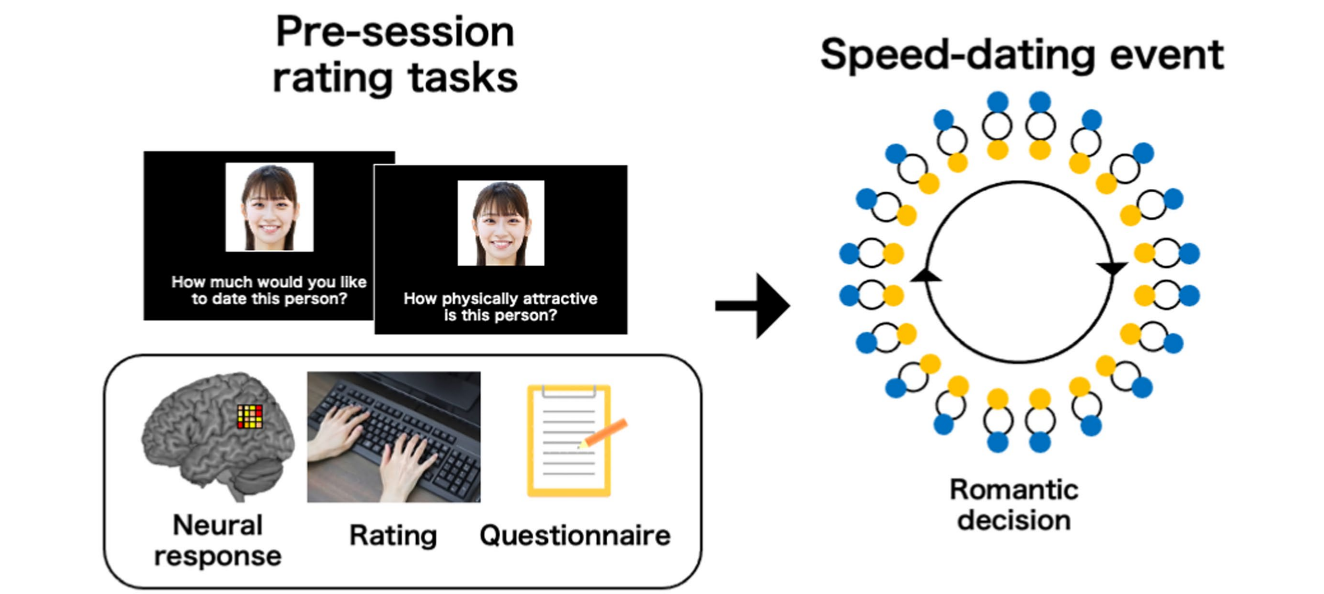


FIGURE 2 | Schematic illustration of the experimental procedure used in Cooper et al. (2012), which combined fMRI with a speed-dating paradigm. Prior to speed-dating events, researchers can collect neural data while initially viewing each speed-dating partner. Other rating data, including perceived facial attractiveness or questionnaire data, can also be recorded. In subsequent speed-dating events, participants met each other and had short conversations. At the end of each date, participants were asked to make romantic decisions about whether they wanted to see each partner again. Combining neural and behavior data that are recorded in a session with romantic decisions allowed us to assess the predictability of neural activation or self-reported ideal partner traits on the romantic decisions. Copyright: ImageNavi, republished with permission.

connectivity profiles obtained from resting-state fMRI data that are predictive of the compatibility of the impressions between the two partners (Kajimura et al., 2021). On the other hand, no previous studies have directly revealed how our brains integrate perceived facial attractiveness into initial romantic attraction. I introduce several analytical methodologies in the following sections that could be useful in addressing this issue.

Disentanglement of the Overlapping Function

Cooper et al. (2012) observed overlapping neural activity associated with facial attractiveness and initial romantic attraction in the paracingulate cortex. Their findings lead to the question of how this region processes each piece of information; that is, does this region process the initial romantic attraction of the potential partner in a different way than that which occurs when only viewing attractive faces? This question is difficult to address through conventional fMRI analysis, which aims to identify brain regions showing a greater response to the experimental conditions than to a control condition. A decoding approach using multivoxel pattern analysis (MVPA) could allow us to test this question by disentangling overlapping neural activity within the region. This method uses a machine-learning technique to decode mental states or cognitive processes from the spatial patterns of neural activity in a target region, thus allowing better interpretation of overlapping functional activations (Norman et al., 2006; Peelen and Downing, 2007). The simplest case of the decoding approach assumes that if one cognitive process induces specific neural activity patterns, the trained classifier can accurately discriminate activation patterns between the conditions. If the paracingulate cortex encodes the formation of initial romantic attraction for a particular person in a different way from perceived facial attractiveness, the neural activity patterns should be distinguishable from those related to individuals with attractive faces who are not selected as potential romantic partners in subsequent decisions. This investigation might help elucidate the early neural processing during romantic evaluation, i.e., how our brains shape preference for a partner at first sight.

One may think that engagement of the paracingulate cortex might also be observed in an evaluation process other than romantic evaluation, including choosing a friend based on the person's facial appearance or evaluation of the likeability on nonsocial objects. In other words, investigation of the specificity of neural processing would be needed. For this kind of investigation with higher-order representational space across multiple dimensions, another type of multivariate method, representational similarity analysis (RSA; Kriegeskorte et al., 2008; Kriegeskorte and Kievit, 2013), may be a suitable approach. RSA quantifies the (dis)similarity of the spatial patterns of neural activity among stimuli within the target region based on the Pearson correlation coefficient or other measures. The (dis)similarity among stimuli can be summarized as a representation (dis)similarity matrix, and then researchers can assess the relatedness between the neural (dis)similarity and the stimulus features. This standard procedure of RSA is well summarized in the review by Popal et al. (2019). In the case of extending findings by Cooper et al. (2012), by including additional conditions (e.g., viewing same-sex unfamiliar individuals or scenery pictures) and/or combining with data from control rating tasks (e.g., evaluation of perceived roundness of faces; Kim et al., 2007), researchers may test the functional specificity of neural processing of preference for a partner at first sight (i.e., examining whether significant association of neural activity patterns with ratings across stimuli in the paracingulate cortex is observed only for romantic evaluation). MVPA including RSA that focuses on neural activity patterns across multiple voxels has an advantage over individual-voxel-based conventional method in increased sensitivity to the cognitive states (Norman et al., 2006). In addition, as described above, disentanglement of overlapping neural activity within the region could contribute to revealing the representational content (Popal et al., 2019). Another strong point of RSA is that it allows a direct comparison of data obtained from different species in a common representation space (Kriegeskorte et al., 2008). Given the importance of bridging the gap with the findings obtained from monogamous prairie vole studies, this investigation would also be valuable.

Modulation of Initial Romantic Attraction Through Interaction

This review has thus far focused on initial romantic attraction. In the last section, I provide suggestions for investigating an advanced question: how could initial romantic attraction be modulated through the subsequent interaction? As described above, a close relationship can be characterized by a persistent attraction to a particular person, and studies have suggested specific neural mechanisms in the reward system (Aragona et al., 2006; Fisher et al., 2016; Walum and Young, 2018; Ueda and Abe, 2021).

Although little is known about the pertinent dynamic neural process, the computational modeling approach used in Cooper et al. (2014) may contribute to revealing this issue. In Cooper et al. (2014), following the completion of speed-dating events, some participants were evaluated by fMRI while they "learned" about their own desirability by receiving information indicating romantic interest from their speed-dating partners. To investigate neural responses associated with learned expectations about partner decisions, the authors implemented a reinforcement learning model. This model assumed that participants learned during the scan how likely partners were to express romantic interest in them and that the expectations would be updated based on the degree to which the expectations were violated at each trial (i.e., prediction error). This kind of computational approach has been widely used to describe the cognitive process of various social behaviors (Dunne and O'Doherty, 2013). Subtraction analysis of neuroimaging data demonstrated significantly greater responses to unexpected romantic interest expressed by a speed-dating partner in the posterior superior temporal sulcus. Furthermore, neural responses to a partner's decision were positively correlated with unsigned prediction errors based on the learning model in the medial prefrontal cortex. These two brain areas have been associated with the mentalizing process: encoding and updating beliefs about the intentions and feelings of others (Frith, 2007). While the model used in Cooper et al. (2014) focused on learning of the person's own desirability, this approach might also be applicable to describe the updating of the subjective preference to each partner through repeated interaction instead of through one-shot speed dating events.

Modulation of initial romantic attraction may involve other cognitive functions, such as emotion processing. People tend to like those who seem to like them (i.e., expectations of reciprocal liking; e.g., Aron et al., 1989; Riela et al., 2010), and fear of rejection could decrease their willingness to approach a potential partner (e.g., Bernstein et al., 1983). Cooper et al. (2014) reported greater neural responses in the anterior cingulate to rejection by a partner in whom a participant was romantically interested. This region is engaged when one experiences "social pain" caused by romantic rejection (Kross et al., 2011). Another study has shown that learning of an attractive potential partner's interest has a greater impact on romantic evaluation (Greitemeyer, 2010), implying that there is interplay between perceived facial attractiveness and emotion processing during interactions. Top-down cognitive control supported by the dorsolateral prefrontal cortex (Heatherton and Wagner, 2011) may also be involved in the modulation process. Specifically, successful devaluation of a potential partner who demonstrates less interest might elicit increased neural responses in the dorsolateral prefrontal cortex and decreased responses in the reward system. Connectivity analyses assessed by psycho-physiological interaction (PPI; Friston et al., 1997) or dynamic causal modeling (DCM; Friston et al., 2003) would allow us to directly test this hypothesis.

CONCLUSION

How we choose a long-term romantic partner has been a central question in studies on close relationships. In the past two decades, human neuroimaging studies have identified engagement of the reward system in the euphoria of viewing attractive faces, which may induce initial romantic attraction. Studies combining neuroimaging with speed dating have further identified specific neural responses that predict subsequent romantic decisions. These findings lead to further questions: How do our brains integrate signals related to perceived facial attractiveness into initial romantic attraction? How do our brains shape a persistent attraction to a particular person through interactions? Future studies will address these issues by elucidating the neural representations, dynamic alterations, and computational process underpinning the behavior.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approves it for publication.

FUNDING

This work was supported by JSPS KAKENHI grant number JP20K20157.

ACKNOWLEDGMENTS

The author thanks Nobuhito Abe for comments on this manuscript.

REFERENCES

Acevedo, B. P., Aron, A., Fisher, H. E., and Brown, L. L. (2012). Neural correlates of long-term intense romantic love. *Soc. Cogn. Affect. Neurosci.* 7, 145–159. doi: 10.1093/scan/nsq092

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., and Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron* 32, 537–551. doi: 10.1016/S0896-6273(01)00491-3

Aragona, B. J., Liu, Y., Yu, Y. J., Curtis, J. T., Detwiler, J. M., Insel, T. R., et al. (2006). Nucleus accumbens dopamine differentially mediates the formation and maintenance of monogamous pair bonds. *Nat. Neurosci.* 9, 133–139. doi: 10.1038/nn1613

Aron, A., Dutton, D. G., Aron, E. N., and Iverson, A. (1989). Experiences of falling in love. *J. Soc. Pers. Relat.* 6, 243–257. doi: 10.1177/0265407589063001

Aron, A., Fisher, H., Mashek, D. J., Strong, G., Li, H., and Brown, L. L. (2005). Reward, motivation, and emotion systems associated with early-stage intense romantic love. *J. Neurophysiol.* 94, 327–337. doi: 10.1152/jn.00838.2004

Asendorpf, J. B., Penke, L., and Back, M. D. (2011). From dating to mating and relating: predictors of initial and long–term outcomes of speed–dating in a community sample. *Eur. J. Personal.* 25, 16–30. doi: 10.1002/per.768

Bartels, A., and Zeki, S. (2000). The neural basis of romantic love. *Neuroreport* 11, 3829–3834. doi: 10.1097/00001756-200011270-00046

Bartels, A., and Zeki, S. (2004). The neural correlates of maternal and romantic love. *NeuroImage* 21, 1155–1166. doi: 10.1016/j.neuroimage.2003.11.003

Bernstein, W. M., Stephenson, B. O., Snyder, M. L., and Wicklund, R. A. (1983). Causal ambiguity and heterosexual affiliation. *J. Exp. Soc. Psychol.* 19, 78–92. doi: 10.1016/0022-1031(83)90006-9

Buss, D. M. (1989). Sex differences in human mate preferences: evolutionary hypotheses tested in 37 cultures. *Behav. Brain Sci.* 12, 1–14. doi: 10.1017/S0140525X00023992

Buss, D. M., and Schmitt, D. P. (1993). Sexual strategies theory: an evolutionary perspective on human mating. *Psychol. Rev.* 100, 204–232. doi: 10.1037/0033-295X.100.2.204

Byrne, D., Ervin, C. R., and Lamberth, J. (1970). Continuity between the experimental study of attraction and real-life computer dating. *J. Pers. Soc. Psychol.* 16, 157–165. doi: 10.1037/h0029836

Chatterjee, A., Thomas, A., Smith, S. E., and Aguirre, G. K. (2009). The neural response to facial attractiveness. *Neuropsychology* 23, 135–143. doi: 10.1037/a0014430

Cooper, J. C., Dunne, S., Furey, T., and O'Doherty, J. P. (2012). Dorsomedial prefrontal cortex mediates rapid evaluations predicting the outcome of romantic interactions. *J. Neurosci.* 32, 15647–15656. doi: 10.1523/JNEUROSCI.2558-12.2012

Cooper, J. C., Dunne, S., Furey, T., and O'Doherty, J. P. (2014). The role of the posterior temporal and medial prefrontal cortices in mediating learning from romantic interest and rejection. *Cereb. Cortex* 24, 2502–2511. doi: 10.1093/cercor/bht102

Cooper, M. L., and Sheldon, M. S. (2002). Seventy years of research on personality and close relationships: substantive and methodological trends over time. *J. Pers.* 70, 783–812. doi: 10.1111/1467-6494.05024

Dunne, S., and O'Doherty, J. P. (2013). Insights from the application of computational neuroimaging to social neuroscience. *Curr. Opin. Neurobiol.* 23, 387–392. doi: 10.1016/j.conb.2013.02.007

Eastwick, P. W., and Finkel, E. J. (2008a). Sex differences in mate preferences revisited: do people know what they initially desire in a romantic partner? *J. Pers. Soc. Psychol.* 94, 245–264. doi: 10.1037/0022-3514.94.2.245

Eastwick, P. W., and Finkel, E. J. (2008b). "Speed-dating: a powerful and flexible paradigm for studying romantic relationship initiation," in *Handbook of Relationship Initiation*. eds. S. Sprecher, A. Wenzel and J. Harvey (Hove: Psychology Press)

Eastwick, P. W., Finkel, E. J., Mochon, D., and Ariely, D. (2007). Selective versus unselective romantic desire. *Psychol. Sci.* 18, 317–319. doi: 10.1111/j.1467-9280.2007.01897.x

Eastwick, P. W., Luchies, L. B., Finkel, E. J., and Hunt, L. L. (2014). The predictive validity of ideal partner preferences: a review and meta-analysis. *Psychol. Bull.* 140, 623–665. doi: 10.1037/a0032432

Feingold, A. (1990). Gender differences in effects of physical attractiveness on romantic attraction: a comparison across five research paradigms. *J. Pers. Soc. Psychol.* 59, 981–993. doi: 10.1037/0022-3514.59.5.981

Finkel, E. J., and Eastwick, P. W. (2015). Attachment and pairbonding. *Curr. Opin. Behav. Sci.* 3, 7–11. doi: 10.1016/j.cobeha.2014.12.006

Finkel, E. J., Eastwick, P. W., and Matthews, J. (2007). Speed dating as an invaluable tool for studying romantic attraction: a methodological primer. *Pers. Relationship* 14, 149–166. doi: 10.1111/j.1475-6811.2006.00146.x

Fisher, H. E. (2004). *Why We Love: The Nature and Chemistry of Romantic Love*. New York, NY: Henry Holt.

Fisher, H. E., Xu, X., Aron, A., and Brown, L. L. (2016). Intense, passionate, romantic love: a natural addiction? How the fields that investigate romance and substance abuse can inform each other. *Front. Psychol.* 7:687. doi: 10.3389/fpsyg.2016.00687

Fisman, R., Iyengar, S. S., Kamenica, E., and Simonson, I. (2006). Gender differences in mate selection: evidence from a speed dating experiment. *Q. J. Econ.* 121, 673–697. doi: 10.1162/qjec.2006.121.2.673

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., and Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* 6, 218–229. doi: 10.1006/nimg.1997.0291

Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *NeuroImage* 19, 1273–1302. doi: 10.1016/s1053-8119(03)00202-7

Frith, C. D. (2007). The social brain? *Phil. Trans. R Soc. B* 362, 671–678. doi: 10.1098/rstb.2006.2003

Greitemeyer, T. (2010). Effects of reciprocity on attraction: the role of a partner's physical attractiveness. *Pers. Relationship* 17, 317–330. doi: 10.1111/j.1475-6811.2010.01278.x

Heatherton, T. F., and Wagner, D. D. (2011). Cognitive neuroscience of self-regulation failure. *Trends Cogn. Sci.* 15, 132–139. doi: 10.1016/j.tics.2010.12.005

Ishai, A. (2007). Sex, beauty and the orbitofrontal cortex. *Int. J. Psychophysiol.* 63, 181–185. doi: 10.1016/j.ijpsycho.2006.03.010

Ito, A., Yoshida, K., Takeda, K., Sawamura, D., Murakami, Y., Hasegawa, A., et al. (2020). The role of the ventromedial prefrontal cortex in automatic formation of impression and reflected impression. *Hum. Brain Mapp.* 41, 3045–3058. doi: 10.1002/hbm.24996

Jankowiak, W. R., and Fischer, E. F. (1992). A cross-cultural perspective on romantic love. *Ethnology* 31, 149–155. doi: 10.2307/3773618

Joel, S., Eastwick, P. W., and Finkel, E. J. (2017). Is romantic desire predictable? Machine learning applied to initial romantic attraction. *Psychol. Sci.* 28, 1478–1489. doi: 10.1177/0956797617714580

Johnson, D. J., and Rusbult, C. E. (1989). Resisting temptation: devaluation of alternative partners as a means of maintaining commitment in close relationships. *J. Pers. Soc. Psychol.* 57, 967–980. doi: 10.1037/0022-3514.57.6.967

Kajimura, S., Ito, A., and Izuma, K. (2021). Brain knows who is on the same wavelength: resting-state connectivity can predict compatibility of a female-male relationship. *Cereb. Cortex* 31, 5077–5089. doi: 10.1093/cercor/bhab143

Karremans, J. C., Pronk, T. M., and van der Wal, R. C. (2015). Executive control and relationship maintenance processes: an empirical overview and theoretical integration. *Soc. Personal. Psychol. Compass* 9, 333–347. doi: 10.1111/spc3.12177

Kim, H., Adolphs, R., O'Doherty, J. P., and Shimojo, S. (2007). Temporal isolation of neural processes underlying face preference decisions. *Proc. Natl. Acad. Sci. U. S. A.* 104, 18253–18258. doi: 10.1073/pnas.0703101104

Kranz, F., and Ishai, A. (2006). Face perception is modulated by sexual preference. *Curr. Biol.* 16, 63–68. doi: 10.1016/j.cub.2005.10.070

Kriegeskorte, N., and Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn. Sci.* 17, 401–412. doi: 10.1016/j.tics.2013.06.007

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2:4. doi: 10.3389/neuro.06.004.2008

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60, 1126–1141. doi: 10.1016/j.neuron.2008.10.043

Kross, E., Berman, M. G., Mischel, W., Smith, E. E., and Wager, T. D. (2011). Social rejection shares somatosensory representations with physical pain. *Proc. Natl. Acad. Sci. U. S. A.* 108, 6270–6275. doi: 10.1073/pnas.1102693108

Kurzban, R., and Weeden, J. (2005). HurryDate: mate preferences in action. *Evol. Hum. Behav.* 26, 227–244. doi: 10.1016/j.evolhumbehav.2004.08.012

Linardatos, L., and Lydon, J. E. (2011). Relationship-specific identification and spontaneous relationship maintenance processes. *J. Pers. Soc. Psychol.* 101, 737–753. doi: 10.1037/a0023647

Lydon, J., and Karremans, J. C. (2015). Relationship regulation in the face of eye candy: a motivated cognition framework for understanding responses to attractive alternatives. *Curr. Opin. Psychol.* 1, 76–80. doi: 10.1016/j.copsyc.2015.01.011

Maldjian, J. A., Laurienti, P. J., and Burdette, J. H. (2004). Precentral gyrus discrepancy in electronic versions of the Talairach atlas. *NeuroImage* 21, 450–455. doi: 10.1016/j.neuroimage.2003.09.032

Maldjian, J. A., Laurienti, P. J., Burdette, J. B., and Kraft, R. A. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage* 19, 1233–1239. doi: 10.1016/S1053-8119(03)00169-1

Maner, J. K., Gailliot, M. T., and Miller, S. L. (2009). The implicit cognition of relationship maintenance: inattention to attractive alternatives. *J. Exp. Soc. Psychol.* 45, 174–179. doi: 10.1016/j.jesp.2008.08.002

Maner, J. K., Rouby, D. A., and Gonzaga, G. C. (2008). Automatic inattention to attractive alternatives: the evolved psychology of relationship maintenance. *Evol. Hum. Behav.* 29, 343–349. doi: 10.1016/j.evolhumbehav.2008.04.003

Meyer, M. L., Berkman, E. T., Karremans, J. C., and Lieberman, M. D. (2011). Incidental regulation of attraction: the neural basis of the derogation of attractive alternatives in romantic relationships. *Cogn. Emot.* 25, 490–505. doi: 10.1080/02699931.2010.527494

Miller, R. S. (1997). Inattentive and contented: relationship commitment and attention to alternatives. *J. Pers. Soc. Psychol.* 73, 758–766. doi: 10.1037/0022-3514.73.4.758

Montoya, R. M., and Horton, R. S. (2014). A two-dimensional model for the study of interpersonal attraction. *Personal. Soc. Psychol. Rev.* 18, 59–86. doi: 10.1177/1088868313501887

Montoya, R. M., Kershaw, C., and Prosser, J. L. (2018). A meta-analytic investigation of the relation between interpersonal attraction and enacted behavior. *Psychol. Bull.* 144, 673–709. doi: 10.1037/bul0000148

Murray, S. L., and Holmes, J. G. (1997). A leap of faith? Positive illusions in romantic relationships. *Pers. Soc. Psychol. B* 23, 586–604. doi: 10.1177/0146167297236003

Murray, S. L., Holmes, J. G., and Griffin, D. W. (1996). The benefits of positive illusions: idealization and the construction of satisfaction in close relationships. *J. Pers. Soc. Psychol.* 70, 79–98. doi: 10.1037/0022-3514.70.1.79

Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430. doi: 10.1016/j.tics.2006.07.005

O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D. M., and Dolan, R. J. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* 41, 147–155. doi: 10.1016/s0028-3932(02)00145-8

Peelen, M. V., and Downing, P. E. (2007). Using multi-voxel pattern analysis of fMRI data to interpret overlapping functional activations. *Trends Cogn. Sci.* 11, 4–5. doi: 10.1016/j.tics.2006.10.009

Popal, H., Wang, Y., and Olson, I. R. (2019). A guide to representational similarity analysis for social neuroscience. *Soc. Cogn. Affect. Neurosci.* 14, 1243–1253. doi: 10.1093/scan/nsz099

Pronk, T. M., Karremans, J. C., and Wigboldus, D. H. J. (2011). How can you resist? Executive control helps romantically involved individuals to stay faithful. *J. Pers. Soc. Psychol.* 100, 827–837. doi: 10.1037/a0021993

Pronk, T. M., and Righetti, F. (2015). How executive control promotes happy relationships and a well-balanced life. *Curr. Opin. Psychol.* 1, 14–17. doi: 10.1016/j.copsyc.2014.11.016

Riela, S., Rodriguez, G., Aron, A., Xu, X., and Acevedo, B. P. (2010). Experiences of falling in love: investigating culture, ethnicity, gender, and speed. *J. Soc. Pers. Relat.* 27, 473–493. doi: 10.1177/0265407510363508

Ritter, S. M., Karremans, J. C., and van Schie, H. T. (2010). The role of self-regulation in derogating attractive alternatives. *J. Exp. Soc. Psychol.* 46, 631–637. doi: 10.1016/j.jesp.2010.02.010

Rorden, C., Karnath, H. O., and Bonilha, L. (2007). Improving lesion-symptom mapping. *J. Cogn. Neurosci.* 19, 1081–1088. doi: 10.1162/jocn.2007.19.7.1081

Rosenfeld, M. J., Thomas, R. J., and Hausen, S. (2019). Disintermediating your friends: how online dating in the United States displaces other ways of meeting. *Proc. Natl. Acad. Sci. U. S. A.* 116, 17753–17758. doi: 10.1073/pnas.1908630116

Schmitt, D. P., Alcalay, L., Allik, J., Angleitner, A., Ault, L., Austers, I., et al. (2004). Patterns and universals of mate poaching across 53 nations: the effects of sex, culture, and personality on romantically attracting another person's partner. *J. Pers. Soc. Psychol.* 86, 560–584. doi: 10.1037/0022-3514.86.4.560

Schmitt, D. P., and Buss, D. M. (2001). Human mate poaching: tactics and temptations for infiltrating existing mateships. *J. Pers. Soc. Psychol.* 80, 894–917. doi: 10.1037/0022-3514.80.6.894

Simpson, J. A., Gangestad, S. W., and Lerma, M. (1990). Perception of physical attractiveness: mechanisms involved in the maintenance of romantic relationships. *J. Pers. Soc. Psychol.* 59, 1192–1201. doi: 10.1037/0022-3514.59.6.1192

Takahashi, K., Mizuno, K., Sasaki, A. T., Wada, Y., Tanaka, M., Ishii, A., et al. (2015). Imaging the passionate stage of romantic love by dopamine dynamics. *Front. Hum. Neurosci.* 9:191. doi: 10.3389/fnhum.2015.00191

Todd, P. M., Penke, L., Fasolo, B., and Lenton, A. P. (2007). Different cognitive processes underlie human mate choices and mate preferences. *Proc. Natl. Acad. Sci. U. S. A.* 104, 15011–15016. doi: 10.1073/pnas.0705290104

Todorov, A., Mende-Siedlecki, P., and Dotsch, R. (2013). Social judgments from faces. *Curr. Opin. Neurobiol.* 23, 373–380. doi: 10.1016/j.conb.2012.12.010

Todorov, A., Said, C. P., Engell, A. D., and Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends Cogn. Sci.* 12, 455–460. doi: 10.1016/j.tics.2008.10.001

Ueda, R., and Abe, N. (2021). Neural representations of the committed romantic partner in the nucleus accumbens. *Psychol. Sci.* 32, 1884–1895. doi: 10.1177/09567976211021854

Ueda, R., Ashida, H., Yanagisawa, K., and Abe, N. (2017). The neural basis of individual differences in mate poaching. *Soc. Neurosci.* 12, 391–399. doi: 10.1080/17470919.2016.1182065

Ueda, R., Yanagisawa, K., Ashida, H., and Abe, N. (2017). Implicit attitudes and executive control interact to regulate interest in extra-pair relationships. *Cogn. Affect. Behav. Neurosci.* 17, 1210–1220. doi: 10.3758/s13415-017-0543-7

Ueda, R., Yanagisawa, K., Ashida, H., and Abe, N. (2018). Executive control and faithfulness: only long-term romantic relationships require prefrontal control. *Exp. Brain Res.* 236, 821–828. doi: 10.1007/s00221-018-5181-y

Walster, E., Aronson, V., Abrahams, D., and Rottman, L. (1966). Importance of physical attractiveness in dating behavior. *J. Pers. Soc. Psychol.* 4, 508–516. doi: 10.1037/h0021188

Walum, H., and Young, L. J. (2018). The neural mechanisms and circuitry of the pair bond. *Nat. Rev. Neurosci.* 19, 643–654. doi: 10.1038/s41583-018-0072-6

Xu, X., Aron, A., Brown, L., Cao, G., Feng, T., and Weng, X. (2011). Reward and motivation systems: a brain mapping study of early-stage intense romantic love in Chinese participants. *Hum. Brain Mapp.* 32, 249–257. doi: 10.1002/hbm.21017

Yuan, G., and Liu, G. (2022). Mate preference and brain oscillations: initial romantic attraction is associated with decreases in alpha- and lower beta-band power. *Hum. Brain Mapp.* 43, 721–732. doi: 10.1002/hbm.25681

Zsok, F., Haucke, M., De Wit, C. Y., and Barelds, D. P. H. (2017). What kind of love is love at first sight? An empirical investigation. *Pers. Relationship* 24, 869–885. doi: 10.1111/pere.12218

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.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 28 September 2022
DOI 10.3389/fpsyg.2022.997498

Check for updates

OPEN ACCESS

EDITED BY
Yoshiyuki Ueda,
Kyoto University, Japan

REVIEWED BY
Yuki Murai,
National Institute of Information and Communications Technology, Japan
Atsunobu Suzuki,
The University of Tokyo, Japan

*CORRESPONDENCE
Daniel N. Albohn
daniel.albohn@chicagobooth.edu

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

RECEIVED 19 July 2022
ACCEPTED 26 August 2022
PUBLISHED 28 September 2022

CITATION
Albohn DN, Uddenberg S and Todorov A (2022) A data-driven, hyper-realistic method for visualizing individual mental representations of faces. *Front. Psychol.* 13:997498.
doi: 10.3389/fpsyg.2022.997498

A data-driven, hyper-realistic method for visualizing individual mental representations of faces

Daniel N. Albohn*, Stefan Uddenberg and Alexander Todorov

Booth School of Business, The University of Chicago, Chicago, IL, United States

Research in person and face perception has broadly focused on group-level consensus that individuals hold when making judgments of others (e.g., "X type of face looks trustworthy"). However, a growing body of research demonstrates that individual variation is larger than shared, stimulus-level variation for many social trait judgments. Despite this insight, little research to date has focused on building and explaining individual models of face perception. Studies and methodologies that have examined individual models are limited in what visualizations they can reliably produce to either noisy and blurry or computer avatar representations. Methods that produce low-fidelity visual representations inhibit generalizability by being clearly computer manipulated and produced. In the present work, we introduce a novel paradigm to visualize individual models of face judgments by leveraging state-of-the-art computer vision methods. Our proposed method can produce a set of photorealistic face images that correspond to an individual's mental representation of a specific attribute across a variety of attribute intensities. We provide a proof-of-concept study which examines perceived trustworthiness/untrustworthiness and masculinity/femininity. We close with a discussion of future work to substantiate our proposed method.

KEYWORDS
StyleGAN, reverse correlation, face judgments, data-driven, face perception

Introduction

What types of faces do individuals draw to mind when they think of a "leader"? A "criminal"? A "thinker"? Understanding the content and downstream consequences of such visual stereotypes has been an area of active study within the field of social perception. However, the dominant approach to characterizing judgments derived from faces and facial attributes involves drawing conclusions based on aggregated measures collected from disparate observers. For example, one highly replicable finding is that neutral faces that have more feminine attributes (e.g., lighter skin, rounded jawline) and appear happier are judged as more trustworthy (see, e.g., Oosterhof and Todorov, 2008; Jaeger and Jones, 2021). However, this observation is based on averaged judgments across both faces and raters (i.e., participants). That is, *on average*, any given face with feminine and happy-like features will likely be judged by a given individual as trustworthy. However, not every feminine/smiling face is necessarily seen as trustworthy. Similarly, not every individual agrees on what a "trustworthy" face might look like. What one person may view as "trustworthy" another might view as "gullible" or perhaps "intelligent." These kinds of idiosyncrasies in face judgment are as numerous as the number of personality traits an individual can judge.

In the present work, we present a rationale for parsing, analyzing, and interpreting both participant-level (i.e., idiosyncratic) and stimulus-level (i.e., shared) contributions to social judgments of faces. First, we review literature that suggests idiosyncratic variance is both pervasive and meaningful in social trait judgments. However, we show that its contribution to social judgments depends on the type of judgment. Specifically, we use two types of social judgments: first-order facial judgments of masculinity and femininity and second-order, more complex judgments of trustworthiness (which rely on integrating many different lower-level perceived attributes in faces). While strict delineation between first- and second-order judgments is difficult to quantify, in the present work we define first-order judgments as those that have clear phenotypic qualities, such as sexually dimorphic features (e.g., jaw shape, face roundness), eye size, and overt or incidental resemblance to emotion expressions. These judgments are likely to have greater inter-rater agreement due to the physically observable features. Similarly, we define second-order judgments as more abstract judgments (e.g., trustworthiness) that are influenced by first-order judgments. Second-order judgments typically have lower inter-rater agreement. For example, the perceived second-order judgment of "babyfacedness" has been shown to be influenced by first-order judgments of face roundness, eye size, and incidental facial resemblance to fear expressions (Marsh et al., 2005; Zebrowitz, 2017). Here, we show that judgments of masculinity and femininity more closely align with our definition of first-order judgments, while judgments of trustworthiness align with what we define as second-order judgments.

Second, we review advances in machine learning and computer vision that can aid in capturing both idiosyncratic and shared variance in face judgments. Third, we introduce a novel, data-driven methodology to visualize idiosyncratic models of faces utilizing machine learning. Finally, we end with a proof-of-concept demonstration of our proposed method. We model idiosyncratic representations of perceived masculinity/femininity and trustworthiness of faces. Consistent with the variance component analyses, we find that these representations are much more similar across individuals in the case of masculinity/femininity than in the case of trustworthiness. We conclude with a set of future directions needed to further validate the proposed method.

Idiosyncratic and shared contributions to face judgments

Intuitively, two or more individuals are likely to disagree to some degree on their opinions about the attractiveness, trustworthiness, or masculinity and femininity of an individual. Indeed, there is a growing literature on individual preferences across diverse domains such as abstract art (Leder et al., 2016; Specker et al., 2020), architecture (Vessel et al., 2018), dancing (Isik and Vessel, 2019), facial beauty (Hönekopp, 2006; Martinez et al., 2020), voices (Lavan et al., 2020), and even technical writing and peer reviews (Jirschitzka et al., 2017). Yet, traditional analyses in person perception aggregate judgments to focus solely on the "shared" contributions of preferences that are similar across all participants. Thus, the majority of past research in face preferences stands in contrast to emerging evidence on the importance of idiosyncratic contributions to preference, taste, and judgments.

Estimating shared and idiosyncratic contributions to social judgments requires calculating variance components at three levels of interest: the stimulus, the participant, and their interaction (Martinez et al., 2020). The stimulus component represents the "shared" variance that is similar across all raters in the sample. For example, if every judge rated all smiling faces as more attractive, this would be reflected in the stimulus component. In other words, the "shared" stimulus attribute of smiling accounts for a certain proportion of the observed variance. On the other hand, the other two components represent participant-level, idiosyncratic contributions to judgments. The participant main effect contributions are idiosyncratic, but are often considered more ambiguous to interpret. For example, Rater A may judge the perceived happiness of two faces as *numerically* different from Rater B, but both raters could still rank order them similarly, resulting in mean differences across participants but identical face rankings. Such a case would be reflected in the participant main effect variance component, but it is unclear whether such differences reflect true idiosyncrasies or simply the fact that different participants interpret the response scale differently. On the other hand, the participant by stimulus interaction component is more straightforward to interpret. This component captures individual ranking preferences for the stimuli. For example, variance at this level will occur if Rater A prefers (or gives a higher perceived attribute rating to) one stimulus over another while Rater B prefers the opposite pairing (for detailed discussion see Hönekopp, 2006; Martinez et al., 2020).

In an early investigation on social judgment idiosyncrasies, Hönekopp (2006) revealed that perceptions of facial attractiveness were explained by *both* individual and shared preferences in taste. Hönekopp first had participants rate the same faces twice on perceived attractiveness one week apart. Next, shared and private (i.e., idiosyncratic) taste in attractiveness was evaluated by determining the proportion of variance explained by the stimulus face image, the rater, and the rater by stimulus interaction. Shared, stimulus-level contributions in taste accounted for approximately 33% of the observed variance while private, participant-level contributions in taste accounted for about 26% to 45% of the observed variance depending on whether the participant main effect variance was taken into account. Hönekopp (2006) determined that both individual- and stimulus-level contributions were important determinants of attractiveness judgments. Other investigations have shown that idiosyncratic variance contributions to attractiveness judgments range from 20 to 40% conservatively (taking into account participant by stimulus interactions only) to well over 50% if the more ambiguous participant main effect variance component is included (Hehman et al., 2017; Martinez et al., 2020).

Despite increasing evidence that there are large idiosyncratic contributions to judgments across a variety of domains, the degree to which there is more idiosyncratic variance over shared variance is likely graded within these specific domains. For example, within the domain of facial judgments, low level, first-order judgments that underlie higher level, second-order judgments are likely to have higher agreement. First-order judgments such as those for masculinity, femininity, skin tone, hair color, face shape, among others are likely to have more shared agreement since these attributes tend to be less perceptually ambiguous. In contrast, there is likely to be less agreement (i.e., more idiosyncratic contributions) for second-order judgments, such as those for attractiveness, trustworthiness, and dominance due to highly individualized preferences for these perceived attributes. In our own work, we have found evidence for such effects (Albohn, Martinez and Todorov, *in prep*). We had participants ($N = 99$) judge the femininity, masculinity, or trustworthiness of 120 neutral faces from the Chicago Face Database (Ma et al., 2015). Next, we computed the shared and idiosyncratic variance components following the procedures outlined by Martinez et al. (2020). When we examined the relative proportion of each type of variance, we observed different patterns dependent on whether the judgment was first-order (feminine or masculine) or second-order (trustworthy) as depicted in Figure 1. Specifically, we found that shared variance in judgments of facial masculinity and femininity accounted for approximately 60% of the observed reliable variance (depicted *via* the large blue bars in the left and center panels) but <4% of the reliable shared variance in judgments of trustworthiness (depicted *via* the small blue bar in the right panel). Importantly, this pattern flips when idiosyncratic contributions are examined. Idiosyncratic variance accounted for <20% of the variance for feminine and masculine face judgments (depicted *via* the black and yellow bars in the left and center panels) and around 20–65% of the idiosyncratic variance for trustworthy judgments (depicted *via* the black and yellow bars in the right panel) depending on whether participant main effect variance components are taken into account. This work aligns with previous research that has similarly found that trustworthiness judgments are often more idiosyncratic compared to judgments for other perceived attributes such as gender and race (Hehman et al., 2017).

Visualization of individual mental representations of faces

There exist myriad data-driven methods for visualizing mental representation of faces (e.g., Gosselin and Schyns, 2001; Mangini and Biederman, 2004; Oosterhof and Todorov, 2008; Schyns et al., 2009; Todorov et al., 2011; Zhan et al., 2021). However, to our knowledge psychophysical reverse correlation that produces noisy or computer generated avatar images is the only approach used for visualizing *individual* representations of faces (Sekuler et al., 2004; Dotsch and Todorov, 2012; Zhan et al., 2021). Typical reverse correlation in social perception overlays base images with random noise to randomly vary features of those base images (Dotsch et al., 2008). Participants then complete a forced-choice task whereby they choose which of two images (overlaid with randomized noise or its inverse) best represents a perceived target attribute (e.g., "Which face looks more "threatening?"). Afterwards, a classification image can be created for each participant by averaging all of the selected images together. In short, reverse correlation can visualize social perceptions of individuals by allowing a meaningful construct to "emerge from the noise."

Reverse correlation has become extremely popular in social perception since its inception. Indeed, past work has utilized reverse correlation to exemplify group averages of ethnic groups (Dotsch et al., 2008), social groups (Tskhay and Rule, 2015; Lloyd et al., 2020), facial emotions (Albohn et al., 2019; Albohn and Adams, 2020), and social categories such as perceived trustworthiness and dominance (Dotsch and Todorov, 2012). While many studies utilize reverse correlation for visualizing higher-level perceived attributes in faces, existing techniques can be improved considerably by leveraging advances from computer vision that have occurred in the last decade. Potential areas of improvement include (1) creating individual-level (rather than group-level) models of social perception; and (2) increasing classification image quality. We expand on both of these points in greater detail below.

Individualized models of social perception

Most studies utilizing reverse correlation only collect and report results from classification images averaged together at the group level. Then, these averaged images are rated by other samples of participants to confirm that such group-level classification images appear as intended (e.g., a "dominant" reverse correlation classification image actually appears "dominant" to other raters). However, the focus on

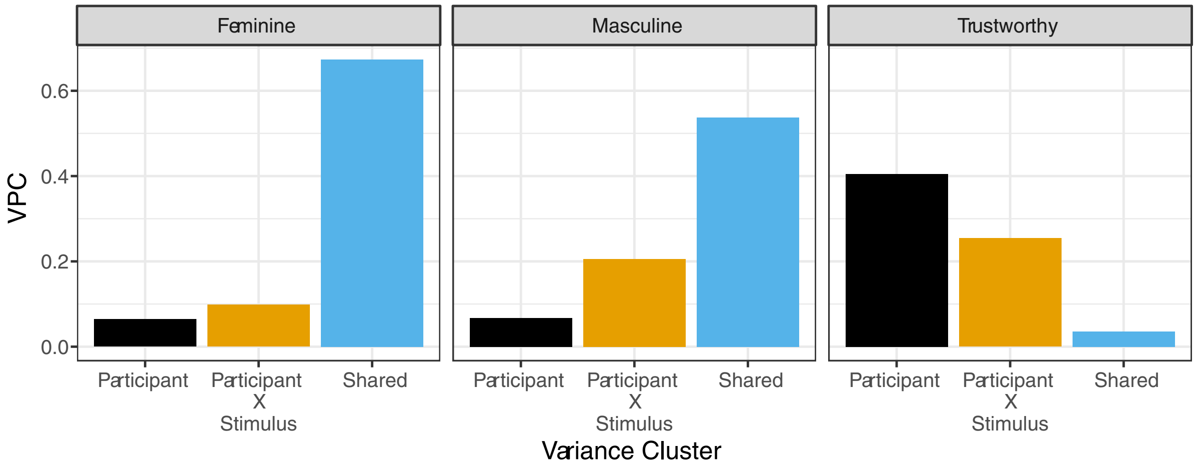


FIGURE 1
The variance partitioning coefficients (VPC) for feminine, masculine, and trustworthy judgment ratings across important variance clusters. The x-axis represents idiosyncratic (participant and participant by stimulus) and shared variance clusters. The y-axis represents the proportion of observed variance explained by each cluster.

characterizing only group-level images has been shown to inflate Type I error rates (Cone et al., 2020). To address this issue, individual classification images can be rated, though studies utilizing this approach typically still only report group means aggregated over all participants in the study and not individual-level effects. While such approaches have revealed important insights into social perception, they reveal little about the idiosyncrasies of social judgments. For example, it is possible that an individual's classification image of a "dominant" face could accurately visualize their mental representation of perceived facial "dominance," yet simultaneously fail to correspond to a group-level consensus representation of perceived "dominance." In such cases, the individual's visualized mental representation would likely either be discarded as an outlier or ignored.

Improved image quality

As noted earlier, the reverse correlation paradigm is, by its very nature, noisy. Such experiments require the application of carefully calibrated image-based noise to some base image in order to generate the distinct stimuli participants must choose between on each trial. Consequently, the resultant classification images — created by averaging across participant-selected noise patterns — are low fidelity. The images look noisy, pixelated, and blurry, or altogether computer-generated. Early visual reverse correlation techniques utilized stimuli that consisted solely of randomized noise (e.g., Gosselin and Schyns, 2003). Such noise-only methods typically required each participant to complete thousands of trials in order for meaningful data to emerge. In light of this difficulty, later iterations of visual reverse correlation made use of a blurred base image, on top of which the noise was applied (Mangini and Biederman, 2004; Dotsch et al., 2008). This helps guide visual classification and reduces the number of required trials by an order of magnitude — from thousands to several hundred. However, the resultant classification images can only be as clear as the images used in their creation. Even when randomized noise is superimposed over a high-fidelity base image, the output classification image is still necessarily noisy and blurry in appearance. Classification images are *prima facie* computer-manipulated and participants are unlikely to believe the images represent "real" faces, although human judges can still consensually discern gross anatomical features (e.g., mouth, eyes, hair, brows, etc), some social categories (e.g., race, gender/sex), and face luminance from noisy classification images.

A promising way forward

Addressing the existing limitations of psychophysical reverse correlation would allow for several important avenues of exploration in future research. First, a reverse correlation approach that yields photorealistic face images would have high face validity. This is because participants would likely be unable to tell that the images they are categorizing (or rating) are synthetic, leading to greater ecological validity. For example, prior work has demonstrated that judging the trustworthiness of computer-generated faces (as compared to real faces) results in weaker observed effects (Balas and Pacella, 2017). Relatedly, computer-generated and artificial faces have been shown to be more difficult to process and remember compared to real faces (Balas and Pacella, 2017; Gaither et al., 2019). Therefore, it is likely that there is less cognitive burden for rating photorealistic face images compared to pixelated images or avatars.

Second, a higher-fidelity approach to reverse correlation would allow for easier and more effective integration with other computational methods in the social sciences. Traditional approaches produce noisy or low-fidelity classification images which are not discernible by pipelines that use computer vision to automatically calculate structural and surface properties of the face in addition to other metrics that may be of interest to researchers. In contrast, real face images can often be substituted one-for-one with photorealistic computer generated face images in algorithms that can automatically detect or estimate facial landmarks, self-identified demographic properties (e.g., race, age, sex/gender), and emotional expression, among other such perceived attributes. In summary, more realistic classification images are ideal not only for human observers to make high-quality judgments, but also for algorithms that can aid in extracting and interpreting lower-level face metrics.

Advances in machine learning applications

One of the most important recent advances to computer vision has been the generation of photorealistic synthetic images. Generative adversarial networks (GAN) accomplish this by pitting two machine learning models in "competition" against each other with the goal of creating better and better output. A generator model produces synthetic output data (often an image) in an attempt to "fool" a separate discriminator model simultaneously trained to discern "real" from "synthetic" data (Goodfellow et al., 2014). Generative adversarial networks have been used to create a variety of image classes, ranging from faces to house facades, cars, and animals, among many others. A generative model learns to produce new images by dynamically updating its output based on whether or not the discriminator model can tell whether its output is a real image or a generated image. Similarly, the discriminator model dynamically improves its performance using the feedback it receives from the generator model.

One GAN that has received considerable attention and research is StyleGAN, a machine learning model capable of producing different types of images at high resolution that are nearly indistinguishable from real world photos (Karras et al., 2018, 2020, 2021). Of particular interest here, StyleGAN has been able to generate human faces with incredible fidelity and precision, mimicking real-world face photographs while also being able to construct new face images of non-existent individuals.

In addition to creating high resolution output, StyleGAN has also proven to be reliable in manipulating faces along a number of dimensions of psychological interest. For example, researchers were able to identify where in the StyleGAN latent space demographic attributes such as age and sex/gender exist. Identifying latent directions for such attributes then allows for any images generated to be moved along those directions and thereby manipulated along age (young to old) or sex/gender (perceived male to perceived female), or both (Shen et al., 2020).

There is nothing limiting the discovery of latent directions within the latent space to such (first-order) demographic features. For example, one recent paper applied modeling techniques previously used for characterizing representations of psychological perceived attributes in 3D computer-generated faces (e.g., Oosterhof and Todorov, 2008) to the StyleGAN2 latent space (Peterson et al., 2022). By collecting over 1 million judgments of 1,004 synthetic faces from over 4,150 participants, the researchers visualized 34 perceived attribute dimensions. These dimensions included such first-order judged dimensions as perceived "masculinity," as well as second-order judged dimensions such as perceived "trustworthiness." These perceived dimensions in particular were able to be modeled with great fidelity, owing to the high inter-rater reliability of participants' judgments. The results are best appreciated with a demonstration: Figure 2 depicts transformations of these two perceived dimensions applied to an averaged neutral face (itself the average of almost 2,500 neutral faces collected from various extant stimulus sets, and described in more detail in the Methods section below).

To our knowledge, this approach represents the state of the art for group-level modeling of social perception in faces (Peterson et al., 2022). However, this set of studies was explicitly designed to characterize the mental representations of faces in the general (online) population and could not model the idiosyncratic representations of individual participants. In principle, the same approach could capture such individualized models simply *via* "scaling up" data collection — by asking a given participant to contribute over a thousand judgments per perceived attribute of interest, it should be possible to characterize their idiosyncratic visual representations. In practice, however, this would prove to be time consuming to the participant, costly to the experimenter, and potentially self-defeating by nature of its greater scale; the approach would likely yield lower and lower quality data as the experiment progresses. Ideally, modeling such perceived dimensions or attributes in face space would not require thousands of trials from each participant. Here, we have developed one promising approach to do so that requires only a few hundred trials, and combines advances in GAN-based image generation with the previously discussed psychophysical reverse correlation techniques.

While our work relies on a particular type of GAN for the generation and transformation of faces, the field of machine learning continues to rapidly advance and new methods emerge regularly. Recently, research groups from companies such as OpenAI (Ramesh et al., 2022) and Google (Saharia et al., 2022)

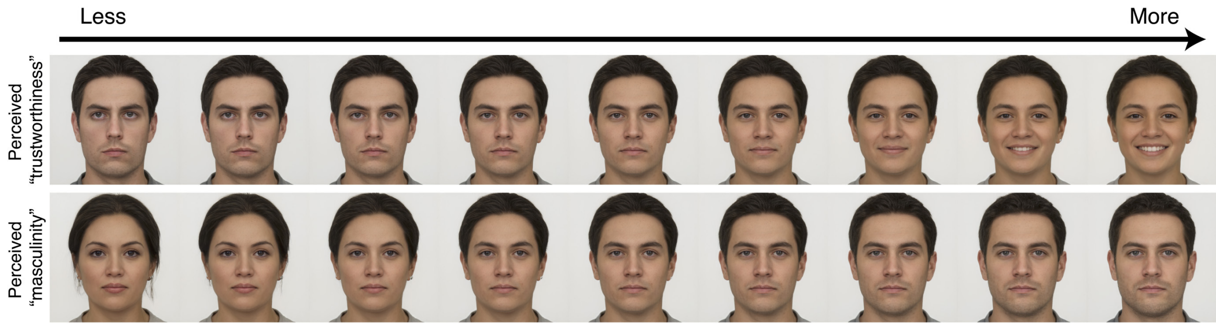


FIGURE 2
An average of 2,484 neutral faces projected into the StyleGAN2 latent space and transformed along dimensions of perceived "trustworthiness" and "masculinity." These dimensions were modeled using average ratings for 1,004 faces each, collected from a general online participant population (Peterson et al., 2022). The faces along each step of the group-level continua are highly realistic and evoke vivid impressions of the transformed perceived attribute.

have created especially powerful generative diffusion models (DALL-E 2 and Imagen, respectively) *via* a technique known as Contrastive-Language-Image-Pretraining (CLIP; Radford et al., 2021). Among other things, these models allow for the creation of arbitrary images from text prompts, which can themselves be quite descriptive (e.g., "a racoon detective in New York City wearing a trench coat under a street light."). In addition, it is possible to quickly generate many variants of a given image, or to edit the content of an image by applying a mask to an area of the image and asking for a desired change with further text (e.g., "add a bed" to a masked area of a scene, which will then be filled with a bed). The availability of such models to the general public is currently limited for various reasons (including concerns of potential abuse by unethical actors) but may prove enormously useful to the broader psychological research community, both for stimulus generation and analysis/exploration of participant data. It is worth noting that although the human faces generated by these types of models are currently not as realistic as those generated by the likes of StyleGAN, it seems likely that they will close the gap in the months to come; future work should explore the utility of such techniques, as they may allow for even more diverse and naturalistic stimuli.

Summary and overview

To summarize the arguments above, typical research in person perception analyzes mental constructs of interest by aggregating over both stimuli and individual participant raters. As such, reported results are only interpretable at the highest group level (e.g., on average a given face image is viewed by a typical participant as "trustworthy"). While aggregated results are informative, emerging research suggests that within person perception a large portion of the variance is attributable to idiosyncratic differences in raters, particularly for complex perceived attributes such as trustworthiness.

Despite both increased attention to idiosyncratic contributions to social perceptions and advances in technology, there is still no reliable, high-fidelity method for visually inspecting individual mental representations of perceived personality attributes. In the present work, we introduce a novel, data-driven method for visualizing hyper-realistic mental representations of social judgments of perceived facial attributes. To accomplish this, we leverage state-of-the-art machine learning models to generate manipulable, highly realistic faces.

We first formally introduce our proposed methodology and then apply it to a small proof-of-concept study. Our main goal is to show that our procedure can visually capture individual social judgments in a predictable manner. For our initial investigation, we utilized both a first-order social judgment ("feminine/masculine") and a second-order social judgment ("trustworthy"). The two judgments were selected intentionally to visualize attributes that should theoretically vary with respect to the shared and idiosyncratic contributions to judgments (see Figure 1). More specifically, while we predicted that all individual mental representations would be idiosyncratic to some degree, we also predicted that first-order social judgments (i.e., feminine/masculine) would vary less across individuals compared with second-order social judgments (i.e., trustworthy). This prediction aligns with previous work on shared and idiosyncratic contributions to social judgments.

Initial proof-of-concept study

Methods

The proposed methodology for our hyper-realistic visualization procedure loosely follows that of a typical psychophysical reverse correlation procedure (Dotsch and Todorov, 2012). However, there are several major differences whereby we leverage state-of-the-art technical innovations. Specifically, our methodological procedure consists of four steps: (1) image inversion; (2) stimulus creation; (3) stimulus selection (by participants); and (4) stimulus analysis (i.e., classification image creation). In what follows, we first briefly introduce each of our methodological steps and then detail the results from a proof-of-concept investigation using our method.

Step 1: Image inversion

The first step consists of inverting a set of faces into the StyleGAN2 latent space. In short, GAN inversion is a process whereby a real face image is reconstructed (located) within a pre-trained GAN latent space (Xia et al., 2022). A successful inversion results in an image that is photorealistic, similar in appearance to the original image, and editable (i.e., maintains the same characteristics of the GAN latent space into which it was inverted so that attributes present in the latent space can be applied to the inverted image). Inverting an image results in a 18×512 matrix of numeric values that represents that face in the StyleGAN2 latent space.

We inverted 2,484 neutral faces from various available databases into the StyleGAN2 latent space using a modified VGG encoder adapted by Peterson et al. (2022). The neutral faces were taken from several face databases: the Chicago Face Database (Ma et al., 2015), FACES (Ebner et al., 2010), NIMSTIM (Tottenham et al., 2009), RAFD (Langner et al., 2010), Face Database (Minear and Park, 2004), Face Research Set London (DeBruine and Jones, 2017), FERET (Phillips et al., 2000), and RADIATE (Conley et al., 2018) image sets, as well as a number of internal face resources. We focused on inverting real neutral face images because the original StyleGAN2 latent space is oversaturated with smiling faces due to the nature of the original training data (Karras et al., 2020). An overrepresentation of smiling faces (or any other type of face/attribute) is undesirable for obtaining an accurate classification image (i.e., individual face prototype or representation). In our tests, an oversampling of smiling faces in the image pool shown to participants resulted in classification images that also overrepresented "smiley" attributes.

Step 2: Stimulus creation

In a typical reverse correlation study, stimuli are created by overlaying random sinusoidal noise over a standardized, singular base image. Much like the original reverse correlation procedure, we sought to create stimuli for our experiment by randomly generating random neutral faces from the GAN latent space and adding a small amount of Gaussian noise. To generate random, unique neutral faces from the latent space, we first averaged together the latents of a subset of 10 randomly selected faces from the 2,484 faces inverted into the model latent space in the previous step. Next, we added a small amount of random Gaussian noise to the averaged latent to further differentiate it from the pool of inverted faces. This two-step process was repeated for each stimulus generated.

We generated 300 neutral face stimuli utilizing the procedure outlined above. We sampled noise from a Gaussian distribution with parameters $\mu = 0$ and $\sigma = 0.4$ to be added to each generated average image. Examples of the generated stimuli with these parameters can be seen in Figure 3B.

Step 3: Stimulus selection

Our stimulus selection procedure involves displaying each image sequentially to participants and asking them to categorize the face based on a specific set of attributes designated by the researcher. In contrast to a standard reverse correlation procedure, which typically uses a two-alternative forced-choice design (i.e., selecting between two images), we propose utilizing a design with three potential categorizations for a single image–that is, a single image is displayed to the participant along with three response options. In our task, each of the categories is as follows: (1) the attribute of interest (e.g., perceived "trustworthy"), (2) the conceptual opposite of this attribute (e.g., perceived "untrustworthy"), and (3) "neutral" (or "neither"). For example, if a researcher is interested in visualizing an individual's mental representation of perceived "masculinity," participants would be asked to select whether they think each face appears "masculine," "feminine," or "neither." The rationale for including a "neutral" or "neither" category is to obtain an unbiased, individual starting point within the latent space for each participant. That is, much like the target categories, what one individual categorizes as "neither" is likely to differ from one participant to the next. The participantHeading 3,H3,APA Level 2s selections are binned into each of the three categories and used for analysis in the next step.

The current experiment had two sets of judgments: "trustworthy/untrustworthy/neither" and "masculine/feminine/neither." Participants were assigned to one of the conditions and tasked with categorizing each face stimulus into one of the three categories. Every participant saw the same 300 faces generated in the previous step, though presentation order was randomized between participants.

Step 4: Stimulus analysis

Stimulus analysis involves matrix arithmetic on the latents of the selected images for each participant. First, the image latents for each selected category are averaged together. Second, the latent matrix of the non-target category is subtracted from the latent matrix of the target category. This process isolates the unique features attributable to the perceived target attribute in the latent space. For example, subtracting the average "untrustworthy" latent matrix from the average "trustworthy" latent matrix yields a new latent matrix that represents the

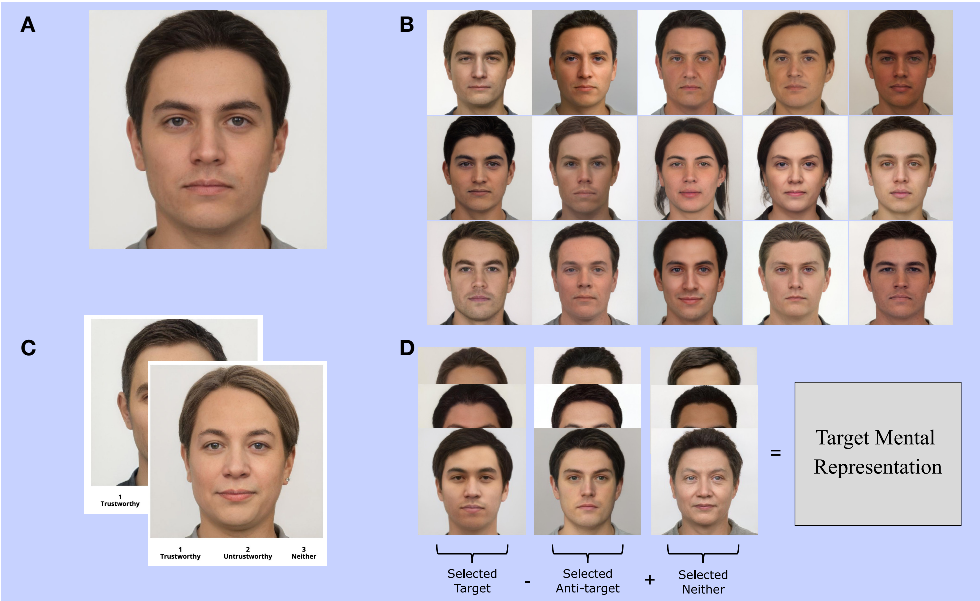


FIGURE 3
Overview of proposed method. **(A)** Depicts the average of 2,484 neutral faces inverted into the StyleGAN2 latent space. **(B)** Provides example neutral face images created by adding random noise to inverted neutral face images. **(C)** Displays example trials participants saw during "stimulus selection" (Step 3). **(D)** Provides an overview of calculating the average directional vector of a target trait applied to an average face to create a target mental representation.

qualities unique to perceived trustworthiness for a given participant. The result of this operation represents a directional vector in the latent space that can be used to sample the mental representation of the desired perceived attribute at varying levels of intensity. This is accomplished by adding the directional vector to the averaged latent matrix of the "neither" selections, which represents a starting point in the StyleGAN2 latent space for estimating an individual's mental model for a particular trait. Consequently, multiplying the directional vector matrix by a constant before adding it to the averaged "neither" latent matrix produces visualizable mental representations for the trait at different levels of intensity.

More formally, $A \in \mathbb{R}^{m \times n}$ is an 18 × 512 dimensional matrix that represents the averaged latents for the faces selected to represent a target trait. Similarly, $B \in \mathbb{R}^{m \times n}$ is a 18 × 512 matrix of the averaged latents for the non-target or non-selected faces. The directional vector matrix, **A**, for a particular subject, *i*, can be computed as,

$$\mathbf{A}_i = A_i - B_i$$

This directional matrix can be applied to the averaged neither latent matrix, $N \in \mathbb{R}^{\mathbf{m \times n}}$, to compute a starting point of an individual's mental representation of the target trait,

$$M_i = N_i + \mathbf{A}_i$$

Finally, when the directional vector matrix is multiplied by a constant, C, the mental representation image, M_{ic}, can be estimated at varying intensities,

$$M_{ic} = N_i + \mathbf{A}_i\, C$$

The extrapolated mental representations exemplify the individual's internal prototypes for the particular trait measured at various levels of intensity.

We computed individual classification images for each of our participants following the procedure outlined above. If participants did not categorize any faces as "neither," a random sample of 20 faces were drawn from the pool of 300 faces and averaged together as a proxy starting point in the latent space. Experiments utilizing different random subsets of images did not meaningfully change the visual results of the output images.

Participants

Eleven participants ($M_{age} = 36$, $SD_{age} = 5.24$) were recruited to complete the study *via* CloudResearch. Participants self-identified as the following: 6 women, 5 men; 8 White, 1 Black, 1 Asian, and 1 Middle Eastern or North African. Six participants were assigned to the "trustworthy/untrustworthy" condition and five participants to the "masculine/feminine" condition. One participant categorized all images as "trustworthy" and was thus not included in our analyses. Two participants did not categorize any face as "neither," though "neither" selections were low across both conditions ($M_{masculine/feminine} = 6.8\%$, $M_{trustworthy} = 10.6\%$). This study was reviewed and approved by the Institutional Review Board at the authors' institution. All participants agreed to take part in this study and were compensated $5 for completing the study.

Participant procedure

Participants were instructed that they would be shown several hundred face images and would be asked to categorize each face into one of three categories depending on which condition they were assigned. Participants were also instructed to take a moment to imagine what a trustworthy/masculine and untrustworthy/feminine face looked like to them (depending on the condition to which they were assigned).

Next, participants completed the main portion of the experiment whereby they categorized each face. During each of the 300 trials, participants were presented with a face toward the center of the screen along with each of the three categories listed beneath the face. Participants were instructed to use the number keys to make their selection (1 = "trustworthy"/"masculine," 2 = "untrustworthy"/"feminine," and 3 = "neither," as was appropriate for their assigned condition). After completing this portion of the experiment, participants answered a basic demographic questionnaire (e.g., listing their age, race, and gender) and were debriefed.

Results and discussion

Figure 4 presents the results from each of the ten participants in our study at several levels of target trait intensity. Individual mental representations were computed following the steps outlined in the Method section. Through experimentation we determined that the directional vector within the latent space could be extrapolated with constants between −8 and +8 without degrading the clarity of the internal face portion of the image or collapsing the latent space (i.e., creating an unrecognizable image). Visual inspection of the produced images suggests that changes between models become readily apparent beginning at +/−4 from the center (or averaged "neither") starting image. Hence, we suggest that internal prototypes should be examined at values +/−4 or greater to secure an adequate visual representation for each individual mental representation.

Visual inspection of these results confirms the face validity of our proposed procedure. Specifically, individuals assigned to the "trustworthy/untrustworthy" condition yielded visualized mental models of faces that appeared happier and more feminine for trustworthy (positive constant values) as well as angrier/stoic and more masculine for untrustworthy (negative constant values). Participants assigned to the "masculine/feminine" condition yielded visualized mental models of faces that appeared more masculine (positive constant values) as well as more feminine (negative constant values).

While interpreting results from such a small sample of participants warrants caution, there are a few additional aspects of the current results that deserve discussion. For one, it is interesting that prototypical, toothy smiles still occurred at the extremes of perceived trustworthiness despite the neutral appearance of all the stimuli that participants categorized. This pattern of results suggests that overt, robust physical features are still able to be derived from faces that only incidentally resemble them.

Most importantly in terms of idiosyncratic contributions to judgments, participants' mental representations of extremely trustworthy faces varied dramatically in physical appearance (e.g., presenting as both "male" and "female" and at differing ages), underscoring the idiosyncrasies of how each participant imagines "trustworthy" in their mind's eye. Conversely, the mental representations of "masculine" and "feminine" at the extremes appear quite similar in physical appearance (e.g., all appearing feminine with round faces and long hair), underscoring agreed upon characteristics for these judgments.

Critically, and in line with our hypotheses, average correlations among the trustworthy/untrustworthy latents [$r_{-8}(8) = 0.53$, $r_8(8) = 0.54$] were lower than the average correlations among the masculine/feminine latents [$r_{-8}(8) = 0.82$, $r_8(8) = 0.84$] at both extremes. Lower correlations among participants in the "trustworthy/untrustworthy" condition suggest that they had more idiosyncratic mental representations for what they considered trustworthy compared to participants in the "masculine/feminine" condition. Similarly, higher correlations among participants in the "masculine/feminine" condition suggest that they had more shared representations of what they considered masculine or feminine.

Discussion

In the present work we highlight the importance of understanding both shared and idiosyncratic contributions to the variance of social judgements. That is, how much of the variance in social judgment ratings can be explained by stimulus-level features (i.e., shared) vs. participant-level preferences (i.e., idiosyncratic). Emerging research suggests that both levels explain an important amount of variance, though the proportion of variance explained by each level appears to be influenced by the type of judgment. For example, shared contributions to judgements of first-order, lower-level judgments such as perceived "masculinity" or "femininity" appear to be larger than idiosyncratic contributions. On the

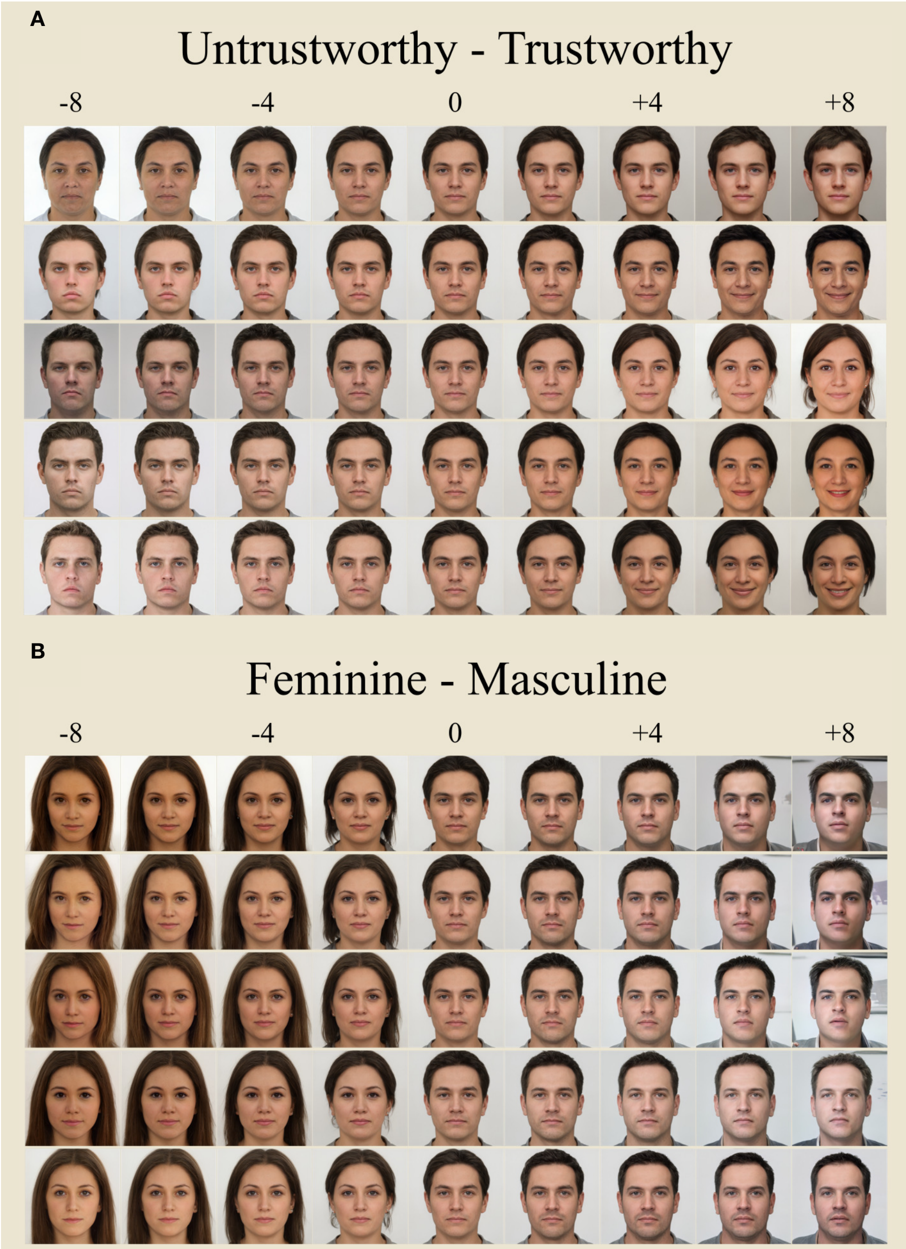


FIGURE 4
(A) Visualizes the mental representations for each participant in the "trustworthy" condition. **(B)** Visualizes the mental representations for each participant in the "masculine" condition. Each row represents a single participant. Each column visualizes mental representations when the directional vector is applied to the average of all faces that were selected as "neither" (0) multiplied by specific constants (−8 to +8).

other hand, second-order, higher-level judgments such as perceived "trustworthiness" or "attractiveness" seem to be better explained by idiosyncratic contributions.

While emerging work has demonstrated the importance of understanding contributions to judgments at multiple levels of interest, there has yet to be a method for visually capturing high-fidelity individual mental representations. Here, we attempt to fill this gap by introducing a novel, data-driven, and hyper-realistic method for visualizing individual mental representations of perceived attributes inferred from (or ascribed to) faces. We leverage current machine learning technology to create a set of photorealistic face stimuli, each of which is manipulable along a specific directional vector of interest. For example, our pipeline can compute directional vectors of a perceived attribute, such as perceived "trustworthiness," at the participant level, which can then be applied to an averaged base face to visualize how an individual represents prototypical "trustworthy" (and anti-trustworthy) faces in their mind's eye.

We conducted a proof-of-concept study utilizing our proposed method whereby we had participants create directional vectors (i.e., mental representations) of "un/trustworthy" and "masculine/feminine" faces. Inspection of our results confirms the face validity of our procedure. Specifically, those assigned to create "trustworthy" mental representations produced face images that appeared happier, younger, and more feminine. Likewise, those assigned to the "masculine/feminine" condition produced face images with sexually dimorphic characteristics (see Figure 4). More importantly, and in line with our predictions, there was much more agreement in the images produced for feminine/masculine-looking faces (a first-order judgment) compared to un/trustworthy-looking faces (a second-order judgment). Such effects were expected given previous work on differential variance contributions to specific types of social judgments.

While correlations between feminine/masculine latents were higher than those between trustworthy/untrustworthy latents, it should be noted that the correlations between trustworthy/untrustworthy latents were still relatively high ($r = 0.5$). These moderate correlations may be due to a number of causes, including "shared" characteristics among created images (e.g., smiling), the entanglement of StyleGAN2 latent space, or the relatively low sample size. One potential cause for moderate correlations among the trustworthy/untrustworthy latents is that many of the trustworthy/untrustworthy images share common features, such as frowning or smiling, which would likely be reflected in a similar pattern among the image latents. Similarly, the StyleGAN2 latent space is known to be entangled, i.e., each latent alters more than one visual attribute as its value changes (Wu et al., 2021). The entangled latent space may result in moderately correlated latents even when they are sampled at random. Finally, we may have observed moderate correlations between the trustworthy/untrustworthy latents due to our sample size. It is likely that all three factors are influencing the relationship between the created image latents and future work should aim at detailing how the latents are related and when highly correlated patterns emerge.

In sum, our proposed pipeline for producing mental representations of social judgments of faces produces high quality, photorealistic visualizations of prototypes at the individual level. Our methodology advances how individual mental representations can be visualized in several important ways. First, our procedure allows for researchers to estimate individual prototypes at various levels of intensity. Our technique allows us to take any number of categorizations and build a vector to move through the latent space in a given direction that represents that individual's internal prototype.

Second, our procedure produces high-fidelity images. The resultant images are photo-realistic and indistinguishable in most cases from non-computer generated images. These images can then be rated by human participants or passed to additional machine learning applications that measure facial attributes or metrics. A method that results in a face image that is indistinguishable from a real face allows for participants to provide an unbiased estimate of the stimulus. Similarly, a face that appears realistic can be interpreted by other machine learning algorithms which would allow additional benefits afforded by them, such as facial landmarks estimation as well as face identification, sex/gender, ethnicity, age, emotion, texture, and color estimation. These additional results would not be possible with lower resolution images.

Finally, higher resolution/fidelity images allow for researchers to identify more subtle differences that can occur across social judgments at the individual level. For example, the images created through our procedure in the anti-masculine (feminine) condition resulted in prototypes that globally looked similar but still differed in minor ways such as hair style, gaze, and mouth curvature. Such differences would likely be indistinguishable when the resultant image is of lower resolution.

Future work

While our preliminary study provides compelling proof-of-concept visualizations, it is important to follow up this work with additional research to confirm the validity of our methodology and determine any boundary conditions that might exist. We identify four important next steps that we plan to undertake to confirm our mental representation pipeline: (1) validate created mental representations by both the creator and naive raters, (2) replicate our results with more participants across a more diverse set of target attributes and judgments; (3) determine the importance of the underlying distribution of faces used to create stimuli; (4) determine the influence of the number of trials on the outcome image.

A critical and missing step in our pipeline is the validation of the individual models. Following the logic of previous validation studies of consensus judgments (Todorov et al., 2013; Todorov and Oh, 2021), participants' judgments should be more sensitive to differences between faces manipulated by their own model than differences between faces manipulated by models of other participants. Specifically, if the intensity of the perceived attribute is manipulated at multiple levels, the slopes of participants' judgments should be steeper for faces manipulated by their own model than by models of other participants. However, we would expect the magnitude of this effect to be attenuated if the judgment of interest is first-order and thus more likely to have higher inter-individual agreement. Further, validating the image transformations *via* ratings from a group of naive participants would also provide insight into whether the generated mental representation images are consensually interpreted as the judgment of interest. In summary, validation by both the classification image creator and other raters is an important validation step for fully understanding the unique contributions of both idiosyncratic and shared variance in face judgments for this proposed methodology.

Second, it is important to establish that our methodology provides consistent and replicable results across a diverse set of target perceived attributes. In our pilot study we provided evidence that our procedure works for both first order and second order judgments of perceived attributes. However, a more diverse portfolio of social judgments that can be visually derived from our pipeline would only further confirm its validity. While we only tested two social judgments and a handful of participants, the results were visually striking. As such, we are confident that studies with additional participants and attributes will provide equally satisfying results.

Third, it is important to understand the effect of the underlying pool of real face images used to provide a foundation for the generated stimuli. In the current pilot we were not concerned with a balanced underlying pool of faces and instead opted to secure a pool of as many high resolution and standardized neutral faces as possible. Therefore, the underlying distribution of faces used to create our experimental stimuli were unbalanced in terms of race, sex/gender, age, and ethnicity. While the social judgments we selected to examine in our pilot study are less likely to be influenced by such imbalances, other target categories that can be examined through our procedure will likely need stimuli drawn from an underlying balanced face distribution. For example, if a researcher is interested in determining whether perceived emotion differs between mental representations of specific ethnic categories, the underlying distribution of faces that the stimuli are created from needs to be balanced in terms of the target ethnicities. A balanced underlying face set would allow for experimental stimuli to be drawn from (or around) an ethnically-ambiguous base face and reduce any unintentional bias that might result from an unbalanced stimulus set.

Finally, the number of trials used in our procedure needs to be experimentally examined to determine a minimal number of trials to produce optimal images. Based on previous work utilizing reverse correlation in social psychology, we determined that our pilot study should use 300 experimental trials. While 300 trials produced visually appealing results, the number of trials could also be increased or decreased to achieve equal or better results. If the number of experimental trials could be reduced while still producing comparable visual results, it would reduce experiment duration and help reduce participant fatigue. Similarly, if increasing the number of trials could increase the final mental prototype image fidelity, it would produce higher quality results. Either way, understanding the optimal number of experimental trials is critical for operationalizing a final procedure for our methodology.

Conclusion

It is unlikely that two individuals will judge the same person identically on how trustworthy, attractive, or dominant they appear. Even if these two individuals did agree on a numeric value, it is even more unlikely that they would agree upon the reasons why they arrived at such conclusions. The only way to truly understand what physical features in a face individuals use to inform their impressions is to visually capture what they imagine for any given social judgment prototype. Until now, such possibilities were limited to low-fidelity, pixelated, avatar representations of faces or self-reports. Now, by leveraging state-of-the-art computer vision, we have provided a methodology for visualizing high-fidelity mental representations of social attributes inferred from faces at the individual participant level. Our work represents a critical advance in understanding how social judgements are formed and how they can dramatically differ or remain consistent from individual to individual.

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 Chicago Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

DA, SU, and AT contributed to the conception and design of the methods and studies. DA conducted the pilot studies and performed statistical analyses. SU contributed to the pilot study design and infrastructure. DA and SU wrote the first draft of the manuscript. All authors revised, read, and approved the submitted version of the manuscript.

Funding

This work was supported by the Richard N. Rosett Faculty Fellowship at the University of Chicago Booth School of Business.

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Albohn, D. N., and Adams, R. B. Jr. (2020). Everyday beliefs about emotion perceptually derived from neutral facial appearance. *Front. Psychol.* 11, 264. doi: 10.3389/fpsyg.2020.00264

Albohn, D. N., Brandenburg, J. C., and Adams, R. B. Jr. (2019). "Perceiving emotion in the "Neutral" face: a powerful mechanism of person perception," in *The Social Nature of Emotion Expression,* eds U. Hess and S. Hareli (Springer International Publishing) 25–47. doi: 10.1007/978-3-030-32968-6_3

Albohn, D. N., Martinez, J. E., and Todorov A. (2022). *What Influences Shared and Individual Variance Contributions to Judgments of Neutral Faces?* [Unpublished].

Balas, B., and Pacella, J. (2017). Trustworthiness perception is disrupted in artificial faces. *Comput. Hum. Behav.* 77, 240–248. doi: 10.1016/j.chb.2017.08.045

Cone, J., Brown-Iannuzzi, J. L., Lei, R., and Dotsch, R. (2020). *Type I error is inflated in the two-phase reverse correlation procedure. Soc. Psychol. Personal. Sci.* 12, 1948550620938616. doi: 10.1177/1948550620938616

Conley, M. I., Dellarco, D. V., Rubien-Thomas, E., Cohen, A. O., Cervera, A., Tottenham, N., et al. (2018). The racially diverse affective expression (RADIATE) face stimulus set. *Psychiatry Res.* 270, 1059–1067. doi: 10.1016/j.psychres.2018.04.066

DeBruine, L., and Jones, B. (2017). *Face Research Lab London Set*. Available online at: https://figshare.com/articles/dataset/Face_Research_Lab_London_Set/5047666 (accessed January 7, 2017).

Dotsch, R., and Todorov, A. (2012). Reverse correlating social face perception. *Soc. Psychol. Personal. Sci.* 3, 562–571. doi: 10.1177/1948550611430272

Dotsch, R., Wigboldus, D. H. J., Langner, O., and van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychol. Sci.* 19, 978–980. doi: 10.1111/j.1467-9280.2008.02186.x

Ebner, N. C., Riediger, M., and Lindenberger, U. (2010). FACES—a database of facial expressions in young, middle-aged, and older women and men: development and validation. *Behav. Res. Methods* 42, 351–362. doi: 10.3758/BRM.42.1.351

Gaither, S. E., Chen, J. M., Pauker, K., and Sommers, S. R. (2019). At face value: Psychological outcomes differ for real vs. computer-generated multiracial faces. *J. Soc. Psychol.* 159, 592–610. doi: 10.1080/00224545.2018.1538929

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). *Generative Adversarial Networks*. Available online at: http://arxiv.org/abs/1406.2661

Gosselin, F., and Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Res.* 41, 2261–2271. doi: 10.1016/S0042-6989(01)00097-9

Gosselin, F., and Schyns, P. G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychol. Sci.* 14, 505–509. doi: 10.1111/1467-9280.03452

Hehman, E., Sutherland, C. A. M., Flake, J. K., and Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *J. Personal. Soc. Psychol.* 113, 513–529. doi: 10.1037/pspa0000090

Hönekopp, J. (2006). Once more: is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 199–209. doi: 10.1037/0096-1523.32.2.199

Isik, A. I., and Vessel, E. A. (2019). Continuous ratings of movie watching reveal idiosyncratic dynamics of aesthetic enjoyment. *PLoS ONE* 14, e0223896. doi: 10.1371/journal.pone.0223896

Jaeger, B., and Jones, A. L. (2021). Which Facial Features Are Central in Impression Formation? *Soc. Psychol. Personal. Sci.* 194855062110349. doi: 10.1177/19485506211034979

Jirschitzka, J., Oeberst, A., Göllner, R., and Cress, U. (2017). Inter-rater reliability and validity of peer reviews in an interdisciplinary field. *Scientometrics* 113, 1059–1092. doi: 10.1007/s11192-017-2516-6

Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., et al. (2021). *Alias-Free Generative Adversarial Networks*. Available online at: http://arxiv.org/abs/2106.12423

Karras, T., Laine, S., and Aila, T. (2018). *A Style-Based Generator Architecture for Generative Adversarial Networks*. Available online at: http://arxiv.org/abs/1812.04948

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). *Analyzing and Improving the Image Quality of StyleGAN*. Available online at: http://arxiv.org/abs/1912.04958

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., and Van Knippenberg, A (2010). Presentation and validation of the radboud faces database. *Cogn. Emotion* 24, 1377–1388. doi: 10.1080/02699930903485076

Lavan, N., Mileva, M., Burton, M., Young, A., and McGettigan, C. (2020). Trait evaluations of faces and voices: comparing within- and between-person variability. *J. Exp. Psychol. Gen.* 150, 1854–1869. doi: 10.1037/xge0001019

Leder, H., Goller, J., Rigotti, T., and Forster, M. (2016). Private and shared taste in art and face appreciation. *Front. Hum. Neurosci.* 10, 155. doi: 10.3389/fnhum.2016.00155

Lloyd, E. P., Sim, M., Smalley, E., Bernstein, M. J., and Hugenberg, K. (2020). Good cop, bad cop: race-based differences in mental representations of police. *Personal. Soc. Psychol. Bulletin* 46, 1205–1218. doi: 10.1177/0146167219898562

Ma, D. S., Correll, J., and Wittenbrink, B. (2015). The Chicago face database: a free stimulus set of faces and norming data. *Behav. Res. Methods* 47, 1122–1135. doi: 10.3758/s13428-014-0532-5

Mangini, M. C., and Biederman, I. (2004). Making the ineffable explicit: estimating the information employed for face classifications. *Cogn. Sci.* 28, 209–226. doi: 10.1207/s15516709cog2802_4

Marsh, A. A., Adams, R. B. Jr., and Kleck, R. E. (2005). Why do fear and anger look the way they do? Form and social function in facial expressions. *Personal. Soc. Psychol. Bulletin* 31, 73–86. doi: 10.1177/0146167204271306

Martinez, J. E., Funk, F., and Todorov, A. (2020). Quantifying idiosyncratic and shared contributions to judgment. *Behav. Res. Methods* 52, 1428–1444. doi: 10.3758/s13428-019-01323-0

Minear, M., and Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behav. Res. Methods* 36, 630–633. doi: 10.3758/BF03206543

Oosterhof, N. N., and Todorov, A. (2008). The functional basis of face evaluation. *Proc. Natl. Acad. Sci. U. S. A.* 105, 11087–11092. doi: 10.1073/pnas.0805664105

Peterson, J. C., Uddenberg, S., Griffiths, T. L., Todorov, A., and Suchow, J. W. (2022). Deep models of superficial face judgments. *Proc. Natl. Acad. Sci. USA*. 119, e2115228119. doi: 10.1073/pnas.2115228119

Phillips, P. J., Moon, H., Rizvi, S. A., and Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Machine Intell.* 22, 15. doi: 10.1109/34.879790

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., et al. (2021). *Learning Transferable Visual Models From Natural Language Supervision*. Available online at: http://arxiv.org/abs/2103.00020

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., and Chen, M. (2022). *Hierarchical Text-Conditional Image Generation with CLIP Latents*. Available online at: https://arxiv.org/abs/2204.06125

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., et al. (2022). Photorealistic text-to-image diffusion models with deep language understanding. *ACM Siggraph*. 15, 1–10. doi: 10.1145/3528233.3530757

Schyns, P. G., Gosselin, F., and Smith, M. L (2009). Information processing algorithms in the brain. *Trends Cogn. Sci.* 13, 20–26. doi: 10.1016/j.tics.2008.09.008

Sekuler, A. B., Gaspar, C. M., Gold, J. M., and Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Curr. Biol.* 14, 391–396. doi: 10.1016/j.cub.2004.02.028

Shen, Y., Yang, C., Tang, X., and Zhou, B (2020). *InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs.* Available online at: http://arxiv.org/abs/2005.09635

Specker, E., Forster, M., Brinkmann, H., Boddy, J., Immelmann, B., Goller, J., et al. (2020). Warm, lively, rough? Assessing agreement on aesthetic effects of artworks. *PLoS ONE* 15, e0232083. doi: 10.1371/journal.pone.0232083

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., and Falvello, V. B (2013). Validation of data-driven computational models of social perception of faces. *Emotion* 13, 724–738. doi: 10.1037/a0032335

Todorov, A., Dotsch, R., Wigboldus, D. H. J., and Said, C. P (2011). Data-driven methods for modeling social perception: modeling social perception. *Soc. Personal. Psychol. Compass* 5, 775–791. doi: 10.1111/j.1751-9004.2011.00389.x

Todorov, A., and Oh, D. (2021). "The structure and perceptual basis of social judgments from faces," in *Advances in Experimental Social Psychology* (Elsevier) S0065260120300290. doi: 10.1016/bs.aesp.2020.11.004

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res.* 168, 242–249. doi: 10.1016/j.psychres.2008.05.006

Tskhay, K. O., and Rule, N. O. (2015). Emotions facilitate the communication of ambiguous group memberships. *Emotion* 15, 812–826. doi: 10.1037/emo0000077

Vessel, E. A., Maurer, N., Denker, A. H., and Starr, G. G. (2018). Stronger shared taste for natural aesthetic domains than for artifacts of human culture. *Cognition* 179, 121–131. doi: 10.1016/j.cognition.2018.06.009

Wu, Z., Lischinski, D., and Shechtman, E. (2021). StyleSpace analysis: disentangled controls for StyleGAN image generation. *2021 IEEE/CVF Conf. Comput. Vision Pattern Recogn.* 12858–12867. doi: 10.1109/CVPR46437.2021.01267

Xia, W., Zhang, Y., Yang, Y., Xue, J.-., H., Zhou, B., and Yang, M.-H. (2022). *GAN Inversion: A Survey*. Available online at: http://arxiv.org/abs/2101.05278 (accessed May 13, 2022).

Zebrowitz, L. A. (2017). First impressions from faces. *Curr. Directions Psychol. Sci.* 26, 237–242. doi: 10.1177/0963721416683996

Zhan, J., Liu, M., Garrod, O. G. B., Daube, C., Ince, R. A. A., Jack, R. E., et al. (2021). Modeling individual preferences reveals that face beauty is not universally perceived across cultures. *Curr. Biol.* 31, 2243–2252.e6. doi: 10.1016/j.cub.2021.03.013

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 13 October 2022
DOI 10.3389/fpsyg.2022.948142

Check for updates

OPEN ACCESS

EDITED BY
Yoshiyuki Ueda,
Kyoto University,
Japan

REVIEWED BY
Cynthia Whissell,
Laurentian University,
Canada
James Strachan,
Italian Institute of Technology (IIT), Italy

*CORRESPONDENCE
Giulia Mattavelli
giulia.mattavelli@iusspavia.it

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

RECEIVED 20 May 2022
ACCEPTED 14 September 2022
PUBLISHED 13 October 2022

CITATION
Mattavelli G, Gorrino I, Cesana E,
De Angelis J and Ricciardelli P (2022)
Illumination and gaze effects on face
evaluation: The Bi-AGI database.
Front. Psychol. 13:948142.
doi: 10.3389/fpsyg.2022.948142

Illumination and gaze effects on face evaluation: The Bi-AGI database

Giulia Mattavelli[1,2]*, Irene Gorrino[1], Elisabetta Cesana[3], Jacopo De Angelis[3] and Paola Ricciardelli[3,4]

[1]IUSS Cognitive Neuroscience (ICoN) Center, Scuola Universitaria Superiore IUSS, Pavia, Italy, [2]Istituti Clinici Scientifici Maugeri IRCCS, Cognitive Neuroscience Laboratory of Pavia Institute, Pavia, Italy, [3]Department of Psychology, Università di Milano-Bicocca, Milan, Italy, [4]Milan Centre for Neuroscience, Milan, Italy

Face evaluation and first impression generation can be affected by multiple face elements such as invariant facial features, gaze direction and environmental context; however, the composite modulation of eye gaze and illumination on faces of different gender and ages has not been previously investigated. We aimed at testing how these different facial and contextual features affect ratings of social attributes. Thus, we created and validated the Bi-AGI Database, a freely available new set of male and female face stimuli varying in age across lifespan from 18 to 87 years, gaze direction and illumination conditions. Judgments on attractiveness, femininity-masculinity, dominance and trustworthiness were collected for each stimulus. Results evidence the interaction of the different variables in modulating social trait attribution, in particular illumination differently affects ratings across age, gaze and gender, with less impact on older adults and greater effect on young faces.

KEYWORDS

face evaluation, gaze direction, age, illumination, face database

Introduction

Face perception is a widely explored human ability in experimental psychology. Faces convey crucial information for social interaction and humans are highly skilled in face processing. Indeed, few milliseconds of exposure are sufficient to detect and evaluate facial features (Bar et al., 2006; Willis and Todorov, 2006) and neuroimaging studies suggest that brain areas in the occipital and temporal cortices are specifically deputed to face processing (Haxby et al., 2000; Ishai, 2008). Face images are stimuli commonly used in many studies investigating specific aspects of face processing, or addressing questions related to social interaction and attention. Yet, the appearance of the same face can change from one image to another in terms of pictures properties and facial characteristics: for example, lighting, head rotation and expression can vary, while the features of the camera (e.g., focal length, shutter, lens settings) influence the overall image quality (Burton, 2013). The interplay between these sources of variation contributes to the development of subsequent impressions, which may actually diverge between images of the same face (Todorov and Porter, 2014).

To address researchers' needs for face stimuli varied in different features, many databases are available in the literature (a collection of references and links could be found at www.face-rec.org). Facial expression is one of the most extensively varied condition in databases presenting models posing different emotions (e.g., FEEST set, Young et al., 2002; KDEF, Lundqvist et al., 1998; Bosphorus database, Savran et al., 2008; Radboud Faces Database, Langner et al., 2010). Most databases have young/middle-age models portrayed with direct gaze, but there are also available stimuli sets with models of different ages across lifespan (Minear and Park, 2004; O'Reilly et al., 2016) and the FACES-database (Ebner et al., 2010) included young, middle-age and older models posing six different emotions. Age is indeed a crucial dimension in face processing, it is linked with attractiveness (Thornhill and Gangestad, 1999; Perrett et al., 2002) and face preference (Ebner, 2008; Ito et al., 2022), it correlates with perceived height, masculinity and dominance (Batres et al., 2015), and a same-age effect is reported for face recognition (Bäckman, 1991); namely, better performances in recognizing individuals similar to own age. Moreover, a recent factor analysis carried out to model the structure of perception of personality traits from highly variable face stimuli (Sutherland et al., 2013) resulted in a three-dimensional model with approachability, dominance and youthful-attractiveness as factors predicting face evaluation. These studies showed that age is a relevant dimension in face processing, thus its impact should be considered in studying human face perception.

Another critical aspect perceived from faces during human interaction is gaze direction. In light of its essential social function, the ability to process gaze cues has been extensively studied in neuroimaging, developmental, and social cognition research (Allison et al., 2000; Emery, 2000; Frischen et al., 2007). Several studies have shown that gaze direction affects facial expression processing (Adams and Kleck, 2003, 2005; Sander et al., 2007; Milders et al., 2011) and social judgments (Mason et al., 2005; Bayliss and Tipper, 2006; Mattavelli et al., 2021). In particular, faces with gaze directed to the perceivers are rated as more dominant (Main et al., 2009), more attractive and trustworthy (Ewing et al., 2010; Kaisler and Leder, 2016; Mattavelli et al., 2021) than faces with averted eyes. Eye gaze is also an effective cue to orient visuospatial attention towards the same direction where other people are attending, a phenomenon known as joint attention (Driver et al., 1999). In this regard, different experiments have explored joint attention in healthy and clinical populations by means of the gaze-cueing paradigm, namely tasks in which participants are asked to detect or discriminate a target presented laterally to a face gazing congruently or incongruently to the region in which the target will appear (Frischen et al., 2007; Mattavelli et al., 2021). Faces with averted gaze direction have been also used in studies exploring attentional resources and memory processes (Frischen and Tipper, 2004, 2006; Artuso et al., 2012; Ricciardelli et al., 2012), and investigating the impact of emotion and gaze cue on the attribution of social traits to faces (Manssuer et al., 2015a,b; Mattavelli et al., 2021). Despite this wide use of faces with manipulated eye gaze in different experimental paradigms, most authors create their own gaze-cue stimuli with image editing software, which is a time-consuming procedure for researchers and the final effect could be not completely naturalistic. Few databases portrayed faces with right- and left-directed eye gaze. Specifically, the Radboud Faces Database (Langner et al., 2010) includes images of adults and children with eight emotional expressions and three gaze directions; Weidenbacher et al. (2007) presented a database of 20 individuals with various combinations of head pose and eye gaze; the ADFES by Van Der Schalk et al. (2011) includes young adults with different expression and head orientation. However, to the best of our knowledge there are no databases of face stimuli with different gaze directions covering different ages across lifespan, and such a stimulus set would be a valuable resource for future work examining in greater detail how perceived age modulates gaze following behaviour (see Ciardo et al., 2014).

Moving from features of the face to contextual variables, it has been shown that within-person variability among different pictures has a critical impact on face processing, in particular when ambient images are employed, namely pictures of naturally occurring faces with surrounding environment (Burton, 2013). Face judgments in terms of attractiveness or social attribution as trustworthiness, competence and intelligence can show within-person variability as much as between-person variability, and when different ambient pictures of the same face are presented for matching tasks, they are often wrongly sorted as belonging to different people (Jenkins et al., 2011; Todorov and Porter, 2014). Previous studies have investigated the effect of changing in illumination condition on face recognition (Hill and Bruce, 1996; Braje et al., 1998; Braje, 2003). Indeed, illumination can affect the amount and/or type of information gathered from faces, with impact on human visual processing, including colour perception (e.g., Shin et al., 2004; Kelber et al., 2017), visual acuity (e.g., Sheedy et al., 1984; Ferwerda, 1998; Hiraoka et al., 2015) and contrast sensitivity (e.g., Amesbury and Schallhorn, 2003; Alghwiri and Whitney, 2012; Wood, 2020). Consistently, studies on facial memory tests suggested an overall negative impact on face identification when illuminated by dimmer light (DiNardo and Rainey, 1989; Nyman et al., 2019; Lim et al., 2022). Moreover, lighting is accurately controlled in professional photography, different light effects are used to emphasize face properties in portraits (Arena, 2012), and specific methods are applied to manipulate and digitally correct the desired effects (Shu et al., 2017). Two examples of widely used artistic effects in portraits are the Rembrandt and split styles, which create particular shadows under the eye or on half face, respectively (Jin et al., 2010). Interestingly, the Rembrandt effect derives from the style of the famous Dutch painter who made portraits with the same features, and it has been shown that such particular position and lighting could impact aesthetic/social judgments and emotional expressions with differences related to hemispheric laterality and gender of faces (Schirillo, 2000, 2014; Powell and Schirillo, 2011). These studies opened interesting questions on the relationship between illumination effects, that can be created with professional photography, and face evaluation. Some face databases are available with pictures taken with different illumination conditions (e.g., Martinez and Benavente, 1998; Georghiades et al., 2001; Sim et al., 2003; Gao et al., 2004), however they do not report the effect of this variable on face evaluation. Crucially, the impact of illumination conditions on different age and gender faces requires further investigation.

The different research areas on face perception reviewed above, evidence that face judgments result from the composite processing of different facial features and contextual factors. In order to investigate the possible combined influence of age, gaze direction and photographic effects on face judgment and person perception (Fiske et al., 2007). We developed a new set of stimuli evaluated for different social dimensions. We present here the University of Milano-Bicocca Age, Gaze and Illumination (Bi-AGI) database, a new set of face images with male and female models of wide age range, portrayed with three different photographic lighting conditions and three gaze directions. Stimuli were rated on the social dimensions of trustworthiness, dominance, attractiveness and femininity-masculinity. The study aimed at providing a high-quality set of face stimuli for future studies in face perception, social cognition and cognitive and social neuroscience. The ratings collected for stimuli validation were also analysed to empirically explore how the different facial features such as gaze direction, gender and age combined with different illumination conditions affect the perception of social attributes.

Materials and methods

Image set and apparatus

The database comprises 270 portrait images of 30 Caucasian models selected from three age ranges with five males and five females in each subgroup: young adults (mean age = 22.29, *sd* = 2.91, range 18–25 years), middle-age adults (mean age = 38.84, sd = 3.55, range 35–45 years), older adults (mean age = 72.16, *sd* = 11.7, range 55–87 years). We considered the criterion >55 years old for the older adults age class, to provide the database with stimuli covering a wide age range across lifespan. All models were asked to maintain a neutral expression while they were portrayed nine times: three gaze directions (direct, right-oriented, left-oriented) each in three illumination conditions (flat, Rembrandt, split). Models were asked not to wear glasses or hat, but no other constraints were specified for hairdos, make up or earrings to maintain natural occurring variance in face images. The study was approved by the Ethic Committee of the University Milano-Bicocca and all models signed informed consent and consented to the use of their images for experimental research.

High-quality digital photographs were taken with a professional camera digital Reflex Canon Eos 6D with a full-frame sensor, a 1:1 lens crop factor and a Sigma 24–60 mm f/2.8 lens assembled on the camera body. The camera was fixed on a tripod while models stood against a white opaque background. The eye gaze was manipulated asking models to look towards the camera or above their left or right shoulder without turning the head. Lighting conditions were obtained by means of an external Nissin D622 flash assembled on the body camera or separated as a lateral flashing light triggered by a wireless Yongnuo RF-603C II. In the flat condition, the flash was on the body camera together with a softbox to create a uniform and natural light diffusion (see Figure 1A). In the split condition, the external flash was located at 90° on the right side of the models producing a darkened effect on the left side of model face (see Figure 1B). In the Rembrandt condition, the external flash was located at 75° on the right side of the models producing a partial darkness on the model face with the characteristic illuminated triangle under the eye (see Figure 1C). All the pictures included in the database are in color mode RGB with a 72 × 72 ppi resolution.

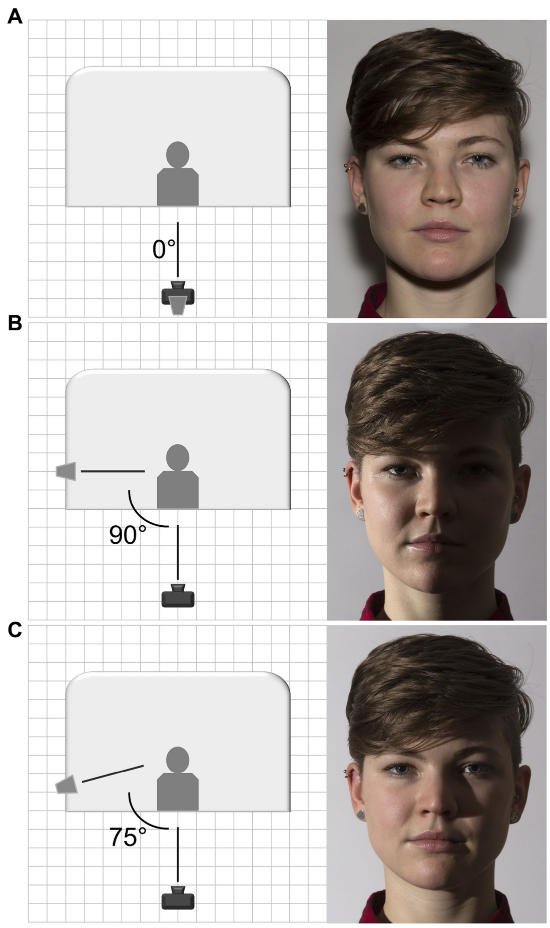


FIGURE 1
Photograph setting and example of one image for the flat **(A)**, split **(B)**, and Rembrandt **(C)** lighting conditions.

Stimuli validation and statistical analyses

One hundred thirty-five volunteers (69 females, 66 males, mean age 42.79, *sd* = 16.62, range 20–86 years, mean education = 13.33, *sd* = 3.37) participated in the validation study. Ratings for four socially relevant dimensions were collected: trustworthiness, dominance, attractiveness and femininity-masculinity. These ratings were presented in four separated blocks in counterbalanced order across participants; within each block the 270 images were presented in random order. Data were collected using E-prime software (Psychology Software Tools, Pittsburgh, PA, USA). The images, resized to 170 × 232 pixels, were presented at the center of the screen with below the 7-point Likert scale with labels of the two extremes of the dimension above the number 1 and number 7 (not trustworthy at all–very trustworthy; not dominant at all–very dominant; not attractive at all–very attractive; very feminine–very masculine). The participants were asked to rate each image pressing the corresponding button from 1 to 7.

Furthermore, since we asked the model to pose with a neutral expression, we were interested to verify whether faces were actually perceived as expressing neutral emotion. Thus, a different sample of 30 healthy participants (15 females, age range 18–30 years old) was involved in an emotion recognition task. Only face stimuli with direct gaze were presented in the three illumination conditions for a total of 90 trials. Stimuli were presented at the center of the screen with below the seven options of emotional expression (neutral, surprise, happiness, fear, disgust, anger, sadness) and the corresponding button to press (from 1 to 7). The order of the emotional labels was counterbalanced across participants, who were asked to press the button corresponding to the expression of the presented faces.

Statistical analyses were performed in R programming environment (R Development Core Team, 2008) throughout R studio (version 2022.2.3.492). Rating scores for attractiveness, femininity-masculinity, dominance and trustworthiness were submitted to a series of linear mixed-effects regressions using the LMER procedure (Baayen et al., 2008), introducing illumination (flat, Rembrandt, split), gaze (central, averted), gender (male, female), and age (young, middle-age, older adults) of face stimuli as fixed factors, while the random-effects structure included by-subject (i.e., participants who rated the stimuli) and by-model (i.e., identity of face images) intercepts to account for inter-subject and inter-stimuli variability. In line with previous experiments (Mattavelli et al., 2021), evaluations of faces looking to the right- or left-side were considered as a unique 'averted' level compared to 'central' gaze to assess the overall impact of faces looking at (*vs* away from) the observer. Likelihood ratio tests were used to evaluate whether the introduction of the *fixed factors* and random effects significantly increased the models' goodness of fit (Gelman and Hill, 2006), then only factors significant as main effect, or in interaction with other factors, were included in the final model. The "phia" R package (version 0.2–1, De Rosario-Martinez, 2015) was used for *post-hoc* pairwise contrasts on significant main effects and interactions on the final best-fitting models, applying Bonferroni correction for multiple comparisons. For the sake of simplicity, we report in the results section for each rating the significant main effects and higher order interactions looking at contrasts between illumination conditions. Tables on model selection and *post-hoc* contrasts are reported in Supplementary material. Appendix A includes mean rating values for each model separately.

Results

Control experiment on emotional expression

The aim of the experiment was to check if faces included in the database were perceived with a neutral expression. To do this, we compared participant responses (i.e., RESPs) with an expected distribution of equiprobability. Specifically, the equal distribution of the responses would show that participants attributed an emotional value to the faces rather than a neutral one. Conversely, a skewed distribution, with a greater frequency of neutral responses, would show a greater tendency of the participants to attribute a neutral valence to faces. Thus, a Chi-Square test of goodness was applied to determine whether RESPs distribution was likely to come from a distribution where RESPs were equally distributed or not – given a total of 7 RESPs the equal distribution would result into a 14.28% of responses for each emotion.

All the assumptions of the Chi-Square test were met and the test was applied by setting the threshold for statistical significance at 0.05. The results of the test led to reject the null hypothesis of equal distribution and to accept the alternative one supporting a greater frequency for specific RESPs categories ($\chi^2(6) = 1121.108$, $p < 0.0001$). Responses distribution is shown in the Supplementary Table S1. This shows an asymmetry towards neutral responses (30.4%) and none of the stimuli was excluded from the database on the basis of emotional expression control experiment. All the percentage values corresponding to emotional categories are indeed below 14.28% except for happiness (26.6%). Similar results were obtained by applying a binomial test to each neutral-emotional RESPs pair (Supplementary Table S2) and setting the equal distribution percentage at 50%. As in the case of the chi-squared results, the percentage of neutral responses is significantly higher than "emotional" responses except for the case of happiness.

Attractiveness

The final model on the attractiveness rating included the main effects of illumination [$\chi^2(2) = 23.33$, $p < 0.001$], gender [$\chi^2(1) = 10.98$, $p < 0.001$], age [$\chi^2(2) = 39.68$, $p < 0.001$] and gaze [$\chi^2(1) = 169.41$, $p < 0.001$] as well as the three-way interactions illumination x gender x age [$\chi^2(4) = 12.7$, $p = 0.012$] and gender x age x gaze [$\chi^2(2) = 6.19$, $p = 0.045$] (see Supplementary Table S3 for the model selection). Faces with central gaze were rated as more attractive than faces with averted gaze, and females received higher scores than males. *Post-hoc* tests on main effects showed that flat and Rembrandt illumination received higher score than split illumination ($ps < 0.001$), whereas young models received higher scores than middle age ($p = 0.03$) and old adults ($p < 0.001$), and middle age received higher scores than older adults ($p < 0.001$). *Post-hoc* comparison for the interaction illumination x gender x age showed that split illumination reduced attractiveness score compared to flat illumination in young ($p < 0.001$) and middle age ($p = 0.002$) female models, whereas middle age male models received lower scores for split compared to both flat ($p = 0.001$) and Rembrandts ($p = 0.02$) illumination. Pairwise contrasts on central vs. averted gaze for the gender x age x gaze interaction, revealed significant higher score for central gaze in male and female models of all age classes (all $ps < 0.001$) (see Figure 2).

Femininity-masculinity

The final model on the femininity-masculinity rating included the four-way interaction [$\chi^2(4) = 154.14$, $p < 0.001$]. The main effects of illumination [$\chi^2(2) = 6.03$, $p = 0.049$], gender [$\chi^2(1) = 726.17$, $p < 0.001$] and age [$\chi^2(2) = 21.87$, $p < 0.001$] were significant (see Supplementary Table S4 for the model selection). As expected, females were clearly rated as more feminine than male models, moreover *post-hoc* tests for illumination main effect revealed a trend for a higher masculine evaluation of models with split than flat illumination ($p = 0.06$) and *post-hoc* on age effect revealed that older adults were rated as more masculine than middle-age ($p = 0.03$) and young adults ($p < 0.001$), and middle-age were rated as more masculine than young adults ($p = 0.02$). Contrasts on the four-ways interaction showed that illumination impacted scores of male and female middle-age models with central gaze, since Rembrandt condition increased masculinity score of females (i.e., females appeared less feminine), and oppositely decreases masculinity score of males (i.e., males appeared less masculine), compared to flat and split conditions (all $ps < 0.001$). Moreover, in the case of averted gaze, female young models were rated as more masculine with split compared to flat illumination ($p = 0.045$) (see Figure 3).

Dominance

The final model on the dominance rating included the four-way interaction [$\chi^2(4) = 40.63$, $p < 0.001$]. The main effects of illumination [$\chi^2(2) = 11.58$, $p = 0.003$] and gaze [$\chi^2(1) = 30.99$, $p < 0.001$] were significant (see Supplementary Table S5 for the model selection). Faces with central gaze were rated more dominant, and scores increased for split compared to flat ($p = 0.02$)

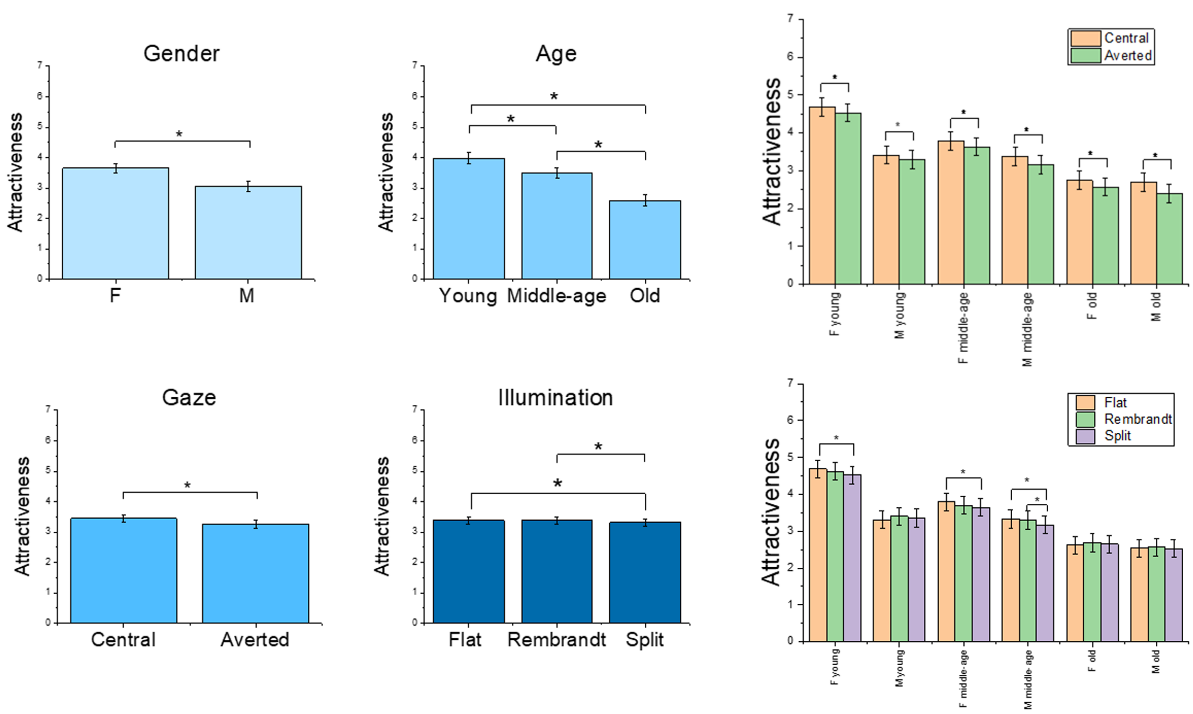


FIGURE 2
Plot of the estimated mean scores from the attractiveness rating, depicting significant main effects and higher order interactions. The asterisks highlight significant main effects and significant results from the *post hoc* analyses. C=central gaze, A=averted gaze, F=female models, M=male models. Error bars represent mean standard errors.

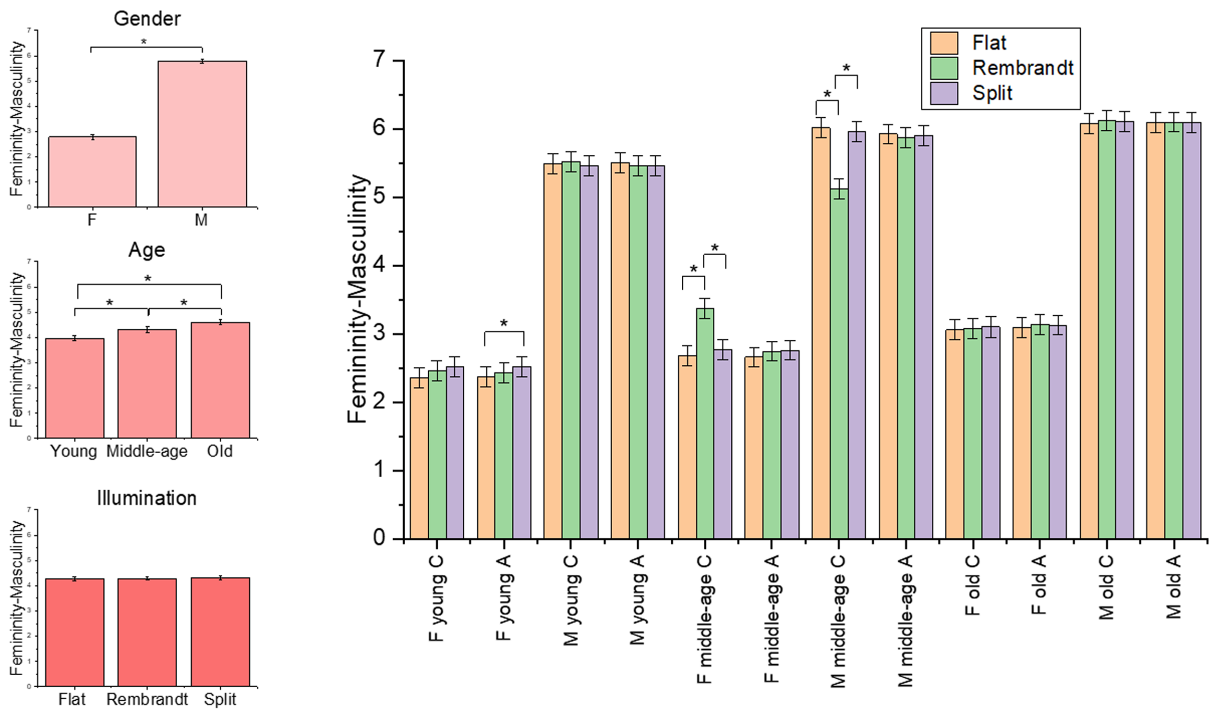


FIGURE 3
Plot of the estimated mean scores from the femininity-masculinity rating, depicting significant main effects and higher order interaction. The asterisks highlight the significant main effect and significant results form *post hoc* analyses. C=central gaze, A=averted gaze, F=female models, M=male models. Error bars represent mean standard errors.

and Rembrandt ($p = 0.02$) illumination. *Post-hoc* analyses on the interaction revealed that in middle-age female models, Rembrandt illumination increased dominance scores compared to flat illumination in case of central gaze ($p = 0.03$), whereas split illumination increased dominance compared to Rembrandt illumination in case of averted gaze ($p = 0.01$). On the other hand, male middle-age models with central gaze were rated more dominant with flat and split illumination compared to Rembrandt condition ($p < 0.001$ and $p = 0.001$, respectively) (see Figure 4).

Trustworthiness

The final model on the trustworthiness rating included the four-way interaction [$\chi^2(4) = 13.24$, $p = 0.01$]. The main effects of illumination [$\chi^2(2) = 31.42$, $p < 0.001$] and gaze [$\chi^2(1) = 829.03$, $p < 0.001$] were significant (see Supplementary Table S6 for the model selection). Faces with central gaze were rated more trustworthy than faces with averted gaze, and scores were lower for split compared to flat and Rembrandt illumination ($ps < 0.001$). Contrasts between illumination conditions for the four-way interaction showed that in the case of female models, illumination affected trustworthiness score only in the younger age class, in which scores were lower for split compared to flat ($p = 0.02$) and Rembrandt condition ($p = 0.04$) when faces had central gaze, whereas scores were lower for split compared only to flat ($p = 0.01$) condition when faces had averted gaze. Different modulations were present in male models: faces with split illumination were rated less trustworthy compared to flat illumination in older adults with central gaze ($p = 0.02$) and in middle-age adults with averted gaze ($p = 0.005$); significant lower scores for split compared to Rembrandt illumination resulted in middle-age males with central gaze ($p < 0.001$) and in young males with averted gaze ($p = 0.007$), moreover Rembrandt illumination received higher scores than flat illumination in middle-age males with central gaze ($p = 0.03$) (see Figure 5).

Discussion

This study presents the Bi-AGI Database, a new set of 30 individual faces with male and female models of different age across lifespan portrayed in three different lighting conditions with central and averted eye gaze. The stimuli were validated with rating questionnaires assessing how illumination conditions and gaze direction, across different ages and gender, affect face evaluation in terms of attractiveness, femininity-masculinity, dominance and trustworthiness. A control experiment was also performed to verify that faces were perceived with neutral expression as expected.

Data from the control experiment confirmed a predominantly neutral perception of faces, although the percentage of happy

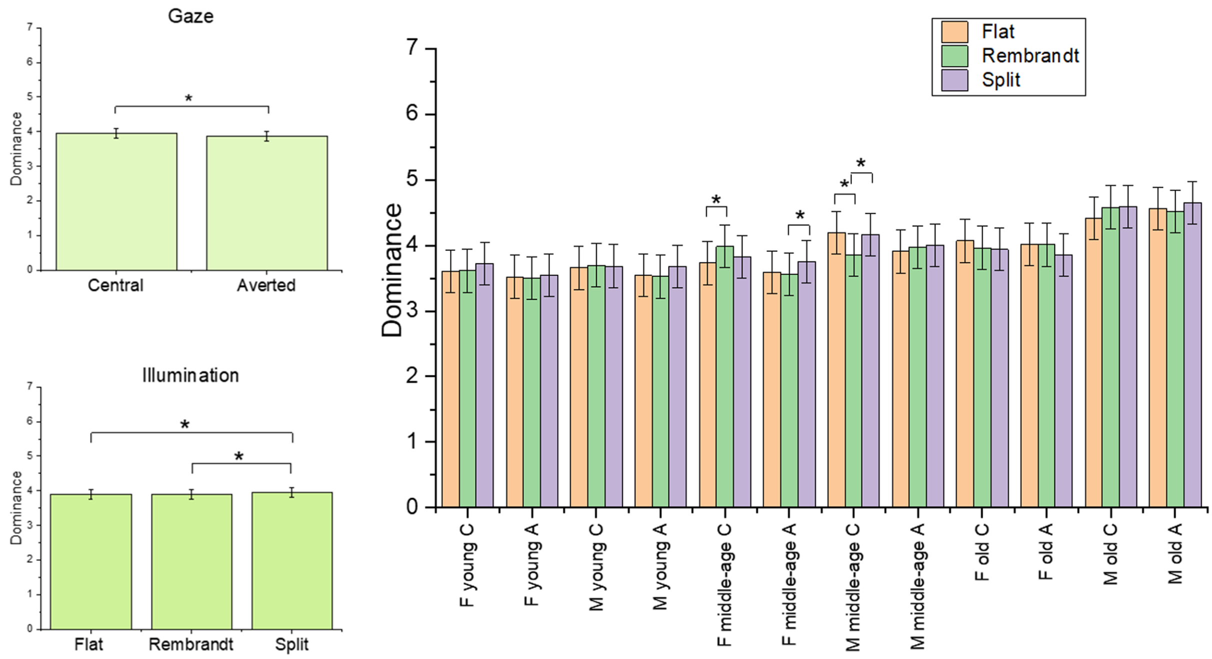


FIGURE 4
Plot of the estimated mean scores from the dominance rating, depicting significant main effects and higher order interaction. The asterisks highlight the significant main effect and significant results form *post hoc* analyses. C=central gaze, A=averted gaze, F=female models, M=male models. Error bars represent mean standard errors.

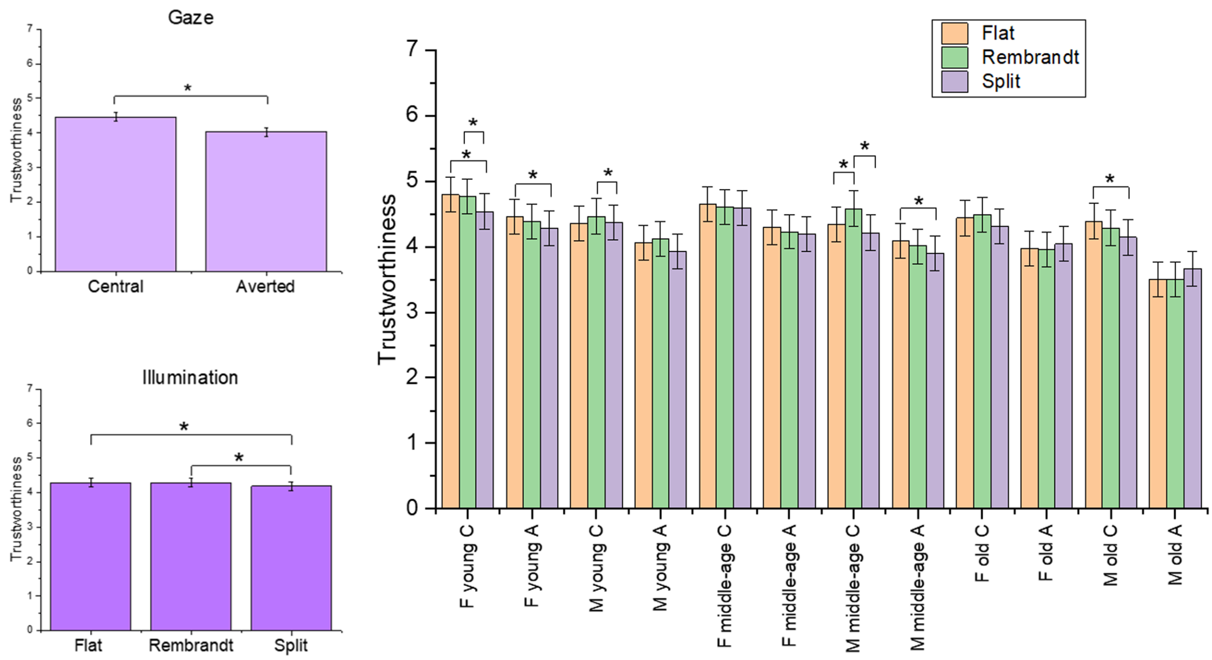


FIGURE 5
Plot of the estimated mean scores from the trustworthiness rating, depicting significant main effects and higher order interaction. The asterisks highlight the significant main effect and significant results form *post hoc* analyses. C=central gaze, A=averted gaze, F=female models, M=male models. Error bars represent mean standard errors.

responses showed that distribution to non-neutral responses was not totally random across the other emotions. This appeared in line with studies adopting a multidimensional perspective, which have shown that the identification of emotional expressions would take place within a dimensional space, rather than within discrete categories, and that neutral expression is located in a space closer to the expression of happiness (Shah and Lewis, 2003).

In line with previous literature, we found that gender, eye gaze, and age influenced face judgments. In particular, as expected, when considering gender and gaze variables, we found that females were rated as more attractive, feminine, trustworthy and less dominant than males (Perrett et al., 2002; Sutherland et al., 2013; Batres et al., 2015), whereas models with central eye gaze were generally judged more attractive, trustworthy and dominant than models with averted gaze (Main et al., 2009; Ewing et al., 2010; Kaisler and Leder, 2016; Mattavelli et al., 2021). Moreover, younger models scored higher on attractiveness (Thornhill and Gangestad, 1999), whereas judgments along the femininity-masculinity dimension increased with increasing age (Batres et al., 2015).

In addition, we investigated whether face judgments were affected by illumination condition, which was varied using three types of photography lighting. Flat, Rembrandt and split lighting conditions were obtained by moving the flash at three different angles from the camera (0°, 75° and 90°, respectively). We considered these settings as they are commonly used in portrait photography to emphasize different facial characteristics or effects conveyed by the pictures (Arena, 2012). Notably, previous studies explored the evaluation of Rembrandt's paints (from which the lighting features of Rembrandt type portrait were derived) in original or mirror-reversed position, highlighting significant impact of such particular position on cognitive and emotional processing of the stimuli (Schirillo, 2000, 2014; Powell and Schirillo, 2011). However, to the best of our knowledge, there are no previous studies investigating the impact of Rembrandt and split lighting, both consisting in lateralized illumination, on judgment of photographic portrait. Our results showed that these conditions affected evaluations on attractiveness, femininity-masculinity, dominance and trustworthiness of the stimuli, although with different impact in relation to age, gender and gaze of the faces. Split lighting has the ability to create a sharp effect on faces (Grey, 2004), indeed, it overall increased ratings of dominance, while decreased ratings of attractiveness and trustworthiness; furthermore, it increased masculinity in young female with averted gaze.

On the other hand, the Rembrandt illumination is often used in professional photography to give an intense and warm effect to faces (Grey, 2004). The present results support the modulatory role of this type of face illumination using systematic ratings in experimental setting. In particular, compared to split lighting, Rembrandt lighting increased the attractiveness, and trustworthiness rating in young and middle-aged models. Furthermore, this illumination appeared to have consistent effects for dominance and femininity-masculinity judgments related to gender of the stimuli: especially in middle-aged models it reduced the gender characterization of faces increasing dominance and masculinity of female faces and decreasing dominance and masculinity of male faces. This appeared in line with previous evidence showing that lateralized portrait of faces differently affects judgments of male and female models, which was exploited by painters (e.g., Rembrandt) to emphasize gender-related features (Schirillo, 2000). Indeed, males were more often portraited in a right-cheek position to increase social appealing, while females were mainly portraited in left-cheek position to give a demure appearance (Schirillo, 2000). Our models were all photographed with the right-side of the face illuminated and data add further evidence on the interaction between gender and lateralized face presentation. However, further investigation directly comparing right- and left-side illumination are required to support our findings.

These effects of illumination condition support previous evidence concerning within-person variability in face processing and social attribution (Jenkins et al., 2011; Todorov and Porter, 2014) and confirm that light is a critical feature affecting evaluation of ambient pictures in forming first impression (Hill and Bruce, 1996; Burton, 2013). Interestingly, in the analyses of the different ratings we found significant interactions among the factors of interest, suggesting that multiple facial features are considered together to create and evaluate face representations (Thornhill and Gangestad, 1999; Sutherland et al., 2013).

In particular, illumination differently affected attractiveness rating across gender and age: flat was rated more attractive than split condition with young and middle-age females, whereas middle-age males were more attractive with both flat and Rembrandt illumination compared to split; all faces were rated more attractive when posed with a central gaze, but with gender differences associated to age. It is also worth noting that, apart from the effect of split condition on trustworthiness of older males, *post hoc* tests revealed that illumination did not significantly affect rating scores in older adults, suggesting that evaluation of this age class was less influenced by contextual features.

In summary, the present study supports the evidence that face evaluation depends on multiple variables related to individual characteristics, changing aspects of faces and context, which interact to modulate social judgments (Burton, 2013). Our findings add novel data on the role of specific photographic portrait styles, and stimuli from the database could be used in further research exploring the impact of the different types of lateralized illumination on face evaluation. Indeed, the high-quality stimuli will be available for studies of face processing in a wide range of research fields from experimental psychology, perception, or face morphing to computational modeling studies for training face recognition algorithms (Adini et al., 1997; Georghiades et al., 2001). Moreover, the natural manipulation of gaze direction provides new stimuli ready to be used in joint attention paradigms, thus sparing the time-consuming editing procedure, and ensuring the ecological validity of stimuli depicting faces of different ages.

In conclusion, the Bi-AGI database offers the advantages of freely available face stimuli with manipulation of illumination in photographic settings and natural gaze directions, covering a wide age range across lifespan. The average ratings for each individual model are provided in the Appendix A, facilitating the selection of faces with particular characteristics and making Bi-AGI a feasible new tool for the scientific community.

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. Stimuli of the Bi-AGI database are available upon request at: https://board.unimib.it/datasets/rx6kpwmvtf/1.

Ethics statement

The studies involving human participants were reviewed and approved by Ethic Committee of the University Milano-Bicocca. 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 identifiable images or data included in this article.

Author contributions

GM and PR contributed to conception and design of the study. EC organized the database. IG and JD performed the statistical analysis. GM wrote the first draft of the manuscript. IG and JD wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The study was funded by a grant from University of Milano-Bicocca to PR. IG was partially supported by grants from Italian Ministry of Health to GM (GR-2016-02361283).

Acknowledgments

The authors thank Radice, Stifanelli, and Scordamaglia for their help in data collection.

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.948142/full#supplementary-material

References

Adams, R. B., and Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychol. Sci.* 14, 644–647. doi: 10.1046/j.0956-7976.2003.psci_1479.x

Adams, R. B., and Kleck, R. E. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion* 5, 3–11. doi: 10.1037/1528-3542.5.1.3

Adini, Y., Moses, Y., and Ullman, S. (1997). Face recognition: the problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. Mach. Intell.* 19, 721–732. doi: 10.1109/34.598229

Alghwiri, A. A., and Whitney, S. L. (2012). "Balance and falls" in *Geriatric Physical Therapy*. eds. A. Guccione, R. Wong and D. Avers. *3rd Edn.* (Mosby: St. Louis), 331–353. doi: 10.1016/B978-0-323-02948-3.00027-4

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4, 267–278. doi: 10.1016/S1364-6613(00)01501-1

Amesbury, E. C., and Schallhorn, S. C. (2003). Contrast sensitivity and limits of vision. *Int. Ophthalmol. Clin.* 43, 31–42. doi: 10.1097/00004397-200343020-00006

Arena, S. (2012). *Lighting for Digital Photography: from Snapshots to Great Shots (using Flash and Natural Light for Portrait, still Life, Action, and Product Photography)*, London: Pearson Education.

Artuso, C., Palladino, P., and Ricciardelli, P. (2012). How do we update faces? Effects of gaze direction and facial expressions on working memory updating. *Front. Psychol.* 3, 1–10. doi: 10.3389/fpsyg.2012.00362

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Bäckman, L. (1991). Recognition memory across the adult life span: the role of prior knowledge. *Mem. Cogn.* 19, 63–71. doi: 10.3758/BF03198496

Bar, M., Neta, M., and Linz, H. (2006). Very first impressions. *Emotion* 6, 269–278. doi: 10.1037/1528-3542.6.2.269

Batres, C., Re, D. E., and Perrett, D. I. (2015). Influence of perceived height, masculinity, and age on each other and on perceptions of dominance in male faces. *Perception* 44, 1293–1309. doi: 10.1177/0301006615596898

Bayliss, A. P., and Tipper, S. P. (2006). Predictive gaze cues and personality judgments: should eye trust you? *Psychol. Sci.* 17, 514–520. doi: 10.1111/j.1467-9280.2006.01737.x

Braje, W. L. (2003). Illumination encoding in face recognition: effect of position shift. *J. Vis.* 3, 161–170. doi: 10.1167/3.2.4

Braje, W. L., Kersten, D., Tarr, M. J., and Troje, N. F. (1998). Illumination effects in face recognition. *Psychobiology* 26, 371–380. doi: 10.3758/BF03330623

Burton, M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Q. J. Exp. Psychol.* 66, 1467–1485. doi: 10.1080/17470218.2013.800125

Ciardo, F., Marino, B. F. M., Actis-Grosso, R., Rossetti, A., and Ricciardelli, P. (2014). Face age modulates gaze following in young adults. *Sci. Rep.* 4:4746. doi: 10.1038/srep04746

De Rosario-Martinez, H. (2015). Phia: *post-hoc* interaction analysis. R Package Version 0.2-1

DiNardo, L., and Rainey, D. (1989). Recognizing faces in bright and dim light. *Percept. Mot. Skills* 68, 836–838. doi: 10.2466/pms.1989.68.3.836

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., and Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Vis. Cogn.* 6, 509–540. doi: 10.1080/135062899394920

Ebner, N. C. (2008). Age of face matters: age-group differences in ratings of young and old faces. *Behav. Res. Methods* 40, 130–136. doi: 10.3758/BRM.40.1.130

Ebner, N. C., Riediger, M., and Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: development and validation. *Behav. Res. Methods* 42, 351–362. doi: 10.3758/BRM.42.1.351

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* 24, 581–604. doi: 10.1016/S0149-7634(00)00025-7

Ewing, L., Rhodes, G., and Pellicano, E. (2010). Have you got the look? Gaze direction affects judgements of facial attractiveness. *Vis. Cogn.* 18, 321–330. doi: 10.1080/13506280902965599

Ferwerda, J. A. (1998). Fundamentals of spatial vision. *Applic. Visual Percept. Comput. Graphics* 140, 1–27.

Fiske, S. T., Cuddy, A. J. C., and Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends Cogn. Sci.* 11, 77–83. doi: 10.1016/j.tics.2006.11.005

Frischen, A., Bayliss, A. P., and Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition and individual differences. *Psychol. Bull.* 133, 694–724. doi: 10.1037/0033-2909.133.4.694

Frischen, A., and Tipper, S. P. (2006). Long-term gaze cueing effects: evidence for retrieval of prior attentional states from memory. *Vis. Cogn.* 14, 351–364. doi: 10.1080/13506280544000192

Frischen, A., and Tipper, S. P. (2004). Orienting attention via observed gaze shift evokes longer term inhibitory effects: implications for social interactions, attention, and memory. *J. Exp. Psychol. Gen.* 133, 516–533. doi: 10.1037/0096-3445.133.4.516

Gao, W., Cao, B., Shan, S., Zhou, D., Zhang, X., and Zhao, D. (2004). The CAS-PEAL large-scale Chinese face database and evaluation protocols. Technical report no. JDL_TR_04_FR_001, Joint Research and Development Laboratory. CAS

Gelman, A., and Hill, J. (2006). *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge: University Press

Georghiades, A. S., Belhumeur, P. N., and Kriegman, D. J. (2001). From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Machine Intellig.* 23, 643–660. doi: 10.1109/34.927464

Grey, C. (2004). *Master Lighting Guide for Portrait Photographers* Amherst Amherst Media, Inc.

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

Hill, H., and Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 986–1004. doi: 10.1037/0096-1523.22.4.986

Hiraoka, T., Hoshi, S., Okamoto, Y., Okamoto, F., and Oshika, T. (2015). Mesopic functional visual acuity in Normal subjects. *PLoS One* 10:134505. doi: 10.1371/journal.pone.0134505

Ishai, A. (2008). Let's face it: it's a cortical network. *NeuroImage* 40, 415–419. doi: 10.1016/j.neuroimage.2007.10.040

Ito, A., Yoshida, K., Aoki, R., Fujii, T., Kawasaki, I., Hayashi, A., et al. (2022). The role of the ventromedial prefrontal cortex in preferential decisions for own- and other-age faces. *Front. Psychol.* 13, 1–13. doi: 10.3389/fpsyg.2022.822234

Jenkins, R., White, D., Van Montfort, X., and Burton, A. (2011). Variability in photos of the same face. *Cognition* 121, 313–323. doi: 10.1016/j.cognition.2011.08.001

Jin, X., Zhao, M., Chen, X., Zhao, Q., and Zhu, S. C. (2010). "Learning artistic lighting template from portrait photographs" in *Computer Vision – ECCV 2010. ECCV 2010. Lecture Notes in Computer Science*. eds. K. Daniilidis, P. Maragos and N. Paragios, *Vol. 6314* (Berlin, Heidelberg: Springer).

Kaisler, R. E., and Leder, H. (2016). Trusting the looks of others: gaze effects of faces in social settings. *Perception* 45, 875–892. doi: 10.1177/0301006616643678

Kelber, A., Yovanovich, C., and Olsson, P. (2017). Thresholds and noise limitations of colour vision in dim light. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160065. doi: 10.1098/rstb.2016.0065

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and validation of the Radboud faces database. *Cognit. Emot.* 24, 1377–1388. doi: 10.1080/02699930903485076

Lim, D., Lee, A., and Or, C. (2022). Incongruence in lighting impairs face identification. *Front. Psychol.* 13:834806. doi: 10.3389/fpsyg.2022.834806

Lundqvist, D., Flykt, A., and Öhman, A. (1998). Karolinska directed emotional faces. *Cognit. Emot.*

Main, J. C., Jones, B. C., DeBruine, L. M., and Little, A. C. (2009). Integrating gaze direction and sexual dimorphism of face shape when perceiving the dominance of others. *Perception* 38, 1275–1283. doi: 10.1068/p6347

Manssuer, L. R., Roberts, M. V., and Tipper, S. P. (2015a). The late positive potential indexes a role for emotion during learning of trust from eye-gaze cues. *Soc. Neurosci.* 10, 635–650. doi: 10.1080/17470919.2015.1017114

Manssuer, L. R., Pawling, R., Hayes, A. E., and Tipper, S. P. (2015b). The role of emotion in learning trustworthiness from eye-gaze: evidence from facial electromyography. *Soc. Neurosci.* 7, 82–102. doi: 10.1080/17588928.2015.1085374

Martinez, A. M., and Benavente, R. (1998). The AR face database. CVC Technical Report, 24.

Mason, M. F., Tatkow, E. P., and Macrae, C. N. (2005). The look of love. Gaze shifts and person perception. *Psychol. Sci.* 16, 236–239. doi: 10.1111/j.0956-7976.2005.00809.x

Mattavelli, G., Romano, D., Young, A. W., and Ricciardelli, P. (2021). The interplay between gaze cueing and facial trait impressions. *Q. J. Exp. Psychol.* 74, 1642–1655. doi: 10.1177/17470218211007791

Milders, M., Hietanen, J. K., Leppänen, J. M., and Braun, M. (2011). Detection of emotional faces is modulated by the direction of eye gaze. *Emotion* 11, 1456–1461. doi: 10.1037/a0022901

Minear, M., and Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behav. Res. Methods Instrum. Comput.* 36, 630–633. doi: 10.3758/BF03206543

Nyman, T. J., Antfolk, J., Lampinen, J. M., Tuomisto, M., Kaakinen, J. K., Korkman, J., et al. (2019). A stab in the dark: the distance threshold of target identification in low light. *Cogent Psychol.* 6:2047. doi: 10.1080/23311908.2019.1632047

O'Reilly, H., Pigat, D., Fridenson, S., Berggren, S., Tal, S., Golan, O., et al. (2016). The EU-emotion stimulus set: a validation study. *Behav. Res. Methods* 48, 567–576. doi: 10.3758/s13428-015-0601-4

Perrett, D., Penton-Voak, I. S., Little, A. C., Tiddeman, B. P., Burt, D. M., Schmidt, N., et al. (2002). Facial attractiveness judgements reflect learning of parental age characteristics. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 269, 873–880. doi: 10.1098/rspb.2002.1971

Powell, W. R., and Schirillo, J. A. (2011). Hemispheric laterality measured in rembrandt's portraits using pupil diameter and aesthetic verbal judgements. *Cognit. Emot.* 25, 868–885. doi: 10.1080/02699931.2010.515709

R Development Core Team (2008). Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Ricciardelli, P., Iani, C., Lugli, L., Pellicano, A., and Nicoletti, R. (2012). Gaze direction and facial expressions exert combined but different effects on attentional resources. *Cognit. Emot.* 26, 1134–1142. doi: 10.1080/02699931.2011.638907

Sander, D., Grandjean, D., Kaiser, S., Wehrle, T., and Scherer, K. R. (2007). Interaction effects of perceived gaze direction and dynamic facial expression: evidence for appraisal theories of emotion. *Eur. J. Cogn. Psychol.* 19, 470–480. doi: 10.1080/09541440600757426

Savran, N., Alyüz, H., Dibeklioglu, O. C, Eliktutan, B., Gökberk, B., Sankur, B., et al. (2008). Bosphorus database for 3D face analysis. The First COST 2101 Workshop on Biometrics and Identity Management (BIOID 2008). Roskilde University, Denmark.

Schirillo, J. A. (2014). Pupil dilations reflect why Rembrandt biased female portraits leftward and males rightward. *Front. Hum. Neurosci.* 7, 1–11. doi: 10.3389/fnhum.2013.00938

Schirillo, J. A. (2000). Hemispheric asymmetries and gender influence Rembrandt's portrait orientations. *Neuropsychologia* 38, 1593–1606. doi: 10.1016/S0028-3932(00)00063-4

Shah, R., and Lewis, M. B. (2003). Locating the neutral expression in the facial-emotion space. *Vis. Cogn.* 10, 513–526. doi: 10.1080/13506280244000203

Sheedy, J. E., Bailey, I. L., and Raasch, T. W. (1984). Visual acuity and chart luminance. *Am. J. Opto. Vision Sci.* 61, 595–600. doi: 10.1097/00006324-198409000-00010

Shin, J. C., Yaguchi, H., and Shioiri, S. (2004). Change of color appearance in Photopic, Mesopic and scotopic vision. *Opt. Rev.* 11, 265–271. doi: 10.1007/s10043-004-0265-2

Shu, Z., Hadap, S., Shechtman, E., Sunkavalli, K., Paris, S., and Samaras, D. (2017). Portrait lighting transfer using a mass transport approach. *ACM Trans. Graph.* 37, 1–15. doi: 10.1145/3095816

Sim, T., Baker, S., and Bsat, M. (2003). The CMU pose, illumination, and expression database. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 1615–1618. doi: 10.1109/AFGR.2002.1004130

Sutherland, C. M., Oldmeadow, J., Santos, I. M., Towler, J., Michael, B. D., and Young, A. W. (2013). Social inferences from faces: ambient images generate a three-dimensional model. *Cognition* 127, 105–118. doi: 10.1016/j.cognition.2012.12.001

Thornhill, R., and Gangestad, S. W. (1999). Facial attractiveness. *Trends Cogn. Sci.* 3, 452–460. doi: 10.1016/S1364-6613(99)01403-5

Todorov, A., and Porter, J. M. (2014). Misleading first impressions: different for different facial images of the same person. *Psychol. Sci.* 25, 1404–1417. doi: 10.1177/0956797614532474

Van Der Schalk, J., Hawk, S. T., Fischer, A. H., and Doosje, B. (2011). Moving faces, looking places: validation of the Amsterdam dynamic facial expression set (ADFES). *Emotion* 11, 907–920. doi: 10.1037/a0023853

Weidenbacher, U., Layher, G., Strauss, P.-M., and Neumann, H. (2007). A comprehensive head pose and gaze database in 3rd IET International Conference on Intelligent Environments, Ulm (Germany), 455–458.

Willis, J., and Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychol. Sci.* 17, 592–598. doi: 10.1111/j.1467-9280.2006.01750.x

Wood, J. M. (2020). Nighttime driving: visual, lighting and visibility challenges. *Ophthalmic Physiol. Opt.* 40, 187–201. doi: 10.1111/opo.12659

Young, A. W., Perrett, D. I., Calder, A. J., Sprengelmeyer, R., and Ekman, P. (2002). *Facial Expressions of Emotion: Stimuli and Test (FEEST)*. Bury St. Edmunds: Thames Valley test Company.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 03 November 2022
DOI 10.3389/fpsyg.2022.979341

Check for updates

OPEN ACCESS

EDITED BY
Ayahito Ito,
Kochi University of Technology, Japan

REVIEWED BY
Ryuhei Ueda,
Kyoto University, Japan
Yousuke Kawachi,
Tohoku University, Japan

*CORRESPONDENCE
Masashi Komori
komori@osakac.ac.jp

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

RECEIVED 15 July 2022
ACCEPTED 17 October 2022
PUBLISHED 03 November 2022

CITATION
Komori M, Teraji T, Shiroshita K and Nittono H (2022) Examination of morphological traits of children's faces related to perceptions of cuteness using Gaussian process ordinal regression. *Front. Psychol.* 13:979341. doi: 10.3389/fpsyg.2022.979341

Examination of morphological traits of children's faces related to perceptions of cuteness using Gaussian process ordinal regression

Masashi Komori[1]*, Teppei Teraji[1], Keito Shiroshita[1] and Hiroshi Nittono[2]

[1]Graduate School of Engineering, Osaka Electro-Communication University, Neyagawa, Japan,
[2]Graduate School of Human Sciences, Osaka University, Suita, Japan

Konrad Lorenz, an ethologist, proposed that certain physical elements are perceived as cute and induce caretaking behavior in other individuals, with the evolutionary function of enhancing offspring survival. He called these features Kindchenschema, baby schema. According to his introspection, these include a large forehead, chubby round features, and chubby cheeks. Previous studies are limited to examining the effects of these facial features on perceived cuteness. However, other morphological factors may be related to perceived cuteness. This study uses Bayesian optimization, one of the global sequential optimization methods for estimating unknown functions, to search for facial morphological features that enhance the perceptions of facial cuteness. We applied Bayesian optimization incorporating Gaussian process ordinal regression (GPOR), which allows an estimation of the latent cuteness function based on evaluations using the Likert scale. A total of 96 preschool children provided the facial images used in this study. We summarized the facial shape variations using methodologies of geometric morphometrics and principal component analysis (PCA) up to the third principal component (PC), which we refer to as the face space. A total of 40 participants evaluated the images created by warping the average facial texture of the children's faces with randomly generated parameters in the face space. Facial traits related to perceived cuteness were estimated based on the averaged cuteness function. Perceived cuteness was linked to the relative lower position of facial components and narrower jawline but not to the forehead height.

KEYWORDS

Gaussian process ordinal regression, Bayesian optimization, face, baby schema, cuteness

10.3389/fpsyg.2022.979341

1. Introduction

Ethologist Konrad Lorenz pointed out that certain physical features of animals induce specific behaviors and affective responses related to caretaking behaviors (Lorenz, 1943). Further, he called such universal features across species the baby schema ("Kindchenschema"). Lorenz argued that baby schema includes features such as a large head, big eyes, high and protruding forehead, chubby cheeks, small nose, and mouth, that are perceived as cute. Many empirical studies have supported this idea, suggesting a biological and evolutionary basis for cognitive and emotional responses to specific features typical of children's faces (Kringelbach et al., 2016). For example, some studies have shown that infants' level of facial cuteness is associated with a large forehead, small chin, full lips, and chubby round features (Hildebrandt and Fitzgerald, 1979; Almanza-Sepúlveda et al., 2018) suggesting that faces that are perceived as cute have some traits in common.

However, facial traits that elicits the perception of cuteness are not identical to the facial features possessed by human infants. This is because the emphasis on infant facial traits does not necessarily increase the levels of perceived cuteness of children's faces (Komori and Nittono, 2013). Therefore, the factors that induce the perception of cuteness cannot be clarified only by examining the objective characteristics of human infants. Thus far, different approaches have been used to investigate which facial features affect the perception of cuteness. One approach is to take specific facial elements, such as roundness of the face, the height of forehead and eyes, nose and mouth size, and manipulate the images directly to examine their effects on perceived cuteness (Sternglanz et al., 1977; Alley, 1981; Glocker et al., 2009; Borgi et al., 2014; Endendijk et al., 2018; Löwenbrück and Hess, 2021). However, there is a limit to this method of determining *a priori* which features to manipulate since there may be unknown factors that increase the levels of cuteness. Another approach is to create facial images having features that enhance cuteness by computationally compositing images of faces with high cuteness (Sprengelmeyer et al., 2009; Lobmaier et al., 2010; Hahn et al., 2013; Nittono et al., 2022). However, such a method cannot clarify which features are involved in cuteness perception and the degree to which they are involved. The other approach is analyzing the correlation between the locations of facial landmarks and perceived cuteness. Almanza-Sepúlveda et al. (2018) examined the relationship between infants' facial traits and cuteness based on a linear regression analysis. However, it is possible that the relationship between facial traits and cuteness is non-linear, similar to that between facial traits and attractiveness (Komori et al., 2009b). Thus, this study aims to solve these problems of previous methodologies by using a method that can comprehensively examine the non-linear relationship between morphological facial traits and perceived cuteness.

We assume that people have their own psychological function $f(\mathbf{x})$ (the present study refers to this function as the utility function) that maps the multivariate input $\mathbf{x}$ (i.e., certain traits of a presented face) to a scalar value representing the degree of perceived cuteness of the person whose face has been assessed. Thus, the problem of elucidating the impression evaluation mechanism for others' faces can be regarded as the problem of estimating the parameters of the utility function $f(\mathbf{x})$.

The input $\mathbf{x}$ of the utility function corresponds to the morphological features of the faces. Instead of arbitrarily targeting specific facial features for investigation, such as forehead size and eye position, this study uses a methodology of geometric morphometrics (Bookstein, 1991; Dryden and Mardia, 1998) to represent the features of infants' facial shapes in a few dimensions. Geometric morphometrics is a technique used to represent facial shape as multivariate data (Komori et al., 2009b).

Even if facial features are represented in a small number of dimensions, the number of combinations is so large that it is not easy to reveal the utility function f. One way to thoroughly examine a psychological utility function $f(\mathbf{x})$ is to explore the search space via a grid search. However, it is challenging to obtain responses to multidimensional stimuli by performing a grid search because of the high cost of the evaluation task. Moreover, since the utility function of cuteness perception may be complicated, it is inappropriate to apply methods, such as conjoint analysis, which does not consider feature combinations, to study the cuteness perception. Since the utility function of perceived cuteness may be multi-peaked or have a complex shape, it is also inappropriate to apply methods such as conjoint analysis (Rao et al., 2014), which does not take into account feature combinations, to the study of cuteness perception.

This study aimed to estimate the relationship between child facial features and perceived cuteness using the Gaussian process regression (GPR) model (O'Hagan, 1978; Neal, 1997), a non-parametric Bayesian approach to metric regression. Gaussian process regression is a method for estimating an unknown black-box function using a kernel function and is often used to find the maximum/minimum value of the function. Unlike linear regression analysis, GPR has the advantage of being able to estimate non-linear functions, such as multi-peaked functions, and also provide probabilistic predictions.

However, there are problems in using GPR to estimate psychological utility functions. In a typical GPR method, the return value (i.e., the responses from the participants) from an unknown function is expected to be a continuous quantity. Conversely, psychological research commonly uses methods that require discrete responses from participants, such as Likert scales, dichotomous choices, and paired comparison methods. Therefore, using discrete responses to estimate psychophysical functions based on natural human responses is desirable.

We used the Gaussian process ordinal regression (GPOR) (Chu et al., 2005), which can estimate non-linear function f from discrete responses (i.e., responses to a Likert scale).

Let $\mathcal{D} = (\mathbf{x}_i, y_i)|i = 1, \ldots, N = (\mathbf{X}, \mathbf{y})$ denote the set of facial traits $\mathbf{x}_i$ and the corresponding cuteness evaluation label $y_i \in L$, where L is a finite set of R ordinal categories, denoted $L = \{1, 2, \ldots R\}$. We assume that the utility function f follows a zero mean Gaussian process prior defined by a kernel function $\mathcal{K}$.

$$f \sim \mathcal{GP}(0, \mathcal{K}) \tag{1}$$

Here, we used an RBF ARD kernel (MacKay, 1996).

$$\mathcal{K}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \exp\left(-\frac{1}{2}\sum_{d=1}^{\mathbf{D}} \eta_d \left(x_{di} - x_{dj}\right)^2\right) \tag{2}$$

Let $\mathbf{b}$ be the threshold variable, where $b_0 < b_1 < \cdots < b_r$ and $b_0 = -\infty, b_r = \infty$, which map $f(x_i)$ to the discrete variable y_i. Under noise-free conditions, the ideal likelihood function would be defined as $\mathcal{P}_{ideal}(y_i|f(\mathbf{x}_i)) = 1$ when $b_{y_i - 1} < f(\mathbf{x}_i) < b_{y_i}$ (0 otherwise). In the presence of noise from inputs or targets, the latent functions are contaminated by Gaussian noise with zero mean and unknown variance, denoted $\mathcal{N}(\delta; 0, \sigma^2)$, where δ is a Gaussian random variable. The ordinal likelihood function becomes

$$\mathcal{P}\left(y_i \mid f\left(x_i\right)\right) = \Phi\left(z_i^{(y_i)}\right) - \Phi\left(z_i^{(y_i - 1)}\right) \tag{3}$$

$$z_i^{(s)} = \frac{b_s - f\left(x_i\right)}{\sigma_\delta} \tag{4}$$

where $\Phi(z)$ is the cumulative unit Gaussian (CDF) whereby $\Phi(z) = \int_{-\infty}^{z} \mathcal{N}(\gamma; 0, 1)d\gamma$. Based on Bayes' theorem, the posterior probability can then be written as

$$\mathcal{P}(f \mid \mathcal{D}) = \frac{1}{\mathcal{P}(\mathcal{D})} \prod_{(i=1)}^{n} \mathcal{P}\left(y_i \mid f\left(x_i\right)\right) \mathcal{P}(f) \tag{5}$$

To approximate the posterior distribution and model evidence $\mathcal{P}(\mathcal{D})$ we used the Laplace approximation at the maximum a posteriori (MAP) estimate (Williams and Barber, 1998). Under the Laplace approximation, the predictive distribution of the utility function can be described as a Gaussian $\mathcal{N}(f(x); \mu, \sigma^2)$ where the predictive mean ($\mu_{\mathbf{x}*}$) and variance ($\sigma^2_{\mathbf{x}*}$) for which the response y_* is unknown.

$$\mu_{\mathbf{x}}* = \mathbf{k}*^T \mathbf{K}^{-1} f_{MAP} \tag{6}$$

$$\sigma^2_{\mathbf{x}*} = \mathcal{K}(\mathbf{X}, \mathbf{X}) - \mathbf{k}_*^T \left(\mathbf{K} + \Lambda_{\mathbf{MAP}}^{-1}\right)^{-1} \mathbf{k}_* \tag{7}$$

where; $\mathbf{k}_*$ is the covariance between the test case and the training data. Further, $f_{\mathbf{MAP}}$ is the MAP estimate of the utility function, $\Lambda_{\mathbf{MAP}}$ is a diagonal matrix whose i-th entry is the second derivative of the likelihood function training sample i concerning $f(\mathbf{x}_i)$.

This study examines the effectiveness of GPOR as a novel method for revealing the utility function f, a black box function that describes the relationship between multidimensional facial features and perceived cuteness, from two perspectives.

First, we compare the predictive performance of the GPOR model used in this study with an ordinal logistic regression model, a standard generalized linear regression model for ordinal data, using the leave-one-out cross-validation (LOOCV) approach.

Second, we average the utility functions obtained from the participants in the experiment and examine their shape. Furthermore, by identifying the face shapes corresponding to the maximum and minimum values of the average utility function, we explore the factors that determine the perceived facial cuteness of a child. Finally, these results will be discussed based on whether GPOR can provide useful information in psychological studies.

2. Computational analysis of facial images

2.1. Materials and facial shape measurement

Japanese preschool children (n = 96, 48 boys and 48 girls; age 3–4 years, mean age = 3.98, SD = 0.55) provided the facial frontal images used in this study. Written informed consents were obtained from the legal guardians for the publication of non-personally identifiable images or data included in this article. A neutral expression on each face was captured using a digital camera. Additionally, the foreheads of the models were exposed using a headband, after removing head accessories such as eyeglasses. These images are the same as those used in the authors' previous study (Komori and Nittono, 2013).

Eighty facial landmarks were selected based on a previous study (Komori et al., 2009a). These landmarks consisted of morphological and functional points, such as the pupils, contours of eyes, eyebrows, nose, and mouth, among others. The authors visually measured all landmarks for each of the 96 photographs using a program written by the authors.

2.2. Facial shape standardization

Each face differed in location, size, and orientation. To standardize them, we performed a generalized Procrustes analysis (GPA) on the facial landmarks of all faces. This

FIGURE 1
Average facial image.

method preserves information about the relative spatial relationships of landmarks throughout the standardization. For the standardization of location and size, we used the centroid size technique (Bookstein, 1991). All facial shapes were translated for that the centroid (center of gravity) to have the exact location scaled to the same centroid size, which is the sum of the squared distances from the centroid to each landmark. For alignment of the orientation, rotations around the centroid of the faces (Dryden and Mardia, 1998) were performed to minimize the sum of the squared distances among corresponding feature points between samples. The "Shapes" statistical package, which runs in an R statistical analysis environment, was employed for the analyses.

Facial differences derived from facial asymmetries are not the focus of this study. Therefore, to exclude facial variations derived from facial asymmetries and the coordinates of original facial images, the mirror-reversed versions of the same images were used to create a "symmetrical version" of the individuals following the procedures of a previous study (Komori et al., 2009a). Consequently, each "symmetrical version" of the individuals was represented as a point on a linear space of 160 dimensions (x- and y-coordinates × 80 landmarks). This study refers to this multidimensional space as "face space." The center of this face space corresponds to the average face of the "symmetrical version" faces.

All face images were gray-scaled and warped to the average face landmark coordinates using thin-plate splines transformation (TPS), a non-linear image deformation technique. Then, the average face texture was synthesized by averaging the luminance values of the warped images (Figure 1).

2.3. Facial feature extraction

Variations in children's facial features were summarized using a principal component analysis (PCA). The results of the PCA indicated that the contributions of the first three principal components (PCs) for the total variance were relatively large (PC1: 31.6%; PC2: 19.2%; PC3: 13.7%; cumulative contribution ratio: 64.5%), and thus up to the PC3 was considered in the study. Based on the distributions of children's facial shapes, the facial landmark coordinates changes along each PC were calculated (Figure 2). The PC1 was related to lower chin position and lower hairline, indicating that PC1 is related to relative eye and mouth position. Further, the higher the score, the higher the relative eye and mouth position. The PC2 was associated with the face width face from the cheekbones to the jaw. PC3 was linked to the height or the length of the forehead.

The facial images corresponding to the theoretical values of −2SD, −SD, average, +SD, and +2SD along each PC were made by warping the average facial texture of the children's faces using TPS (Figure 3).

3. Assessment of perceived cuteness

3.1. Participant

Forty undergraduates (20 men and 20 women; mean age = 22.13, SD = 2.20) participated in the assessment of facial cuteness.

3.2. Procedure

The images presented to the participants are the warped images of the average facial image (Figure 3) using TPS so that the facial shapes correspond to given PC scores in the face space. Participants were instructed to evaluate the levels of cuteness of face images. Further, they were rated on a 5-point scale where 1 = not cute (*kawaii* in Japanese) at all and 5 = very cute ((*kawaii*) according to their first impressions of the images, i.e., without contemplating their responses. No time limit was set for the responses in each trial. Facial images were presented on an LCD monitor. The participants responded with a keyboard.

The rating task consists of 20 practice trials and 60 test trials. In the first 45 test trials, the stimulus images were generated from the PC scores randomly selected from a range of ±2SD in the face space. In the subsequent 15 trials, the images were generated from the PC scores at which the upper confidence bound (UCB) values are maximized based on the response history of the participants in the previous test trials where N is the number of trials:

$$\mathrm{UCB}_{\mathbf{x}} = \mu_{\mathbf{x}} + \sqrt{\frac{\log N}{N}}\sigma_{\mathbf{x}}. \tag{8}$$

Upper confidence bound is a type of acquisition function often used in Bayesian optimization. This type of experimental

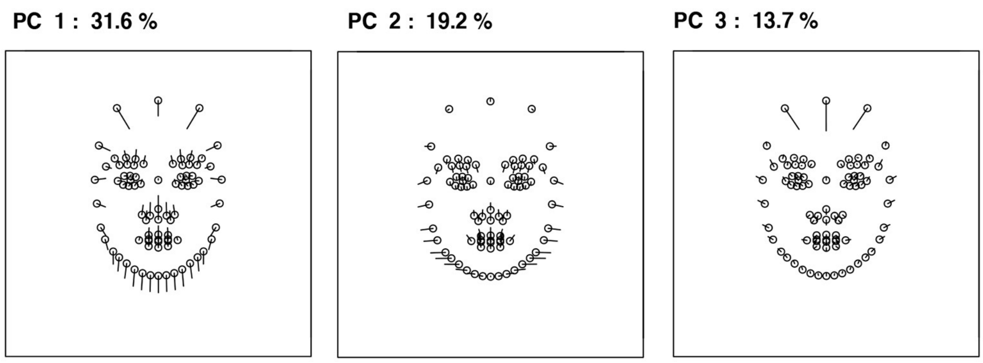


FIGURE 2
Facial shape changes along each PC. When the score of each PC increases, each facial landmark (depicted as a circle) moves along the leading line.

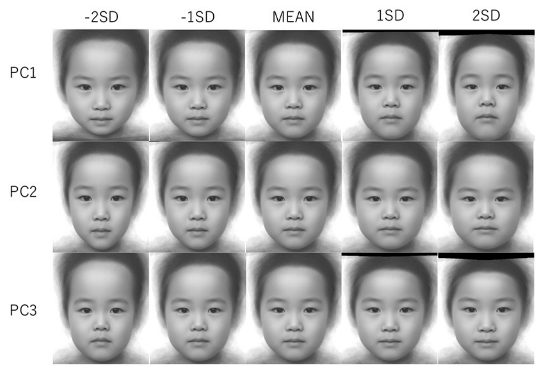


FIGURE 3
Differences in face shape along the each principal component (−2 SD/−SD/MEAN/+SD/+2 SD). The aspect ratios of the images were different from each other due to TPS warping.

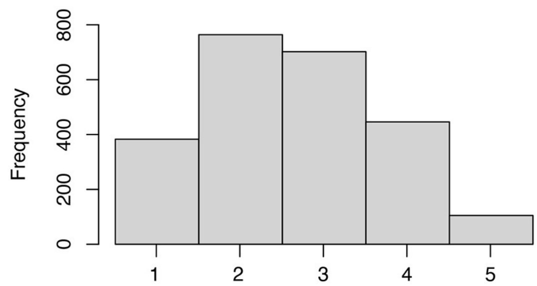


FIGURE 4
Distribution of ratings.

procedure is called a sequential experimental design. The trials were separated by 5 s of a blank gray screen. Each participant completed the rating task in about 30 min. The application used in the experiments was implemented in a PsychoPy environment (Peirce, 2007).

4. Results

4.1. Leave-one-out error analysis

The predictive performance of GPOR was examined using leave-one-out cross-validation (LOOCV). Specifically, we withheld one response from each participant. Further, we fit the models of GPOR and ordinal logistic regression on the remaining responses of the participant and formed a prediction of the held-out response using each of the learned models. We then compared each learned model's predictions with the observed responses for the held-out trial and computed the zero-one error and the squared error. Zero-one error gives an error of 1 to every incorrect prediction. Absolute error is the deviation of the prediction from the actual target. By computing the mean of the held-out zero-one errors (MZE) and the mean of absolute errors (MAE) across all participants, we can evaluate the predictive performance of the models.

- Mean zero-one error (MZE): the fraction of incorrect predictions on test data; $\frac{1}{N}\sum_{i=1}^{N} I\left(\hat{y}_i \neq y_i\right)$, where $I(\cdot)$ denotes an indicator function which gives 1 when the argument is true and 0 otherwise.
- Mean absolute error (MAE): the average deviation of predicted test outputs from the true rank, in which we treat the ordinal scales as consecutive integers, $\frac{1}{N}\sum_{i=1}^{N} \left|\hat{y}_i - y_i\right|$.

The distribution of ratings is shown in Figure 4. The MZE and mean MAE of the predictions by the GPOR model and by ordinal logistic regression were obtained by LOOCV (Table 1).

TABLE 1 Cross-validation results.

	MZE		MAE	
	Test	Train	Test	Train
Ordinal logistic regression	0.723	0.664	1.122	0.942
GPOR	0.655	0.455	0.82	0.525

Mean zero-one error refers to the fraction of classes predicted that differ from the correct response, which is the fraction of incorrect predictions. This study's chance level of MZE is 0.8 because the participants were given a 5-point scale. The MZEs of both models had higher prediction accuracies than chance level [ordinal logistic regression: $t_{(39)} = -4.71$, $p < 0.001$, Cohen's $d = 0.74$; GPOR: $t_{(39)} = -6.71$, $p < 0.001$, Cohen's $d = 1.06$]. The GPOR model had a lower error rate than the ordinal logistic regression model [$t_{(39)} = -2.81$, $p = 0.007$, Cohen's $d = 0.44$]. Mean of absolute error is the average deviation of the prediction from the actual target values in which ordinal scales are treated as consecutive integers, and, as with MZE, GPOR performed better than ordinal logistic regression [$t_{(39)} = -5.87$, $p < 0.001$, Cohen's $d = 0.93$]. These results suggest that GPOR is more effective than the generalized linear model in estimating the utility function that represents the relationship between facial features and perceived cuteness.

4.2. Facial shape that maximizes and minimizes perceived cuteness

The utility function f_s of each participant s is expressed as a set of predicted means estimated by using GPOR. The predicted means and the variances were obtained within ± 2 *SD*. of each dimension of the PCs with an interval of 0.4, resulting in 1,331 points per participant. The maximum value of utility function $f_s(\mathbf{x}_s^*)$ for each participant was shown to be significantly greater than zero [$p(f_s(\mathbf{x}_s^*) \leq 0) < 0.01$ for each participant], where $\mathbf{x}_s^* = \arg\max_{\mathbf{x} \in A} f_s(\mathbf{x})$ in face space A, indicating that differences in face shape had a consistent influence on individual judgments.

Next, we calculated the mean of the predicted mean $\mu_{S\mathbf{x}}$ and the predicted variance $\sigma^2_{S\mathbf{x}}$ for each coordinate $\mathbf{x}$. M denotes the number of participants in this study.

$$\mu_{S\mathbf{x}} = \frac{1}{M} \sum_{s=1}^{M} \mu_{s\mathbf{x}} \tag{9}$$

$$\sigma^2_{S\mathbf{x}} = \frac{1}{M} \sum_{s=1}^{M} \left(\mu^2_{s\mathbf{x}} + \sigma^2_{s\mathbf{x}} \right) - \mu^2_{S\mathbf{x}} \tag{10}$$

Here, $\mu_{S\mathbf{x}}$ is an estimate of the average perceived cuteness for a given facial feature $\mathbf{x}$, and we refer to the set of $\mu_{S\mathbf{x}}$ as the average utility function f_S. Figure 5 the relationship between the first, second, and third PCs of facial traits and the average perceived cuteness. The perceived cuteness is higher for the combination of slightly lower the first and the second PC scores. However, the high scores of the third PC are consistently associated with perceived cuteness.

$\mathbf{x}^*$ denotes the coordinate of a maximum of averaged cuteness in the face space A.

$$\mathbf{x}^* = \arg\max_{\mathbf{x} \in A} \mu_{S\mathbf{x}} \tag{11}$$

The maximum predicted mean $\mu_{S\mathbf{x}^*}$ of the function f_S was significantly greater than zero [$p(f_S(\mathbf{x}^*) \leq 0) < .05$], according to the combined predicted variance $\sigma^2_{S\mathbf{x}^*}$. This suggests that averaging utility functions of the participants is valid in examining the average tendency of judgments.

The coordinates of the maximum value $\mathbf{x}^*$ of the average utility function were obtained (PC1: −0.8, PC2: −0.4, PC3: 0.8). Further, the coordinates of minimal value $\mathbf{x}^-$ of the average utility function were obtained (PC1: 0.0, PC2: 0.4, PC3: −1.6).

$$\mathbf{x}^- = \arg\min_{\mathbf{x} \in A} \mu_{S\mathbf{x}} \tag{12}$$

Figure 6 shows the face images corresponding to the maximum and minimum values of the average utility function.

The face perceived to be the cutest, on average, is characterized by low eye position, smaller chin, narrow jawline and forehead, according to the first PC related to relative eye position, the second to jawline width, and the third to forehead height. Conversely, the face with the minimum utility function for perceived cuteness is characterized by high eye position, wide jawline, and high forehead.

5. Discussion

The present study used GPOR (Chu et al., 2005), a non-linear regression analysis, to estimate utility functions that describe the relationship between multivariate children's facial shape traits and the degree of perceived cuteness. We measured children's face shapes and constructed a face space through feature dimensionality reduction using a combination of methodologies of geometric morphometrics (Bookstein, 1991; Dryden and Mardia, 1998) and a PCA. Participants were instructed to rate the cuteness of the facial images synthesized within the face space on a 5-point scale. Furthermore, the average utility function was calculated,

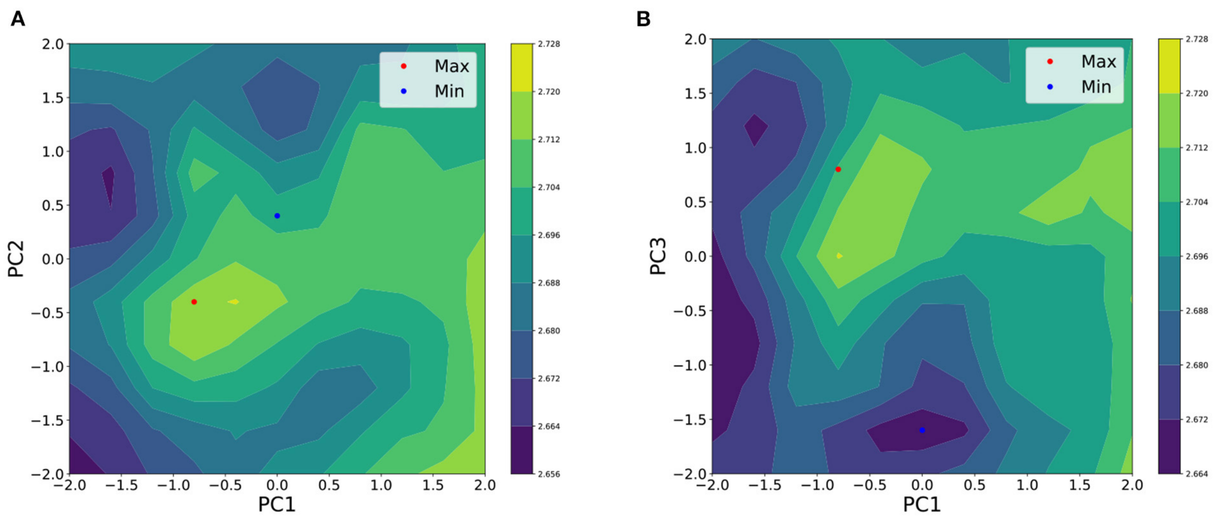


FIGURE 5
Averaged utility function of perceived cuteness on **(A)** PC1 and PC2 and **(B)** PC1 and PC3. The height of each surface describes the strength of the perceived cuteness. Red dots represent the PC score corresponding to the maximum of the function and blue dots represent the PC scores corresponding to the minimum.

and the faces with the maximum and minimum values were synthesized.

5.1. Effectiveness of the proposed method

The results of a LOOCV of ordinal prediction accuracy showed that the GPOR model had better prediction accuracy than the logistic ordinal regression model, indicating that non-linear models are effective in estimating the utility function of perceived cuteness. Another advantage of using non-linear regression models is that it allows the estimation of face shapes with maximum and minimum perceived cuteness. Among non-linear regression models (e.g., LOESS model, Komori et al., 2009b), The GPR model is superior because it can assess prediction uncertainty.

The commonly used method for estimating multivariate utility functions is a technique similar to the analysis of variance, called conjoint analysis (Rao et al., 2014). However, the conjoint analysis generally does not examine higher-order interactions because it requires a more significant number of trials. Moreover, the complicated balance of facial components influences perceived cuteness. Thus, the GPR is more suitable for studies on faces than conjoint analysis.

One of the conventional methods used to examine the morphological characteristics of faces related to perceived cuteness is to create a composite image of face images judged to be relatively cute and explore the characteristics of the composite image (Sprengelmeyer et al., 2009; Lobmaier et al., 2010; Hahn et al., 2013; Nittono et al., 2022). Such methods implicitly assume that the cuteness utility function exhibits a unimodal response, but if this response were multimodal, such procedures could lead to erroneous conclusions. Contrastingly, using GPR, it is possible to examine the relationship between face shape and perceived cuteness even if the utility function is multimodal. The results of this study do not show that the utility function for cuteness was multimodal. Nevertheless, future studies may find multimodality in the estimated function when adding the facial feature dimension.

5.2. Facial features associated with perceived cuteness

The parameters corresponding to the maximum value of the mean utility function suggest that a face with a low eye position, small chin, narrow jawline, and a small forehead are features that enhance perceived cuteness. Almanza-Sepúlveda et al. (2018) examined the relationship between infants' facial traits and cuteness argued that small chin and narrow jawline are related to the roundness of the face, and consequently, the roundness leads to cuteness. The results of this study are consistent with this explanation.

10.3389/fpsyg.2022.979341

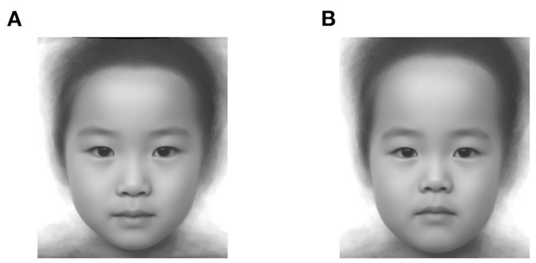


FIGURE 6
Facial images generated from the parameters corresponding to **(A)** the maximum (PC1: -0.8, PC2: -0.4, PC3: 0.8) and **(B)** the minimum of the averaged utility function of cuteness (PC1: 0.0, PC2: 0.4, PC3: −1.6).

Although many previous studies have argued that a large forehead is a cue for perceived cuteness (e.g., Hildebrandt and Fitzgerald, 1979), the results of this study suggest that forehead height itself does not contribute to cuteness enhancement. Rather, the width of forehead (represented in PC1) appears to be associated with perceived cuteness. Contrastingly, the relatively low position of facial parts such as the eyes and mouth contribute to cuteness. These results suggest that high foreheads may impair the roundness of the face, resulting in low cuteness. It should be noted that while previous studies examined infants' faces, this study used faces of 3- to 4-year-old children as stimuli. This discrepancy may be the cause of the inconsistency of the results. This means that different morphological factors may have affected perceived cuteness in the previous and present studies. It is necessary to conduct further research on the possibility that the morphological features of the face that affect perceived cuteness change along with growth.

5.3. Limitations and Future Directions

In this study, the locations of facial landmark coordinates were determined based on previous studies (Kamachi et al., 2001; Komori and Nittono, 2013), in which the shape of the forehead was measured with reference to the hairline. However, the shape of the hairline did not fully and accurately represent or replicate the features of the child's forehead. In the future, 3D measurements would be needed to encompass and replicate the three-dimensional shape of the forehead, which reflects the facial features of children.

Next, children's facial features were reduced to three PCs using PCA. Nonetheless, these three dimensions do not fully explain the variations in the facial features of young children. Notably, facial features not included in the three dimensions might affect perceived cuteness. Therefore, future studies should consider even greater dimensions of facial features. However, it is unclear how many trials would be required to construct a model with sufficient prediction accuracy when the number of dimensions considered was increased.

6. Conclusion

This study examined the relationship between children's face shapes and perceived cuteness in multivariate data using GPOR. Gaussian process ordinal regression is an extension of GPR that can be applied to ordinal scale responses and has not been used in psychological research before. This study, estimated the average utility function of perceived cuteness based on responses to a Likert scale, thereby identifying facial features associated with perceived cuteness.

This method can be applied to both research relating to facial cuteness and that pertaining to the study of various facial assessments, such as facial attractiveness, impressions, and stereotypes, and the shapes of the utility functions estimated from these investigations will provide clues to finding the factors that determine various judgments on faces.

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 Behavioral Research Ethics Committee of the Osaka University School of Human Sciences (HB019-058) and the Ethics Committee of the Osaka Electro-Communication University (18-004). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent for the publication of potentially identifiable images/data was not required as the figures included in the manuscript are composites generated using a large number of individual images.

Author contributions

MK, KS, and HN contributed to the conception, designed the study, and performed the experiment. MK and HN collected, measured, and synthesized facial images. MK, TT, and KS implemented the experimental and analytical programs. Further, MK and TT performed the statistical analysis. Finally, MK and TT wrote the draft of the manuscript and made figures and tables. All authors contributed to manuscript revision and read and approved the submitted version.

Funding

This study was supported by the JSPS KAKENHI (grant nos. 19K03375 to MK, 17H02651, and 21H04897 to HN).

Acknowledgments

We thank Haruka Shibata for her help in conducting the experiment.

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alley, T. R. (1981). Head shape and the perception of cuteness. *Dev. Psychol.* 17:650. doi: 10.1037/0012-1649.17.5.650

Almanza-Sepúlveda, M. L., Dudin, A., Wonch, K. E., Steiner, M., Feinberg, D. R., Fleming, A. S., et al. (2018). Exploring the morphological and emotional correlates of infant cuteness. *Infant Behav. Dev.* 53, 90–100. doi: 10.1016/j.infbeh.2018.08.001

Bookstein, F. L. (1991). *Morphometric Tools for Landmark Data.* New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511573064

Borgi, M., Cogliati-Dezza, I., Brelsford, V., Meints, K., and Cirulli, F. (2014). Baby schema in human and animal faces induces cuteness perception and gaze allocation in children. *Front Psychol.* 5:411. doi: 10.3389/fpsyg.2014.00411

Chu, W., Ghahramani, Z., and Williams, C. K. (2005). Gaussian processes for ordinal regression. *J. Mach. Learn. Res.* 6, 1019–1041.

Dryden, I. L., and Mardia, K. V. (1998). *Statistical Shape Analysis*. Chichester: Wiley. doi: 10.5555/1046920.1088707

Endendijk, J. J., Spencer, H., van Baar, A. L., and Bos, P. A. (2018). Mothers' neural responses to infant faces are associated with activation of the maternal care system and observed intrusiveness with their own child. *Cogn. Affect. Behav. Neurosci.* 18, 609–621. doi: 10.3758/s13415-018-0592-6

Glocker, M. L., Langleben, D. D., Ruparel, K., Loughead, J. W., Gur, R. C., and Sachser, N. (2009). Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults. *Ethology* 115, 257–263. doi: 10.1111/j.1439-0310.2008.01603.x

Hahn, A. C., Xiao, D., Sprengelmeyer, R., and Perrett, D. I. (2013). Gender differences in the incentive salience of adult and infant faces. *Quart. J. Exp. Psychol.* 66, 200–208. doi: 10.1080/17470218.2012.705860

Hildebrandt, K. A., and Fitzgerald, H. E. (1979). Facial feature determinants of perceived infant attractiveness. *Infant Behav. Dev.* 2, 329–339. doi: 10.1016/S0163-6383(79)80043-0

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., and Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception* 42, 875–887. doi: 10.1068/p3131

Komori, M., Kawamura, S., and Ishihara, S. (2009a). Averageness or symmetry: which is more important for facial attractiveness? *Acta Psychol.* 131, 136–142. doi: 10.1016/j.actpsy.2009.03.008

Komori, M., Kawamura, S., and Ishihara, S. (2009b). Effect of averageness and sexual dimorphism on the judgment of facial attractiveness. *Vis. Res.* 49, 862–869. doi: 10.1016/j.visres.2009.03.005

Komori, M., and Nittono, H. (2013). Influence of age-independent facial traits on adult judgments of cuteness and infantility of a child's face. *Proc. Soc. Behav. Sci.* 97, 285–291. doi: 10.1016/j.sbspro.2013.10.235

Kringelbach, M. L., Stark, E. A., Alexander, C., Bornstein, M. H., and Stein, A. (2016). On cuteness: unlocking the parental brain and beyond. *Trends Cogn. Sci.* 20, 545–558. doi: 10.1016/j.tics.2016.05.003

Lobmaier, J. S., Sprengelmeyer, R., Wiffen, B., and Perrett, D. I. (2010). Female and male responses to cuteness, age and emotion in infant faces. *Evol. Hum. Behav.* 31, 16–21. doi: 10.1016/j.evolhumbehav.2009.05.004

Lorenz, K. (1943). The innate forms of potential experience. *Z. Tierpsychol* 5, 235–409. doi: 10.1111/j.1439-0310.1943.tb00655.x

Löwenbrück, F., and Hess, U. (2021). Not all "caregivers" are created equal: Liking, caring and facial expression responses to the baby schema as a function of parenthood and testosterone. *Biol. Psychol.* 163:108120. doi: 10.1016/j.biopsycho.2021.108120

MacKay, D. J. (1996). "Bayesian methods for backpropagation networks," in *Models of Neural Networks III. Physics of Neural Networks*, eds E. Domany, J. L. van Hemmen, and K. Schulten (New York, NY: Springer), 211–254. doi: 10.1007/978-1-4612-0723-8_6

Neal, R. M. (1997). Monte Carlo implementation of Gaussian process models for Bayesian regression and classification. *arXiv Preprint*. doi: 10.48550/arXiv.physics/9701026

Nittono, H., Ohashi, A., and Komori, M. (2022). Creation and validation of the Japanese Cute Infant Face (JCIF) dataset. *Front. Psychol.* 13:819428. doi: 10.3389/fpsyg.2022.819428

O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *J. Roy. Stat. Soc. B Methodol.* 40, 1–24. doi: 10.1111/j.2517-6161.1978.tb01643.x

Peirce, J. W. (2007). Psychopy-psychophysics software in python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017

Rao, V. R. (2014). *Applied Conjoint Analysis*. Heidelberg; New York, NY; Dordrecht; London: Springer. doi: 10.1007/978-3-540-87753-0

Sprengelmeyer, R., Perrett, D. I., Fagan, E., Cornwell, R., Lobmaier, J., Sprengelmeyer, A., et al. (2009). The cutest little baby face: a hormonal link to sensitivity to cuteness in infant faces. *Psychol. Sci.* 20, 149–154. doi: 10.1111/j.1467-9280.2009.02272.x

Sternglanz, S. H., Gray, J. L., and Murakami, M. (1977). Adult preferences for infantile facial features: an ethological approach. *Anim. Behav.* 25, 108–115. doi: 10.1016/0003-3472(77)90072-0

Williams, C. K., and Barber, D. (1998). Bayesian classification with Gaussian processes. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 1342–1351. doi: 10.1109/34.735807

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 07 December 2022
DOI 10.3389/fpsyg.2022.955194

Check for updates

OPEN ACCESS

EDITED BY
Koyo Nakamura,
University of Vienna, Austria

REVIEWED BY
Megumi Kobayashi,
Niigata University, Japan
Tomoko Isomura,
Osaka University, Japan

*CORRESPONDENCE
Yuiko Sakuta
sakuta-yuiko@jissen.ac.jp

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

RECEIVED 28 May 2022
ACCEPTED 20 September 2022
PUBLISHED 07 December 2022

CITATION
Sakuta Y (2022) Face-to-trait inferences in Japanese children and adults based on Caucasian faces. *Front. Psychol.* 13:955194.
doi: 10.3389/fpsyg.2022.955194

Face-to-trait inferences in Japanese children and adults based on Caucasian faces

Yuiko Sakuta*

Faculty of Human Life Sciences, Jissen Women's University, Tokyo, Japan

Recently, many studies have indicated that humans make social evaluations from facial appearances instantaneously and automatically. Furthermore, such judgments play an important role in several social contexts. However, the mechanisms involved in the ability to form impressions from faces are still unknown, as is the extent to which these can be regarded as universal in perceiving impressions. In the current study, computer-generated Caucasian faces were used to assess the universality or cultural differences in impression formation among Japanese children and adults. This study hypothesized that impressions of trustworthiness and dominance may be more fundamental and universal, whereas the impression of competence may be more complex and culture-dependent. In Experiment 1a, 42 children aged 3–6 years were presented with 10 pairs of face images and asked which image in each pair was more trustworthy, dominant, or competent. Overall, it was found that as age increased, the rate of agreement of Japanese participants with the judgment of American participants, obtained in a previous study, increased. However, the agreement rate for competence was relatively low. Experiment 1b, conducted with 46 children, was a replication of Experiment 1a, and the results showed the same tendency. In Experiment 2, 45 Japanese adults made impression judgments on 19 pairs of face images identical to those used in Experiment 1b. The results suggested that while dominance was a dimension not easily influenced by developmental changes or culture, trustworthiness could be influenced by cultural differences in facial expression recognition. Therefore, different judgment criteria are used for children and adults. For competence, the agreement rate with Americans was relatively stable and low among the different age groups. This suggests that depending on the dimension of the trait, certain judgments are influenced by cultural context and, therefore, change criteria, while others are based on more universal criteria.

KEYWORDS
face-to-trait inference, development, impression, cultural differences, trustworthiness

Introduction

When we see a person, we often rapidly draw inferences regarding their traits, such as trustworthiness or competence. Recently, various studies have clarified that humans can immediately make social evaluations from facial appearance and that such judgments play an important role in several social contexts. The ability to rapidly detect a trustworthy person is especially important for forming good relationships within a social community.

Trait inference from a face occurs automatically, spontaneously, and rapidly (Willis and Todorov, 2006; Todorov et al., 2009). Todorov et al. (2009) demonstrated that it even occurred within 50 ms after exposure to a face and that trustworthiness influenced priming, even when the face was presented subliminally. Furthermore, various studies clarified that face-to-trait inference or impression judgment played an important role in several social contexts. For example, it can affect the outcome of the U.S. congressional elections (Todorov et al., 2005; Rule et al., 2010). Judgments regarding a candidate's competency correlate strongly with judgments regarding their facial attractiveness, which in turn can correctly predict approximately 70% of congressional electoral outcomes (Todorov et al., 2005). Thus, first impressions based on facial appearance unconsciously affect our choices or judgments in various social settings.

In social psychology, "warmth" and "competence" are two important decision axes in interpersonal impressions (Fiske et al., 2007; Cuddy et al., 2008). People in various occupations and positions in society are distributed in a two-dimensional space based on a combination of warmth and competence parameters. Oosterhof and Todorov (2008) also proposed trustworthiness and dominance as basic dimensions through a principal component analysis (PCA) as described below.

Recently, data-driven approaches have attracted attention. Oosterhof and Todorov (2008) had participants infer characteristics from faces and applied a PCA to reduce characteristic judgments to two basic dimensions: trustworthiness/valence and dominance. They collected trait ratings for 300 computer-generated faces to cross-validate their model. The principal components of the physical attributes of these faces were provided. Oosterhof and Todorov (2008) mapped the perceived characteristic dimensions onto the "face" defined by these physical dimensions, centered on an average face [based on a procedure by Blanz and Vetter (1999)]. In this model, faces higher than the average of the trustworthiness dimension appeared to smile, while those lower appeared to be increasingly angry. Dominant faces appeared more mature, masculine, and darker (Oosterhof and Todorov, 2008; Todorov and Oosterhof, 2011). Furthermore, the faces generated to be placed on these two dimensions were actually perceived to differ in trustworthiness and dominance by new participants (Oosterhof and Todorov, 2008). A similar approach was proposed by Sakuta et al. (2009). Sakuta et al. (2009) extracted factors that roughly matched the three factors by Osgood et al. (1957). Sutherland et al. (2013) used various face images, which included a wide range of ages, expressions, and poses, and found an additional dimension, "youthful-attractiveness," which related increasing age to perceptions of decreasing attractiveness, and extended model Oosterhof and Todorov (2008).

As visualized by Oosterhof and Todorov (2008), impression judgments from faces are considered to be a generalization of emotion perception. Happy and angry facial expressions increase and decrease trustworthiness judgments, respectively (Oosterhof and Todorov, 2008; Todorov et al., 2008; Sutherland et al., 2017). Even when presented with neutral faces, adults are more likely to use subtle expressive signs of positive or negative affect to judge trustworthiness, a phenomenon known as emotional overgeneralization (Todorov, 2008; Zebrowitz, 2017). In other words, previous findings indicated that adults tended to trust individuals who expressed positive rather than negative emotions, and did so even when the expressive cues were subtle. Furthermore, Verosky and Todorov (2013) showed that participants rated new faces that resembled familiar negative faces more negatively than those that resembled familiar positive faces. The authors stated that generalization learning based on the physical similarity of faces was a powerful and relatively automatic process that could influence face ratings in various situations. In addition, Lee et al. (2021) used artificial object images (greebles) in an experiment with adults to examine whether the origin of first impressions was due to learning. Results showed that first impressions from appearance were rapidly learned, and that once learned, they were detectable in a very short time. This study suggests that first impressions are not necessarily innate. The current study focuses on the development and cultural differences on impression perception.

Several studies have reported cultural universality in the trait-inference from faces (e.g., Zebrowitz et al., 1993, 2012; Cunningham et al., 1995; Rule et al., 2010, 2011), while others reported both universalities and some differences between cultures (Walker et al., 2011; Sutherland et al., 2018). Particularly, the finding that the Tsimane people, who live isolated in the Bolivian rainforest, and American raters generally agreed on facial impressions is noteworthy (Zebrowitz et al., 2012). The cross-cultural similarity in the characteristics of impressions received from faces has been considered to be partially influenced by culture-specific perceptual learning, although there may be universal mechanisms at work. Sutherland et al. (2018) built a data-driven model with Chinese and British people, and found common "approachability" and strong evidence for the "youthfulness" dimension, as well as evidence for a third dimension similar to "competence." Impressions along these dimensions were thought to be based on adaptive cues for threat detection and sexual selection, and this finding was strong evidence that facial impressions were shared across cultures.

In contrast, many studies have shown cultural differences in face perception. For example, several studies that measured eye gaze during face observation found cultural differences in gazing behavior. For example, in Blais et al. (2008), Western Caucasian observers showed a triangular gaze pattern, which consisted of two eyes and a mouth for both races' faces, for all tasks. Conversely, East Asians gazed more at the center of the face. These results indicated that face processing could not be considered universal. Jack et al. (2009) also conducted a gaze measurement experiment and found that Eastern observers gazed relentlessly at the eye area instead of evenly across the face as Westerners did. The authors also pointed out that the strategies and categories of facial expression recognition differed by culture.

Recent studies have pointed out the possibility that infants perceive impressions similar to adults to an extent. Cogsdill et al. (2014) verified that 3-year-old children made impression judgments in a manner similar to adults. Such judgments became more stable at 5 years and older. Recent studies revealed that even infants detected trustworthiness in faces (Jessen and Grossmann, 2016, 2019; Sakuta et al., 2018). Thus, young children could have sensitivity for facial features that evoke facial impression, especially trustworthiness. Cogsdill and Banaji (2015) used face images of adults, children, and macaques to examine the agreement between characteristic inferences from child and adult faces. An agreement was found between child and adult judgments for all face stimuli. From these results, it was concluded that trait inferences from faces were acquired early in childhood. However, it is still unknown how the ability to perceive impressions from faces develops in very young ages, how the impressions are related to facial expressions, and whether the impressions are influenced by culture.

As noted above, first impressions have been reported to be influenced by subtle facial expressions in adults; however, there are conflicting findings in children. According to previous studies, such as Ewing et al. (2019) and Tang et al. (2019), young children's trust judgment was influenced by facial expressions. In contrast, Mondloch et al. (2019) reported that, unlike adults, there was no evidence that children aged 4–11 years used facial expressions when forming impressions. Furthermore, it was noted that the relationship between facial expressions and first impressions was similar for children as for adults in explicit impression judgment tasks. However, no consistent findings were available for tasks that involved implicit processing (Van Der Zant et al., 2021).

Thus, research on impression perception in infancy has received increasing attention recently. However, much remains unknown. To approach cultural differences, a detailed examination of impression perception during infancy, when the influence of culture and experience is still small, may be an effective method.

In the current study, I used computer-generated Caucasian faces to assess the universality of perceiving facial impressions in Asian (Japanese) children. Previous studies have shown that the other-race effect in infants occurs at around 6 months of age and is present at around 9 months of age (Pascalis et al., 2005; Kelly et al., 2007). However, these previous studies were only concerned with face identification and not with impression consensus. Therefore, it cannot be said that other-race effects also occur in infants' impression perception. Japanese infants aged 6–8 months have been found to gaze at faces deemed trustworthy by American adults in experiments with Caucasian faces (Sakuta et al., 2018). Thus, it is possible that at least infants were able to detect some trait that leads to trustworthiness, regardless of the race of the face. Moreover, one of the cues for trust detection is a facial expression that looks like a smile (e.g., raised corners of the mouth), and it is possible that American adults and Japanese infants alike use "raised corners of the mouth" as a cue to perceive trustworthiness. However, the facial stimuli used in this study and in Sakuta et al. (2018) were originally created with neutral expressions and do not show a clear smile, although the faces with high trustworthiness appear to be smiling slightly. It should be noted that at around 6–7 months, infants cannot perceive subtle facial expressions without movement (Ichikawa et al., 2014). Hence, it is possible that the facial images used in the current and Sakuta et al. (2018) studies were not perceived as smiles by the infants. Therefore, it may be necessary to consider cues other than facial expressions in infants' impression perception. To examine whether the impression perceptions in Sakuta et al.'s (2018) were found only in infants or also in older children, it was necessary to use the same stimuli as in Sakuta et al. (2018). Testing whether infants are able to judge impressions on faces of other races and comparing the judged impressions with results obtained in other countries will be important for elucidating the mechanisms of impression perception.

According to previous studies (Oosterhof and Todorov, 2008; Cogsdill et al., 2014), it is possible that there could be both universality and some cultural differences based on the kind of impressions. In Oosterhof and Todorov's (2008), trustworthiness was treated as a dimension independent of dominance. Results of the PCA for trait judgment showed that the first principal component, trustworthiness, had high loadings for adjectives, such as trustworthy, emotionally stable, responsible, and sociable, while the second principal component, dominance, had high loadings for dominant, aggressive, and confident. Cogsdill et al. (2014) also addressed trustworthiness as a basic evaluation and dominance and competence as more specific traits. Considering the above, trustworthiness and dominance could be more fundamental and universal, whereas competence could be more complex and culture-dependent.

Focusing on one's appearance is an adaptive ability that helps us rapidly detect a friend or foe in a social environment. Judgment of trustworthiness can predict an individual's social and economic success (Todorov, 2008;

van 't Wout and Sanfey, 2008). Judgment of competence is important for choosing a leader in adults and children (Todorov et al., 2005; Antonakis and Dalgas, 2009). Todorov et al. (2008) noted that evaluating faces on valence (or trustworthiness) and dominance may be an overgeneralization of adaptive mechanisms, which attempts to estimate others' behavioral intention and status in a hierarchy of power. The current study chose stimuli that varied on the dimensions of trustworthiness, dominance, and competence as these were important social dimensions, at least for adults. It is worth examining whether these are important for young children as well.

To examine what facial impressions were perceived in Asian people, Experiment 1a assessed Japanese children's judgment of computer-generated Caucasian faces. Experiment 1b was conducted with another group of Japanese children to validate the results of Experiment 1a. Lastly, Japanese adults' impression judgment was assessed in Experiment 2.

Experiment 1a: Impression formation in Japanese children

Experiment 1a intended to replicate the prior studies regarding impression formation in children (Cogsdill et al., 2014; Sakuta et al., 2018).

Method

Participants

This study included 42 children (23 girls and 19 boys) divided into two age groups: 3–4-year-olds ($n = 25$, range: 2 years 11 months–4 years 10 months) and 5–6-year-olds ($n = 17$, range: 5 years 4 months–6 years 10 months). The experiment was conducted from February to March 2017. Previous studies have shown that toddlers exhibit the face inversion effect at about 6 years of age and develop holistic processing in face recognition similarly to adults (Brace et al., 2001). Therefore, comparing the developmental process of face recognition by dividing the children into younger and older groups is possible. All the children were Japanese and recruited from a nursery school. The following experiments (Experiments 1a, 1b, and 2) were approved by the Ethical Committee of Jissen Women's University (H28-18). Moreover, the experiments were conducted according to the principles laid down in the Declaration of Helsinki. Written informed consent was obtained from each child's parents prior to their participation. A sample size was determined based on previous research (Oosterhof and Todorov, 2008), where approximately 20 participants made impression judgments in each study.

Stimuli

We used computer-generated face images, in which trustworthiness, dominance, and competence were manipulated, as stimuli. Face images were obtained from a database created in FaceGenModeller 3.2 (Singular Inversions)[1], validated (Oosterhof and Todorov, 2008), and open to the public[2]. These were created based on data-driven, computational models (derived from adults' judgments in the U.S.) of the respective traits (Oosterhof and Todorov, 2008). Oosterhof and Todorov (2008) proposed a two-dimensional (2D) model of face evaluation, which consisted of two orthogonal axes of valence (trustworthiness) and dominance. These two axes were composite factors obtained from a PCA: valence evaluation included positive judgments of attravvctiveness and responsibility and dominance evaluation consisted of judgments of dominance, aggressiveness, and confidence. These axes were the most important dimensions in social judgment, yet other social evaluations were also representable in this 2D model. All faces were bald, Caucasian males.

The stimuli were chosen from two databases: (A) Distinct face identities manipulated on face shape and reflectance, on several dimensions. The current study used the faces with manipulation −3SD and +3SD on competence, dominance, and trustworthiness dimensions. These databases were validated by Todorov et al. (2013). Cogsdill et al. (2014) used this type of stimuli to assess American children's impression formation. (B) The average face was manipulated on face shape and orthogonally on perceived trustworthiness and dominance, parametric face manipulation that ranged from −3 to +3SD. The current study used the faces with manipulation −2SD and +2SD. These orthogonal models were created by Oosterhof and Todorov (2008). Sakuta et al. (2018) used this type of stimuli to assess Japanese infants' impression formation.

All of the face stimuli were created by the common procedure described above; however, they contained different images. We chose six pairs (two each for the three traits: competence, dominance, and trustworthiness) from set A and four pairs (two pairs each for trustworthy-untrustworthy and dominant-submissive) from set B. Altogether, four trustworthy-untrustworthy, four dominant-submissive, and two competent-incompetent pairs were presented to each participant. Examples of the stimuli are shown in **Table 1**. Using Adobe Photoshop software, the images were cropped in contour and superimposed on a uniform white background. These were printed on glossy paper side-by-side.

Procedure

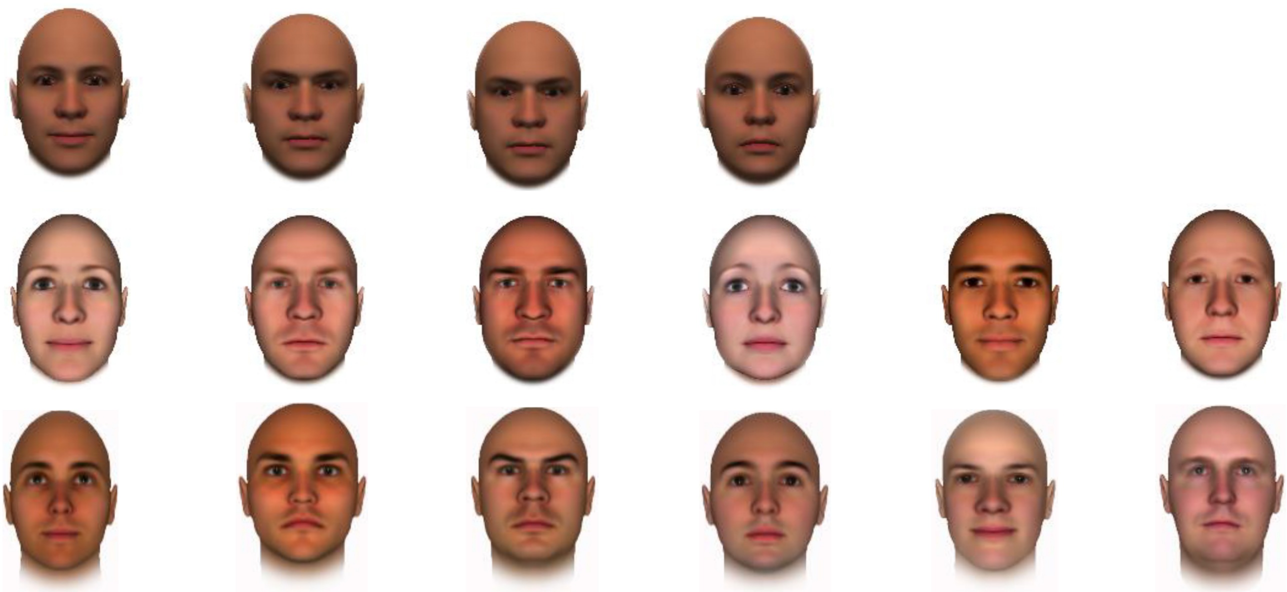
Participants performed two-alternative forced-choice judgments in each pair. First, an experimenter showed two cartoon characters to relax and explain to the child what to do in the experiment, and confirmed that they were able to complete the task. Next, two characters were presented side-by-side on

1 http://www.facegen.com/

2 https://tlab.uchicago.edu/databases/

TABLE 1 Examples of the stimuli. Images in the upper and middle rows were used in Experiment 1a.

Trustworthiness		Dominance		Competence	
High	Low	High	Low	High	Low



The upper four stimuli were identical to the stimuli used in Sakuta et al. (2018) and were taken from Oosterhof and Todorov (2008). Reproduced with permission. The middle six were chosen from Oosterhof and Todorov (2008). Reproduced with permission. Images in the lower row were added in Experiments 1b and 2. They were identical to the stimuli used in Cogsdill et al. (2014). (Source: https://osf.io/p85eb).

a white panel and the participants were asked to choose the one that they thought was better. Further, two stimuli were presented side-by-side as well. Participants were asked to judge which person was better (trustworthy), stronger (dominant), or smarter (competent). Each participant was exposed to 10 trials. Presentation order and position were counterbalanced across the participants. For example, a face that was judged as trustworthy by American adults (Oosterhof and Todorov, 2008) was presented on the left for half of the participants and on the right for the other half. Each participant's behavior was recorded both in the video and recording sheets.

Results

For each face in each stimulus pair, the percentage of children who selected the trustworthy face as the "nice" one was recorded for trustworthy-untrustworthy pair. Higher percentages indicated stronger consensus between Japanese children and American adults. Consensus data are shown in **Figure 1** together with the data from Experiments 1b and 2. Please refer to the **Supplementary material** for the individual data. Frequencies of the choice were shown in **Supplementary Table 1**.

As a result of a one sample t-test against chance level, the difference was significant only for trustworthiness among the 3–4-year-olds [$t(24) = 2.43$, $p = 0.02$, $r = 0.44$], and significant for trustworthiness and dominance among the 5–6-year-olds [trustworthiness: $t(16) = 6.91$, $p < 0.001$, $r = 0.87$; dominance: $t(16) = 4.52$, $p < 0.001$, $r = 0.75$]. That is, both age groups showed significant consensus when identifying faces as mean or nice (66.00% for 3–4-year-olds; 82.35% for 5–6-year-olds). Only older children showed a significant consensus when identifying faces as strong or not strong (60.00% for 3–4-year-olds; 76.47% for 5–6-year-olds). No significant consensus above chance was shown for dominance among the 3–4-year-olds [60.00%, $t(24) = 1.73$, $p = 0.10$, $r = 0.33$] and smart or not smart pairs [56.00% for 3–4-year-olds, $t(24) = 0.65$, $p = 0.52$, $r = 0.13$; 64.71% for 5–6-year-olds, $t(16) = 1.43$, $p = 0.17$, $r = 0.34$].

A two-way analysis of variance (ANOVA) of age group × type of trait for the consensus rate revealed a significant main effect of age group, $F(1, 40) = 5.94$, $p = 0.02$, partial $\eta^2 = 0.13$. Although the 3–4-year-olds responded with American adult like consensus, they were less consistent than the 5–6-year-olds. The main effect of trait and the interaction between age group and trait were not significant, $F(2, 80) = 1.52$, $p = 0.23$, partial $\eta^2 = 0.04$; and $F(2, 80) = 0.16$, $p = 0.82$, partial $\eta^2 = 0.004$, respectively.

Discussion

Some differences were found in the development of impression formation. Based on the t-tests, for trustworthiness

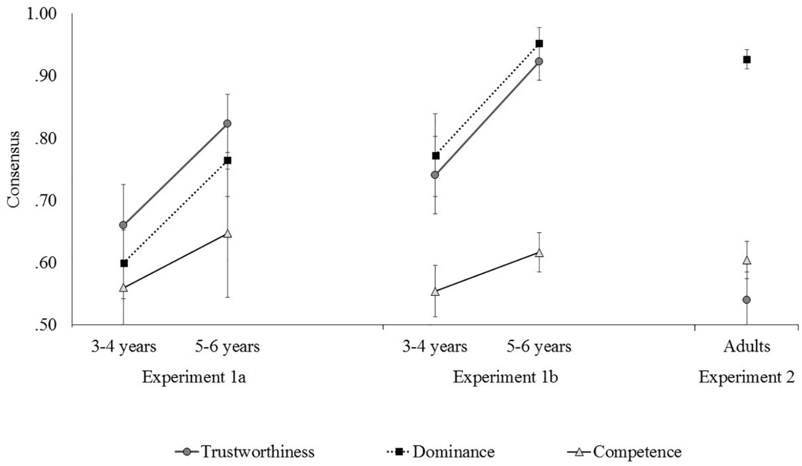


FIGURE 1
Consensus data for Japanese young children (Experiments 1a and 1b) and adults (Experiment 2) with the U.S. adults are shown together. (The error bar indicates standard error).

and dominance, older children showed better agreement with American adults than younger children. For competence, there was no relationship between ages and impression judgment. In sum, trustworthiness and dominance judgment became more consistent as children grew up, whereas competence judgment may have been more difficult for the children. However, as the ANOVA showed that only the main effect of age was significant and the interaction was not, it can be said that the overall agreement rate varies with age, although it cannot be said that the timing of development is different for the different types of impressions.

Experiment 1b: Impression formation in Japanese children (Replication of Experiment 1a)

Experiment 1a used the same impression-manipulated stimulus set as Cogsdill et al. (2014); however, it was not identical. Therefore, in Experiment 1b, a stimulus set identical to that of Cogsdill et al. (2014) was added and the impression judgments from faces in Japanese children examined. If this experiment showed a tendency consistent with the results of Experiment 1a, more robust findings should be obtained on impression judgments in early childhood in Japan.

Method

Participants

This study included 46 children (25 girls and 21 boys) divided into two age groups according to their class: 3–4-year-olds ($n = 22$) and 5–6-year-olds ($n = 24$). Although the year ranges were almost equivalent to Experiment 1a, 29 children did not disclose their birth month. The experiment was conducted from February to March 2018. There were eight children who were categorized into younger group in Experiment 1a and in the older group in Experiment 1b. All the children were Japanese and recruited from a nursery school. Sixteen children had also participated in Experiment 1a. The experiments were conducted according to the principles laid down in the Declaration of Helsinki. Written informed consent was obtained from each child's parents prior to their participation.

Stimuli

The same 10 stimuli pairs as in Experiment 1a were used again. Additionally, 18 stimuli were chosen from Cogsdill et al. (2014). Altogether seven trustworthy-untrustworthy, seven dominant-submissive, and five competent-incompetent pairs were used. Using Adobe Photoshop software, the images were cropped in contour and superimposed on a uniform white background. These were printed on glossy paper side-by-side.

Procedure

The procedure was identical to that of Experiment 1a. Each participant's behavior was recorded both in the video and recording sheets.

Results

As per Experiment 1a, the percentages of consensus were calculated (**Figure 1**). Frequencies of the choice were shown in **Supplementary Table 2**. As a result of a one sample t-test versus chance level, both age groups showed significant consensus when identifying faces as mean or nice [74.03% for 3–4-year-olds, $t(21) = 3.87$, $p < 0.001$, $r = 0.65$; 92.26% for 5–6-year-olds,

$t(23) = 14.20, p < 0.001, r = 0.95$] and as strong or weak [77.27% for 3–4-year-olds, $t(21) = 4.08, p < 0.001, r = 0.67$; 95.24% for 5–6-year-olds, $t(23) = 17.87, p < 0.001, r = 0.97$]. Only older children showed significant consensus when identifying faces as smart or not smart [55.45% for 3–4-year-olds, $t(21) = 1.32$, $p = 0.20$, $r = 0.28$; 61.67% for 5–6-year-olds, $t(23) = 3.69$, $p < 0.01, r = 0.61$].

A two-way ANOVA revealed significant main effects of age group and trait, $F(1, 44) = 12.14$, $p < 0.001$, partial $\eta^2 = 0.22$; $F(2, 88) = 26.24$, $p < 0.001$, partial $\eta^2 = 0.37$, respectively. Although the 3–4-year-olds responded with adult like consensus, they were less consistent than the 5–6-year-olds. Competence judgment showed lower consensus than trustworthiness and dominance. The interaction between age group and trait was not significant, $F(2, 88) = 1.34$, $p = 0.27$, partial $\eta^2 = 0.03$.

Discussion

We found some differences in the development of impression formation, as in Experiment 1a. Compared to Experiment 1a, the overall consensus in judgment with American adults increased; however, the overall trend was similar to that of Experiment 1a. For all impressions, older children showed better agreement with American adults than younger children. Trustworthiness and dominance judgment became more consistent as children grew up, whereas competence judgment may have been more difficult for the children. However, as in Experiment 1a, the interaction between age group and trait was not significant. Therefore, it cannot be said that the rate of consensus in trait judgments differs by age.

Experiment 2: Impression formation in Japanese adults

Using the identical stimuli set, Japanese adults' impression judgment was examined and compared to that of Japanese children (data from Experiments 1a and 1b) and American adults (Cogsdill et al., 2014).

Method

Participants

This study included 45 Japanese adults (33 women and 12 men, mean age: 20.3 years old, range: 18–23 years old). Participants provided informed consent to the procedures, and the study was approved by the Ethical Committee of Jissen Women's University. The experiment was hosted online using a Google form and participants were recruited through the University. The experiments were conducted according to the principles laid down in the Declaration of Helsinki.

Stimuli

We used identical stimulus sets to Experiment 1b. Altogether, seven trustworthy-untrustworthy, seven dominant-submissive, and five competent-incompetent pairs were used in Experiment 2.

Procedure

Participants were asked to judge which person was better (trustworthy), stronger (dominant), or smarter (competent). Different from Experiments 1a and 1b, the adults performed a forced-choice judgment with 5-point confidence rating scales and answered using an online form (Google form). For example, to the question "Which person seems to be more trustworthy?" participants were required to choose one of the following: 1: definitely left, 2: probably left, 3: not sure, 4: probably right, 5: definitely right. Presentation order and position were counterbalanced among the participants.

Results

The percentages of responses were calculated along with the five response categories for each stimulus pair. The five categories indicated the five degrees varied from the unexpected response with higher confidence (i.e., low consensus with the American judgment) to expected response with higher confidence (i.e., high consensus with the American judgment). Thus, they corresponded with the degree of consensus to the American adults' judgment. As a result of a chi-squared test, all traits showed significance [trustworthiness: $\chi^2(4) = 49.33$, $p < 0.01$, dominance: $\chi^2(4) = 442.03, p < 0.01$, and competence: $\chi^2(4) = 62.80$, $p < 0.01$]. As shown in Table 2, dominance judgment showed the highest consensus.

Next, we summed the selection frequencies of "definitely left" and "probably left" and "probably right" and "definitely right" for each stimulus pair to create judgment categories consistent and inconsistent with the American judgments, respectively. The percentage of responses for each judgment categories was calculated, and a two-way ANOVA with age group and type of trait as factors was performed. Participants' data from Experiments 1a and 1b were also used for this analysis. Averages of agreement with U.S. were shown in Supplementary Table 3. As noted above, 16 children participated in both experiments; in Experiment 2, these participants' data were regarded as independent from that of the participants who were different in both experiments. All main effects and an interaction were significant, age group: $F(2, 130) = 10.46$, $p < 0.001$, partial $\eta^2 = 0.14$; trait: $F(2, 260) = 25.40$, $p < 0.001$, partial $\eta^2 = 0.16$; interaction: $F(4, 260) = 9.53$, $p < 0.001$, partial $\eta^2 = 0.13$. Simple main effect tests for interaction showed that differences among traits were significant in all age groups (3–4-year-olds: $p = 0.02$, 5–6-year-olds: $p < 0.001$, adults: $p < 0.001$). Multiple comparisons using Holm's method showed that dominance

TABLE 2 Japanese adults' impression judgment (choice rate; %) and the result of the chi-squared-test in Experiment 2.

	Consensus with U.S.					
	Low			High		
	Very confident (definitely)	Less confident (probably)	Not sure	Less confident (probably)	Very confident (definitely)	$\chi 2$
Trustworthiness	6.98	20.00	19.05	31.75	22.22	49.33, $p < 0.01$
Dominance	1.59	0.95	4.76	30.79	61.90	442.03, $p < 0.01$
Competence	7.11	13.78	18.67	38.67	21.78	62.80, $p < 0.01$

$\chi 2$ indicates the chi-squared test. Choice rate for each of the five categories was calculated for three types of stimuli sets. For example, when a participant chose a left image as trustworthy with high confidence, and the image was a trustworthy face in the U.S., it was qualified as "high consensus with U.S".

judgment (92.7%) was significantly more consistent than trustworthiness (54.0%) and competence (60.4%) judgments in adults (adjusted $ps < 0.01$). Competence judgment (62.9%) was less consistent than the other two traits (trustworthiness: 88.2%, dominance: 87.5%) in 5–6-year-olds (adjusted $ps < 0.01$). This tendency was also seen in 3–4-year-olds, but the adjusted p-value was not significant (trustworthiness: 69.8%, dominance: 68.1%, competence: 55.7%). In terms of traits, the difference in consensus rates among age groups were significant for trustworthiness ($p < 0.001$) and dominance ($p < 0.001$), but not for competence ($p = 0.43$). Adults had the lowest consensus rate for trustworthiness judgments (54.0%), followed by 3–4-year-olds (69.8%) and 5–6-year-olds (88.2%). For dominance judgments, adults (92.7%) and 5–6-year-olds (87.5%) had similar consensus rates, with 3–4-year-olds having the lowest rates (68.1%).

Discussion

When data obtained from Japanese children and adults were analyzed together, it was found that the consensus with the U.S. data increased with age for the dominance judgment. However, the consensus for trustworthiness judgment decreased to near chance level in adults compared to children. Competence judgment was approximately 60% consensus in all age groups and did not vary with age. The results showed that some impression judgments became more similar between Japanese and American judgments as the respondents' age increased, while others decreased in consensus. Trustworthiness was more likely to be influenced by culture and experience, as consensus was lower in adults compared to early childhood.

General discussion

Overall, in Experiment 1a, agreement with the American judgments increased from ages 3–4 to 5–6 years. Competence judgments showed a relatively low agreement with Americans (56–64%). These results were consistent with the prediction that trustworthiness and dominance would be more universal judgment dimensions, and competence would be more influenced by culture and experience. Using the same stimulus set, Sakuta et al. (2018) found that for trustworthiness, Japanese infants also showed preferential gazing at face images that Americans judged trustworthy. Therefore, Japanese infants tended to agree with the American judgments regarding trustworthiness. In addition, Cogsdill et al. (2014) reported that American children judged competence similarly to American adults, with agreement rates of 75% at ages 3–4 and nearly 90% at ages 5–6 years old. This suggested that the low agreement rate in competence judgments in Experiment 1a was not solely due to the underdevelopment of the judgment criteria in early childhood. It was also likely influenced by culture. Experiment 1b was conducted with the same additional stimulus set as in Cogsdill et al. (2014). Overall, agreement with Cogsdill et al. (2014) increased; however, the trend was the same as in Experiment 1a. Trustworthiness and dominance had higher agreement rates, while competence had lower agreement rates. The overall increase in agreement was due to the addition of the same stimulus set as in the previous study. Hence, this was not surprising. However, it is possible that the stimuli judged to be highly trustworthy in the U.S. may also be judged to be highly trustworthy or not in Japan, depending on the images used.

As 16 of the 46 participants in Experiment 1b also participated in Experiment 1a, the possibility that these children's experience with inferring traits from faces in Experiment 1a may have influenced their face-to-trait inferences in Experiment 1b cannot be ruled out. However, in this study, participants were not given individual feedback regarding the results of their judgments. Therefore, it was impossible for them to obtain information such as how the face image they selected was judged in the U.S. (e.g., whether the person was considered more trustworthy). Therefore, it is unlikely that children who participated in both experiments would have a higher rate of agreement with the American participants than children who participated only once. Furthermore, the mean agreement rates for children who participated only in Experiment 1b ($n = 30$) and those who participated in both ($n = 16$) were found to be 77.3 and 74.4%, respectively, and a *t*-test showed no significant difference [$t(44) = 0.61$, $p = 0.55$]. In summary, while it is conceivable that past experience with trait inference may influence subsequent trait inferences, it does not appear to be a reason for the higher agreement rate in Experiment 1b than in Experiment 1a in the present data.

Experiment 2 used the same stimulus set as Experiment 1b and was conducted with Japanese adults. Only dominance was in agreement with Cogsdill et al. (2014) at over 90%, whereas trustworthiness and competence had only 54 and 60% judgmental agreement, respectively. These results were consistent with the prediction that competence would be influenced by culture, yet contrary to the prediction that trustworthiness would be a universal dimension. In particular, the finding that trustworthiness judgments were less consistent with American judgments in adults than in infancy was surprising.

As for the difference in trustworthiness judgment, it is possible that there is a difference in the meaning of a smile between Japan and America. Regarding evaluation, smiling has generally been thought to be positively valued. However, when Krys et al. (2014) compared intellectual impressions of smiling people across seven countries, they found that in Germany and China smiling people were perceived as more intelligent than non-smiling people, while in Iran, on the contrary, they were perceived as less intelligent. It was shown that in some cultures, smiling people may be evaluated more negatively than non-smiling people. Matsumoto and Kudoh (1993) also reported the differences of trait judgment on smiling and non-smiling faces between American and Japanese raters. For example, Americans rated a smiling person as more intelligent than a neutral face, while the Japanese did not show such a difference. Furthermore, while there was agreement that smiling was more sociable than a neutral face, the degree of sociability was greater among Americans than among the Japanese. These findings suggest that due to cultural differences in facial expression recognition, there is a strong possibility that cultural differences also occur in impression perception, considered to be dependent on facial expressions.

There are two possible factors, other than cultural differences in the evaluations themselves, that may explain the present results: the race of the face and terms of the evaluation.

First, it should be noted that this may include the influence of other-race effects. The other-race effect is a phenomenon in which the face of another race is more difficult to distinguish than the face of one's own race, and it has been observed even in infants less than 1 year old (Kelly et al., 2009, 2007). As composite images based on Caucasian faces were used for Japanese children and adults in this study, the possibility that the obtained results included only differences in impression perception due to other-race effects and culture cannot be denied. Among the results obtained in this study, the other-race effect was possibly involved in the fact that the agreement rate with the judgment of Americans was lower for adults than for infants. However, Japanese infants aged 6–8 months old have been found to gaze at faces judged trustworthy by American adults in experiments with Caucasian faces (Sakuta et al., 2018). Thus, at the very least, infants were able to detect some physical cues that induce the impression of trustworthiness, regardless of the race of the face. However, as described in the Introduction, it cannot be said that they merely gazed at faces that appeared to smile. Therefore, it may be necessary to include cues other than facial expressions in regard to infants' impression perception. As, currently, very few studies on cultural differences in impression perception from faces have been conducted, further research is needed.

Second, when examining cultural differences, it is necessary to consider the differences in word connotations. In Cogsdill et al. (2014), trustworthiness was translated as "mean/nice," dominance as "strong/not strong," and competence as "smart/not smart." It may be debatable whether this paraphrasing was appropriate, that is, whether trustworthy had exactly the same meaning as nice for adults. In addition, it is possible that the words "trustworthiness," "dominance," and "competence" in the U.S. and "*shinraikan*," "*shihaisei*," and "*yuunousa*" in Japan, as well as "nice," "strong," and "smart," paraphrases of these words for young children in the U.S., and "*iihitosou*," "*tsuyosou*," and "*atama ga yosasou*" for young children in Japan do not necessarily have the exact same meanings. It is also possible that the meaning of "trustworthy" as used by adults is different between the U.S. and Japan. The English word "nice" and the Japanese word "*ii* (良い)" are thought to correspond almost semantically. However, the Japanese word "*shinraidekiru* (信頼できる=trustworthy)" does not necessarily mean the same thing as a "good" person. We cannot deny the possibility that the intentional impression-judgment task and verbal response may have resulted in subtle differences in nuance, or that the instruction was not well conveyed to younger children. Therefore, in the future, it will be necessary to examine cultural differences in impression perception without language by conducting cognitive tasks that do not depend on language.

Recently, the Western, Educated, Industrialized, Rich, and Democratic (WEIRD) problem has been noted (Henrich et al., 2010). Namely, many articles regarding psychology have been published based on samples drawn entirely from WEIRD societies. Furthermore, in social cognition, many studies are still published from the U.S., Europe, and other countries, and the findings may be biased as mainly Caucasian faces are used (Cook and Over, 2021). The current study examined the impressions that Japanese children and adults had of computer-generated images based on Caucasian faces. By using the same images as in a previous study conducted in the U.S., the responses to the same stimuli could be directly compared. This is a big advantage in examining the basic face perception. In the future, it may be possible to examine the generalizability of the findings by mixing the faces of various races and using realistic human face images. In this study, CG composite images were used; however, realistic facial images can be used to reproduce realistic situations and provide findings with greater ecological validity.

Conclusion

The findings of this study can be summarized in two points: First, while the trend in Japanese young children becomes more similar to that in Americans as they get older, the degree of agreement in the judgment of trustworthiness decreases in adults, suggesting that the criteria for judging impressions from faces changes significantly between early childhood and adulthood. Second, the results suggest that some criteria for such impression judgments are less susceptible to cultural influences than others, depending on the dimension of the trait. In this study, contrary to expectations, it was the trustworthiness impression that was more susceptible to the influence of culture. Together with the adults' data, there could be some differences among cultures and ages. By approaching the cultural differences in impression perception from the aspect of impression perception in early childhood, the differences among impression dimensions were examined in detail. Recognizing that there can be discrepancies in cross-cultural communication due to differences in impression judgments and facial expression recognition from the face by accumulating such studies will be useful for forming smooth cross-cultural communication.

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 the Ethics Committee of Jissen Women's University. Written informed consent to participate in this study was provided by the participants or their legal guardian/next of kin in Experiments 1a and 1b and from the participants themselves in Experiment 2.

Author contributions

YS designed the study, conducted the experiments, analyzed the results, and wrote the manuscript.

Funding

This work was supported by JSPS KAKENHI Grant Numbers JP15K04192, JP19K12188, and JP20K03481.

Acknowledgments

The author would like to thank the nursery school in Tokyo for their help in recruiting the children, the students in the University for their help in data collection, and Editage (www.editage.com) for English language editing.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.955194/full#supplementary-material

References

Antonakis, J., and Dalgas, O. (2009). Predicting elections: Child's play! *Science* 323:1183. doi: 10.1126/science.1167748

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., and Caldara, R. (2008). Culture shapes how we look at faces. *PLoS One* 3:e3022. doi: 10.1371/journal.pone.0003022

Blanz, V., and Vetter, T. (1999). "A morphable model for the synthesis of 3D faces," in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, (New York, NY: SIGGRAPH), 187–194. doi: 10.1145/311535.311556

Brace, N. A., Hole, G. J., Kemp, R. I., Pike, G. E., Van Duuren, M., and Norgate, L. (2001). Developmental changes in the effect of inversion: Using a picture book to investigate face recognition. *Perception* 30, 85–94. doi: 10.1068/p3059

Cogsdill, E. J., and Banaji, M. R. (2015). Face-trait inferences show robust child-adult agreement: Evidence from three types of faces. *J. Exp. Soc. Psychol.* 60, 150–156. doi: 10.1016/j.jesp.2015.05.007

Cogsdill, E. J., Todorov, A. T., Spelke, E. S., and Banaji, M. R. (2014). Inferring character from faces: A developmental study. *Psychol. Sci.* 25, 1132–1139. doi: 10.1177/0956797614523297

Cook, R., and Over, H. (2021). Why is the literature on first impressions so focused on white faces? *R. Soc. Open Sci.* 8:211146. doi: 10.1098/rsos.211146

Cuddy, A. J. C., Fiske, S. T., and Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Adv. Exp. Soc. Psychol.* 40, 61–149. doi: 10.1016/S0065-2601(07)00002-0

Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B., and Wu, C.-H. (1995). 'Their ideas of beauty are, on the whole, the same as ours': Consistency and variability in the cross-cultural perception of female physical attractiveness. *J. Pers. Soc. Psychol.* 68, 261–279. doi: 10.1037/0022-3514.68.2.261

Ewing, L., Sutherland, C. A. M., and Willis, M. L. (2019). Children show adult-like facial appearance biases when trusting others. *Dev. Psychol.* 55, 1694–1701. doi: 10.1037/dev0000747

Fiske, S. T., Cuddy, A. J. C., and Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends Cogn. Sci.* 11, 77–83. doi: 10.1016/j.tics.2006.11.005

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci* 33, 61–83. doi: 10.1017/S0140525X0999152X

Ichikawa, H., Kanazawa, S., and Yamaguchi, M. K. (2014). Infants recognize the subtle happiness expression. *Perception* 43, 235–248. doi: 10.1068/p7595

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., and Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Curr. Biol.* 19, 1543–1548. doi: 10.1016/j.cub.2009.07.051

Jessen, S., and Grossmann, T. (2016). Neural and behavioral evidence for infants' sensitivity to the trustworthiness of faces. *J. Cogn. Neurosci.* 28, 1728–1736. doi: 10.1162/jocn_a_00999

Jessen, S., and Grossmann, T. (2019). Neural evidence for the subliminal processing of facial trustworthiness in infancy. *Neuropsychologia* 126, 46–53. doi: 10.1016/j.neuropsychologia.2017.04.025

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., et al. (2009). Development of the other-race effect during infancy: Evidence toward universality? *J. Exp. Child Psychol.* 104, 105–114. doi: 10.1016/j.jecp.2009.01.006

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., and Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychol. Sci.* 18, 1084–1089. doi: 10.1111/j.1467-9280.2007.02029.x

Krys, K., Hansen, K., Xing, C., Szarota, P., and Yang, M. (2014). Do only fools smile at strangers? cultural differences in social perception of intelligence of smiling individuals. *J. Cross Cult. Psychol.* 45, 314–321. doi: 10.1177/0022022113513922

Lee, R., Flavell, J. C., Tipper, S. P., Cook, R., and Over, H. (2021). Spontaneous first impressions emerge from brief training. *Sci. Rep.* 11:15024. doi: 10.1038/s41598-021-94670-y

Matsumoto, D., and Kudoh, T. (1993). American-Japanese cultural differences in attributions of personality based on smiles. *J. Nonverbal Behav.* 17, 231–243. doi: 10.1007/BF00987239

Mondloch, C. J., Gerada, A., Proietti, V., and Nelson, N. L. (2019). The influence of subtle facial expressions on children's first impressions of trustworthiness and dominance is not adult-like. *J. Exp. Child Psychol.* 180, 19–38. doi: 10.1016/j.jecp.2018.12.002

Oosterhof, N. N., and Todorov, A. (2008). The functional basis of face evaluation. *Proc. Natl Acad. Sci. U.S.A.* 105, 11087–11092. doi: 10.1073/pnas.0805664105

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Champaign, IL: University of Illinois Press.

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., et al. (2005). Plasticity of face processing in infancy. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5297–5300. doi: 10.1073/pnas.0406627102

Rule, N. O., Ambady, N., Adams, R. B., Ozono, H., Nakashima, S., Yoshikawa, S., et al. (2010). Polling the face: Prediction and consensus across cultures. *J. Pers. Soc. Psychol.* 98, 1–15. doi: 10.1037/a0017673

Rule, N. O., Ishii, K., and Ambady, N. (2011). Cross-cultural impressions of leaders' faces: Consensus and predictive validity. *Int. J. Intercult. Relat.* 35, 833–841. doi: 10.1016/j.ijintrel.2011.06.001

Sakuta, Y., Ishi, H., Akamatsu, S., and Gyoba, J. (2009). Psychological evaluation of higher-order facial impressions synthesized by the impression transfer vector method. *Kansei Eng. Int. J.* 9, 1–10. doi: 10.5057/kei.9.1

Sakuta, Y., Kanazawa, S., and Yamaguchi, M. K. (2018). Infants prefer a trustworthy person: An early sign of social cognition in infants. *PLoS One* 13:e0203541. doi: 10.1371/journal.pone.0203541

Sutherland, C. A. M., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., and Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Pers. Soc. Psychol. Bull.* 44, 521–537. doi: 10.1177/0146167217744194

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., and Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition* 127, 105–118. doi: 10.1016/j.cognition.2012.12.001

Sutherland, C. A. M., Young, A. W., and Rhodes, G. (2017). Facial first impressions from another angle: How social judgements are influenced by changeable and invariant facial properties. *Br. J. Psychol.* 108, 397–415. doi: 10.1111/bjop.12206

Tang, Y., Harris, P. L., Zou, H., and Xu, Q. (2019). The impact of emotional expressions on children's trust judgments. *Cogn. Emot.* 33, 318–331. doi: 10.1080/02699931.2018.1449735

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Ann. N. Y. Acad. Sci.* 1124, 208–224. doi: 10.1196/annals.1440.012

Todorov, A., and Oosterhof, N. (2011). Modeling social perception of faces [Social Sciences]. *IEEE Signal Process. Mag.* 28, 117–122. doi: 10.1109/MSP.2010.940006

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., and Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion* 13, 724–738. doi: 10.1037/a0032335

Todorov, A., Mandisodza, A. N., Goren, A., and Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science* 308, 1623–1626. doi: 10.1126/science.1110589

Todorov, A., Pakrashi, M., and Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Soc. Cogn.* 27, 813–833. doi: 10.1521/soco.2009.27.6.813

Todorov, A., Said, C. P., Engell, A. D., and Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends Cogn. Sci.* 12, 455–460. doi: 10.1016/j.tics.2008.10.001

van 't Wout, M., and Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition* 108, 796–803. doi: 10.1016/j.cognition.2008.07.002

Van Der Zant, T., Reid, J., Mondloch, C. J., and Nelson, N. L. (2021). The influence of postural emotion cues on implicit trait judgements. *Motiv. Emot.* 45, 641–648. doi: 10.1007/s11031-021-09889-z

Verosky, S. C., and Todorov, A. (2013). When physical similarity matters: Mechanisms underlying affective learning generalization to the evaluation of novel faces. *J. Exp. Soc. Psychol.* 49, 661–669. doi: 10.1016/j.jesp.2013.02.004

Walker, M., Jiang, F., Vetter, T., and Sczesny, S. (2011). Universals and cultural differences in forming personality trait judgments from faces. *Soc. Psychol. Pers. Sci.* 2, 609–617. doi: 10.1177/1948550611402519

Willis, J., and Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychol. Sci.* 17, 592–598. doi: 10.1111/j.1467-9280.2006.01750.x

Zebrowitz, L. A. (2017). First impressions from faces. *Curr. Dir. Psychol. Sci.* 26, 237–242. doi: 10.1177/0963721416683996

Zebrowitz, L. A., Montepare, J. M., and Lee, H. K. (1993). They don't all look alike: Individual impressions of other racial groups. *J. Pers. Soc. Psychol.* 65, 85–101. doi: 10.1037//0022-3514.65.1.85

Zebrowitz, L. A., Wang, R., Bronstad, P. M., Eisenberg, D., Undurraga, E., Reyes-García, V., et al. (2012). First impressions from faces among U.S. and culturally isolated Tsimane' people in the Bolivian rainforest. *J. Cross Cult. Psychol.* 43, 119–134. doi: 10.1177/0022022111411386

frontiers | Frontiers in Psychology

TYPE Review
PUBLISHED 19 December 2022
DOI 10.3389/fpsyg.2022.988497

Check for updates

OPEN ACCESS

EDITED BY
Tilo Strobach,
Medical School Hamburg, Germany

REVIEWED BY
Quoc Vuong,
Newcastle University, United Kingdom
Marta Zimmer,
Budapest University of Technology
and Economics, Hungary

*CORRESPONDENCE
Kazusa Minemoto
k.minemoto@gmail.com
Yoshiyuki Ueda
ueda.yoshiyuki.3e@kyoto-u.jp

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

RECEIVED 07 July 2022
ACCEPTED 01 December 2022
PUBLISHED 19 December 2022

CITATION
Minemoto K and Ueda Y (2022) Face identity and facial expression representations with adaptation paradigms: New directions for potential applications.
Front. Psychol. 13:988497.
doi: 10.3389/fpsyg.2022.988497

Face identity and facial expression representations with adaptation paradigms: New directions for potential applications

Kazusa Minemoto* and Yoshiyuki Ueda*

Institute for the Future of Human Society, Kyoto University, Kyoto, Japan

Adaptation and aftereffect are well-known procedures for exploring our neural representation of visual stimuli. It has been reported that they occur in face identity, facial expressions, and low-level visual features. This method has two primary advantages. One is to reveal the common or shared process of faces, that is, the overlapped or discrete representation of face identities or facial expressions. The other is to investigate the coding system or theory of face processing that underlies the ability to recognize faces. This study aims to organize recent research to guide the reader into the field of face adaptation and its aftereffect and to suggest possible future expansions in the use of this paradigm. To achieve this, we reviewed the behavioral short-term aftereffect studies on face identity (i.e., who it is) and facial expressions (i.e., what expressions such as happiness and anger are expressed), and summarized their findings about the neural representation of faces. First, we summarize the basic characteristics of face aftereffects compared to simple visual features to clarify that facial aftereffects occur at a different stage and are not inherited or combinations of low-level visual features. Next, we introduce the norm-based coding hypothesis, which is one of the theories used to represent face identity and facial expressions, and adaptation is a commonly used procedure to examine this. Subsequently, we reviewed studies that applied this paradigm to immature or impaired face recognition (i.e., children and individuals with autism spectrum disorder or prosopagnosia) and examined the relationships between their poor recognition performance and representations. Moreover, we reviewed studies dealing with the representation of non-presented faces and social signals conveyed *via* faces and discussed that the face adaptation paradigm is also appropriate for these types of examinations. Finally, we summarize the research conducted to date and propose a new direction for the face adaptation paradigm.

KEYWORDS

face perception, face recognition, representation, aftereffect, norm-based coding, social message, development, autism spectrum disorder (ASD)

1 Introduction

The face is an important visual stimulus for our social life and it has attracted significant interest in the study of psychology. Various paradigms have been used to explore the perception and recognition of face identity and facial expressions. One useful tool is the adaptation and aftereffect method, also referred to as the "psychologist's microelectrode" (Frisby, 1979). Perception of a given object stimulus (target, S2) can be biased or impaired by another stimulus, presented before (S1) the target. *Aftereffect* is a change in perception of a sensory stimulus due to the viewing of S1. At the behavioral level, this phenomenon is brought by *forward masking* (inhibition)/*priming* (facilitation)/*adaptation* (biased perception) (Walther et al., 2013; Mueller et al., 2020 for a review). At the neural level, it is usually observed as signal (activation) reduction for repeated stimulus presentation. Moreover, a very extreme case of adaptation is the so-called repetition suppression, when the same stimulus is presented several times in a row (e.g., Grill-Spector et al., 2006; Schweinberger and Neumann, 2016 for a review). The typical adaptation experimental procedure called S1 as *adaptor* or *adaptation stimuli*, and S2 as *test stimuli*. In this procedure, participants were asked to keep looking at the adaptation stimulus, and to respond to the test stimuli with answers such as who they looked like in the pre-learned individual or which facial expression they expressed. For example, after viewing the adaption stimulus of an individual (e.g., Jim) or a facial expression (e.g., happy), the perception of participant is changed, resulting that they cannot recognize the test stimuli as Jim or happy expressions though they recognize the same stimuli as Jim or happy expressions before the adaptation.

Previously, the adaptation paradigm was used in low-level (or simple feature) perception of visual stimuli, such as perception of color (Webster, 1996) and tilt (Gibson and Radner, 1937); it was later applied to higher-level (or visually more complex stimulus) cognition concerning face perception: configuration of a face (Webster and MacLeod, 2011 for a review), face identity (e.g., Leopold et al., 2001; Rhodes and Jeffery, 2006), facial expression (e.g., Hsu and Young, 2004; Webster et al., 2004), gaze direction (e.g., Jenkins et al., 2006; Seyama and Nagayama, 2006; Clifford and Palmer, 2018 for a review), masculinity/femininity (e.g., Webster et al., 2004; Little et al., 2005; Jaquet and Rhodes, 2008), ethnicity (e.g., Webster et al., 2004), and attractiveness (e.g., Rhodes et al., 2003). Two major problems that must be solved using this method are discussed. One problem is to reveal the common (or shared) process of facial information. If two facial information are processed in the same process, based on the same neural representation, the prior presentation of the face could have an aftereffect on the recognition of the subsequent face. The second is to reveal the neural representations of the face. The latter issue has been examined under the hypothesis of norm-based coding proposed by Valentine (1991), in which faces are assumed to be represented in a mental space, centered on the average of all faces that each person has seen. Each face was identified based on its distance and direction from the center point in this space. According to this, adapting to a certain face recalibrates and temporally shifts the center point to the adaptation stimuli in the space, resulting in a change in subsequent perception (see the detailed discussion in section 3.1).

This study first overview what face adaptation studies (mainly using behavioral paradigms) revealed regarding the representation of face identity (i.e., who it is) and facial expression (e.g., happiness, anger). Then, beyond the scope of previous studies, we introduce the recent findings that can be revealed using the face adaptation paradigm, and discuss new potential applications of it. Specifically, in section 2, we summarize the basic procedures of facial adaptation and its aftereffects. Considering previous eminent reviews of face adaptation (Webster and MacLeod, 2011; Rhodes and Leopold, 2012; Strobach and Carbon, 2013; Mueller et al., 2020), recent findings and other information not presented therein have also been introduced. Some terms are used ambiguously, and we have noted this and redefined them. This section would help readers who have set foot in this field. In section 3, we review studies that have investigated facial representations of individual identity and facial expression, as well as studies that have expanded the scope of applicability of the facial adaptation paradigm. Using the face adaptation paradigms, some cognitive models concerning facial representation have been examined. We discuss what has revealed by face adaptation studies the face and facial expression recognition models (Ekman and Friesen, 1971; Bruce and Young, 1986; Valentine, 1991; Haxby et al., 2000) and what remains unclear yet. Furthermore, studies investigating children and participants with atypical face recognition (i.e., autism and prosopagnosia) are mentioned. Participants with atypical face recognition, in particular, have not been well-mentioned in the existing review articles. In section 4, we further expand the scope of the face adaptation paradigm and suggest that it is also useful for examining representations of non-presented faces (i.e., the mental images and ensemble average of faces) and social signals conveyed through faces. This section is a review of new findings in recent years, showing that the face adaptation paradigm can shed light on what is still unclear in representations related to face. In section 5, we discuss the representation of faces in the current stage and suggest the future applicability of the facial adaptation paradigm.

2 Basic characteristics of face aftereffects

First, we summarize the basic characteristics of face aftereffects to clarify the effectiveness of the adaptation method. In particular, aftereffects have been investigated in low-level visual features, and many studies have explored the relationship between stimuli and whether the two share common mechanisms.

2.1 Adaptation for high- and low-level visual processing

The face includes many low-level visual features (e.g., color, tilt, or shape). In the early face and facial expression adaptation research, there was a great focus on whether the face aftereffect was due to high-level visual processing specific to the face or the results of retinotopy or inheritance from aftereffects of low-level visual processing. One of the popular procedures to examine this is to change the positions or physical sizes of the adaptation and test stimuli. Previous research has reported that face aftereffects persist even when the aftereffects of low-level visual features collapse (Leopold et al., 2001; Rhodes et al., 2007 in face identity; Hsu and Young, 2004; Burton et al., 2016; Zamuner et al., 2017 in facial expressions).

In addition, studies using composite and hybrid faces have demonstrated that face adaptation differs from adaptation in low-level visual processing (Butler et al., 2008; Laurence and Hole, 2012). Face and facial expressions are related to both featural processing, such as eye and mouth, and configural (holistic) processing, such as the spatial arrangement of facial parts. Composite and hybrid faces were used to examine the effect of configural processing on the recognition of faces and facial expressions. A composite face is a photograph in which the top and bottom halves of a face are misaligned, and a hybrid face is a photograph of two different identities or expressions combined into one face (e.g., the top half expresses happiness and the bottom half expresses sadness). It is difficult to correctly recognize identity and facial expressions in these stimuli, although they have the same featural component as normal faces (Young et al., 1987; Calder et al., 2000). If the face aftereffect is based on low-level visual features, adaptation to a composite or hybrid face is expected to produce the same magnitude of aftereffects as does adaptation to normal faces. Butler et al. (2008) reported that a significant aftereffect was observed when participants adapted to normal and hybrid faces made of different images from the same facial expressions. However, an aftereffect was not observed when they adapted to hybrid faces made of different facial expressions. Moreover, Laurence and Hole (2012) reported that the composite faces in which participants could recognize the identity showed an aftereffect, while those in which participants could not recognize the identity did not show the aftereffect. These studies suggest that recognizability is an important factor in the face aftereffects.

However, the contribution of the aftereffect of low-level visual features cannot be denied as many studies have demonstrated it. The face aftereffect remains when the size and position of adaptation and test stimuli change, but it has been shown that the further away facial aftereffects are from the adaptation location, the weaker they become (Afraz and Cavanagh, 2008). In addition, only simple concave or convex curved lines and an isolated mouth from the real face are sufficient to cause the aftereffect of facial expressions, but these effects disappear when the presented position of adaptation and test are different (Xu et al., 2008). In addition, when the orientation of adaptation and test stimuli changed (the orientation of the adaptation and test stimuli rotated by 90°), the aftereffect of the expressions decreased (Swe et al., 2019). Note that the aftereffect remained when the test stimuli were rotated by 90°, suggesting that the orientation aftereffect could not fully explain the face aftereffect.

2.2 Temporal dynamics of face adaptation

After finding the face aftereffect, researchers were interested in its characteristics and found a commonality between the face aftereffect and low-level visual feature aftereffect.

With regard to temporal dynamics, there is a close relationship between the amount of face aftereffect and the presentation time: the increasing duration of the exposure to the adaptation stimuli builds up the aftereffect logarithmically, and the increasing duration of exposure to the test stimuli decays the aftereffect exponentially (Leopold et al., 2005; Rhodes et al., 2007; Burton et al., 2016). These effects were observed within a shorter adaptation period (e.g., 1 s). These dynamics occur in both face identity and facial expression aftereffects (Leopold et al., 2005; Rhodes et al., 2007; Burton et al., 2016). In addition, the time duration between adaptation and test stimuli (i.e., inter-stimulus intervals) also affects the size of the face aftereffect; that is, longer gaps lead to weaker aftereffects (Burton et al., 2016).

Temporal dynamics of face aftereffects suggest an answer to a question: As the faces are social stimuli, when the two faces are presented in succession, some meaning or context will alter the response to the test stimuli. If this is true, it would be expected that their responses would not be affected by the duration of either the presentation or gap durations between adaptation and test stimuli; however, the results showed their influence (Leopold et al., 2005; Rhodes et al., 2007; Burton et al., 2016). In summary, the change in response after prolonged viewing of the face is not the result of context or strategy due to the sequential presentation of the two faces, but is the result of the aftereffect.

Hsu and Young (2004) showed a facial expression aftereffect when the adaptation duration was 5,000 ms but not 500 ms, suggesting that a certain duration is necessary to produce the adaptation effect. However, it could not determine that 500 ms is the minimum length of exposure to elicit an aftereffect because it may change not only by adaptation duration but also by various factors, such as valence or its intensity of facial expressions or repetition of the adaptation stimulus. For example, aftereffects were observed by adaptation to angry expressions for 17 ms and happy expressions for 50 ms (Sou and Xu, 2019), and repeated presentation of a 500 ms adaptation face (Moriya et al., 2013). In addition, some neurophysiological studies have robustly reported the short-term adaptations (around 200 or 300 ms) reduce the face-sensitive neural activations of M170 in magnetoencephalographic (MEG) studies (Harris and Nakayama, 2007, 2008), N170 in event-related potentials (ERP) studies (Eimer et al., 2010; Nemrodov and Itier, 2011), and functional magnetic resonance imaging (fMRI) signal in fusiform cortex and posterior superior temporal sulcus (Winston et al., 2004).

Moreover, there are interaction between adaptation duration and position consistency of adaptation and test stimuli (Zimmer and Kovács, 2011 for a review). Kovács et al., 2005, 2007, 2008 investigated the N170 amplitude under conditions where the adaptation and test stimuli were presented at the same or different positions using an adaptation task of face gender. Results showed that the long-term (5,000 ms) adaptation duration induced the greater reduction in N170 amplitude when the adaptation and test stimuli were presented at the same location compared to the different locations, though no differences by location were observed for short-term (500 ms) adaptation duration. It suggests that the long-term face adaptation is position-specific, while the short-term face adaptation is position-invariant. The same results were reported in fMRI study and indicated that the activations of the right occipital face area reduced when the positions of the adaptation and test stimuli were the same after only long-term (4,500 ms) adaptation, but no differences were observed either when the positions of two stimuli were different or when adaptation duration was short (500 ms) (Kovács et al., 2008). On the other hand, the activations of the right fusiform face area reduced regardless the adaptation duration and the locations of two stimuli. These studies suggest that different adaptation durations are associated with different neural mechanisms.

3 What is revealed by face aftereffect?

Using the face adaptation "paradigms, some cognitive models concerning facial representation have been proposed. Here we introduce the famous face and facial expression models (Ekman and Friesen, 1971; Bruce and Young, 1986; Valentine, 1991; Haxby et al., 2000) and discuss what face aftereffect examined about them and what remains unclear yet.

3.1 Face representation

3.1.1 Face identity

Early experiments on face aftereffect were conducted using distorted images of the face created by a circular Gaussian envelope so that the face elements were expanding or contracting relative to a midpoint on the nose (Webster and MacLin, 1999). It was found that the face appeared biased toward the opposite direction of the preceding presented stimulus. For example, the original face appears to expand after adapting to contracting faces. In addition, this face aftereffect was beyond the mere distortion of visual objects because the face aftereffect decreased when the orientations of the adaptation and test faces were different (i.e., upright vs. inverted). Adaptation to the original (non-distorted) image did not change the face perception. These results suggest that our perception of the face was normalized by what we saw immediately before.

Subsequently, adaptation research has examined an important face recognition framework called "face space" (Valentine, 1991) by using anti-face images (Leopold et al., 2001; Rhodes and Leopold, 2012 for a review). Face space is considered a multidimensional mental space centered on the norm face, and each face is represented in this space (see **Figure 1**). In accordance with this idea, people represent each face identity to refer to the distance from the center of mental space (it is called "norm"[1]), which was investigated using anti-faces (e.g., Leopold et al., 2001; Rhodes and Jeffery, 2006). *Anti-face*s were generated by making a face with features that were the physically opposite of the original face to the average face of multiple faces using *morphing* techniques (here, the average of multiple faces can be regarded as the substitution of the norm). For example, identity A have smaller eyes than the average face while anti-face of A (described "anti-A") have bigger eyes than the average. Although the anti-face does not look like the original person from whom it was created, it lies on the same axis connecting the original face to the average face on the opposite side of the average face from the original face. This feature dimension through the original face, average (norm) face, and anti-face is referred to

1 The central point of the face space is referred as "average," "norm," and "prototype." Moreover, they are often interchangeably used (e.g., Rhodes et al., 1987; Valentine, 1991). Although it is better to use them differently based on the roles or functions of them, the differences among them seem not to be clearly defined yet. In this study, the following remarks are made to ensure that these terms are commonly used in studies of facial identity and facial expression. First, in the facial expression studies, the faces of neutral category are often used as "norm." "Average" is not appropriate in this field because the neutral faces are not made from other facial expressions. Second, "prototype" sometimes refers to typical facial expressions (e.g., Ekman et al., 1983). Therefore, the term "norm" generally refers to the center of the face space, whether it is used for identity or facial expression. In this study, "norm" is used to refer to this, and "average" is defined as the average over multiple faces using morphing techniques. "Prototype" does not used because it could be mistaken for its opposite meaning in facial expression studies.

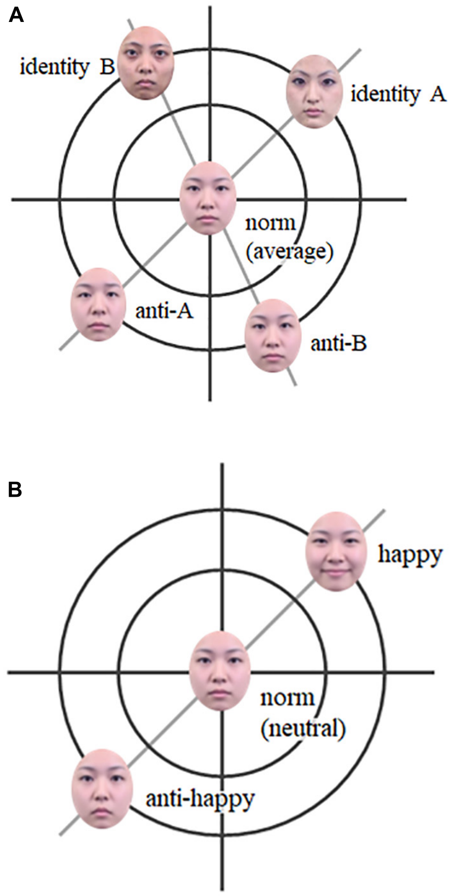


FIGURE 1
Schemas of the face-space. **(A)** Identity: two identities (identity A and B), their anti-faces (anti-A and anti-B), and norm (average). **(B)** Facial expression: a facial expression (happy), its anti-expression (anti-happy), and norm (neutral). The anti-faces were made from each identity or expression and norm (average or neutral) using morphing techniques.

as the *identity trajectory* (Leopold et al., 2001). Leopold et al. (2001) showed that after adapting to the anti-face for 5 s, the intensity of the features needed to identify each person (i.e., stimulus identity thresholds) decreased, and the average face was frequently identified as the original person of the anti-face. These aftereffects seem to be the result of the temporal shift of the central point of the face space toward an adapting face to fit the neural populations, which have a limited response range, depending on the current situation[2]

Furthermore, Rhodes and Jeffery (2006) examined the differences between the aftereffects of two types of anti-faces. One is the *matching* anti-face, which lies on the same trajectory as the test face, and the other is the *mismatching* anti-face, which does not lie on the same trajectory as the test face but on the trajectory of another face. The latter was equally perceptually dissimilar to the original face, but with a different identity trajectory[3]. The results showed that adaptation to the matching anti-face had a larger aftereffect on the recognition of the target face compared to adaptation to the mismatching anti-face. Thus, the aftereffect of the face is a selective bias of the central point toward the adapting face in the face space, but not the simple contrast effect between two faces (see **Figure 1**; Rhodes and Jeffery, 2006).

In summary, the results of the identity aftereffect suggest that perception of faces could be based on the face space in which people refer to the direction and distance between the norm and the individual face. In addition, this face space is temporally and immediately recalibrated with what we see, even for seconds, and it is quite likely that this flexible updating occurs in real life. Accordingly, the aftereffect is not limited to the picture and occurs even when adapting to the video clip of a face (Petrovski et al., 2018).

3.1.2 Facial expressions

Following the studies of the identity aftereffect, adaptation studies have focused on facial expressions to reveal their representation, in which two theories have long been the center of controversy. One is the categorical theory, which states that facial expressions are represented as discrete qualitative categories. The idea is led by Ekman and Friesen (1971) and assumes that there are six basic categories of facial expressions (anger, fear, disgust, happy, sad, and surprise) that are innate and universal and are often used in studies of facial expressions. The other is the continuous theory that argues facial expressions are represented in the circumplex model (Russell, 1980; Russell and Bullock, 1985). The main dimensions of this model are valence (unpleasant-pleasant) and arousal (high arousal-low arousal or sleepiness). In the latter theory, the boundaries of the categories are ambiguous and the category can change depending on the context in which the expression is presented (Carroll and Russell, 1996).

To investigate which of the two hypotheses is appropriate, the literature used the same and different categories of expressions as the adaptation and test stimuli. For the aftereffect between the same category of adaptation and test facial expressions, there is growing consensus that prolonged presentation of facial expressions leads to strong selective aftereffect (Hsu and Young, 2004; Webster et al., 2004; Juricevic and Webster, 2012), in which after adapting to happy facial expressions, the more intensely happy faces are needed to

2 Since identity A and anti-A are in the opposite direction across the average in the face space, the results after adapting to identity A and anti-A are also the opposite. For example, adapting to identity A makes the average look less like A though adapting to anti-A makes the average look more like A.

3 Matching anti-face and mismatching anti-face were referred to as opposite face and non-opposite face, respectively, by Rhodes and Jeffery (2006), but we use the term "anti" in this paper for consistency.

perceive happiness compared to before the adaptation (Hsu and Young, 2004; Webster et al., 2004; Juricevic and Webster, 2012). Likewise, after adapting to anti-happy facial expressions, participants can perceive happiness, even for less happy expressions (Skinner and Benton, 2010; Juricevic and Webster, 2012). On the other hand, for the aftereffect between the different categories of adaptation and test stimuli, some studies performed it but did not reach a consensus on their results. In Experiment 1 of Hsu and Young (2004), the facilitation aftereffect on the recognition of sad facial expressions (i.e., participants can perceive sadness even for less sad expressions) after adapting to happy facial expressions was presented, but this aftereffect was not found in Experiment 2, which employed the same procedure but a different identity face set. Juricevic and Webster (2012) found that adaptation to any of the six basic emotion categories did not facilitate the recognition of any category. The results showed that robust aftereffects were observed only in combinations of the same facial expressions, indicating that people could process facial expressions categorically. However, the aftereffect was observed in some combinations of different facial expressions, suggesting that the categories of facial expressions are not completely independent. Considering these studies, it is suggested that the categories of facial expressions are semantically independent, but their neural representations somewhat overlap.

Although the adaptation paradigm was expected to shed light on the discrete process, it was difficult to make a clear consensus between the two hypotheses because the aftereffects were found in both the same and different categories between the adaptation and test facial expressions (Hsu and Young, 2004; Rutherford et al., 2008; Pell and Richards, 2011; Juricevic and Webster, 2012). Therefore, the aim of this study shifted from showing which of the two hypotheses is correct to examining the representation of facial expressions, incorporating the concept of face space.

To elaborate on the representation of facial expressions, Cook et al. (2011) conducted an experiment based on the idea of Rhodes and Jeffery (2006), as mentioned in subsection 3.1.1. They used the anti-facial expression and the original facial expression as a pair (A and B) and the orthogonal faces pair (C and D), whose features are in the orthogonal direction against an A and B pair in the facial expression space. The results showed that selectivity aftereffects were observed when the adaptation stimulus was congruent with the test stimulus (e.g., the adaptation stimulus was either A or B, and the test stimulus was one of AB continua). When incongruent pairs were presented as the adaptation and test stimuli (e.g., the adaptation stimulus was either A or B, and the test stimulus was one of CD continua), no or fewer aftereffects were observed. These results suggest that facial expressions are represented by multidimensional facial spaces.

In addition, it has been reported that adaptation stimuli causing stronger aftereffects are associated with more intense facial expressions (Skinner and Benton, 2010; Burton et al., 2015; Rhodes et al., 2017; Hong and Yoon, 2018). These results suggest that we do not perceive facial expressions regardless of their intensity (if so, then an adaptation to a facial expression should always lead to a constant intense aftereffect), but that the perceptual bias due to adaptation shifts gradually as a function of the intensity of the adaptation stimuli. These results also support the norm-based coding of facial expressions.

3.1.3 Relationship between identity and facial expression

As we can recognize facial expressions even if we do not know who the person is, earlier face models suggest that face identity and facial expression are processed distinctively (Bruce and Young, 1986; Haxby et al., 2000). However, some recent studies have argued that they are not independent of each other (Kaufmann and Schweinberger, 2004; Winston et al., 2004; Calder and Young, 2005). The adaptation paradigm is a good way to examine this using the same and different models of the same facial expression. Simultaneously, it is also possible to examine whether face identity and facial expressions can be mapped onto the same mental space. To investigate these questions, Fox and Barton (2007) compared the aftereffect with four combinations of adaptation and test stimuli: two types were from the same model with the same and different images, and the other two types were from different models with the same or different genders displaying facial expressions. The results showed that all combinations led to the same direction of the aftereffect, but the size of the aftereffect was smaller in the different model conditions than in the same model with the same image conditions. To determine whether differences in the aftereffect size were due to differences in image or model identity, they further compared conditions in which the same and different images of the same model showed comparable aftereffects across these conditions. The same patterns have been observed across different emotional expressions (Campbell and Burke, 2009; Pell and Richards, 2013). Ellamil et al. (2008) examined this aftereffect using faces that had the same shape (featural patterns) but were wrapped with the surfaces of different models. Although the featural pattern was the same, the size of the aftereffect decreased when the wrapped model differed from the adaptation model. These results suggest that there are identity-independent and identity-dependent representations of facial expressions. In addition, Song et al. (2015) reported that when adaptation and test stimuli had the same identity and expression configurations, the aftereffect was larger than when adaptation and test stimuli had the same identity but different expression configurations. By contrast, when their identities were different, the aftereffect was comparable, regardless of the expression configuration. This indicates that sensitivities to expression configurations were different between the identity-dependent and independent representations.

Furthermore, a positive correlation was observed between face identity aftereffect and facial expression aftereffect, while these aftereffects did not correlate with other features of the face (gaze direction) and the aftereffect using orientation stimuli (i.e., not face, but a geometric feature) (Rhodes et al., 2015). Their results suggest that there is a common representation of identity and expression. However, Rhodes et al. (2015) reported that the recognition ability of identity and expressions were predicted by identity aftereffects regardless of expression and expression aftereffects regardless of identity, suggesting identity-selective or expression-selective dimensions.

In summary, most studies have agreed that there are two components of face adaptation: one is the identity-independent representation, which results in the aftereffect regardless of model, and the other is the identity-dependent representation, which results in a larger aftereffect when adapting to the same model rather than a different model. These two components are represented as a multi-dimensional space because anti-expressions led to the aftereffect of whether the models in the adaptation and test stimuli were the same or different, and the weaker adaptation stimuli led to smaller aftereffects (Skinner and Benton, 2012). Interestingly, asymmetric results were observed in an identity recognition study; Fox et al. (2008) demonstrated that the identity aftereffect did not change depending on whether the conditions of adaptation and test stimuli were the same or different facial expressions. This suggests that identity can be independently represented by a facial expression component. As few studies have investigated this asymmetric effect, it is important to modify and revise the model of face recognition in the future.

3.2 Development and impairment investigated using face adaptation

People are experts at recognizing and discriminating between face identity and facial expressions. The norm-based representation supports this skill, as reviewed in the previous section, and some research has reported that there is a correlation between the recognition ability and the size of the aftereffect (Rhodes et al., 2014b, 2015; Palermo et al., 2018). In this section, we review this face representation in people who are not as proficient as healthy adults, such as children or people who have impaired face recognition function. As developmental research on identity aftereffects was reviewed by Jeffery and Rhodes (2011), we summarize it simply and focus on the facial expressions and interaction of face identity and facial expressions in this section.

3.2.1 Development of norm-based recognition

Face-recognition ability develops according to age (Vicari et al., 2000; Mondloch et al., 2003). An adaptation study has explored the developmental improvement in norm-based coding, and most studies have shown the same patterns of aftereffect in children aged 4–10 years as well as adults (Jeffery and Rhodes, 2011 for a review). Further, matching anti-face leads to larger aftereffects than mismatching anti-face even when matching anti-face and mismatching anti-face had the same perceptual dissimilarity, and the more intense adaptation stimuli led to larger aftereffects, suggesting that children over 4 years old may have multidimensional face space (Nishimura et al., 2008 with 8 years of age; Jeffery et al., 2010 with 4–6 years of age; Jeffery et al., 2011 with 5–9 years of age; Jeffery et al., 2013 with 4 years of age).

To our knowledge, only two studies have examined the development of aftereffects of facial expressions (Vida and Mondloch, 2009; Burton et al., 2013). Vida and Mondloch (2009) reported that the aftereffects of facial expressions were observed in the children's group, but this depended on the pairs' expressions (i.e., happy/sad or fear/anger). While children aged 5 years did not show adult-like aftereffects in happy/sad pairs, children aged 7 years showed adult-like aftereffects in these pairs. However, children aged 7 years did not show adult-like aftereffects in fear/angry pairs, whereas children aged 9 years showed adult-like aftereffects in these pairs. Considering that a previous study indicated that the development of sensitivity to fear and anger was slower than happiness and sadness (Vicari et al., 2000), the results suggest that the aftereffect of facial expression might depend on the developmental sensitivity of facial expressions. Burton et al. (2013) investigated whether the representation of facial expressions in 9 years-old children is norm-based or not, using anti-facial expressions with two types of strength. They found that the aftereffect occurred after adaptation to anti-facial expressions, and stronger adaptation led to a larger impact of aftereffects. These results supported the idea that children of these ages could represent facial expressions in multidimensional facial space as well as adults.

Studies with adults have suggested that there are identity-dependent and identity-independent components for the representation of facial expressions, while identity representation is independent of facial expressions (Fox and Barton, 2007; Fox et al., 2008; Skinner and Benton, 2012). Likewise, children may have the same two types of components because adaptation to the same identity as the test stimuli resulted in a larger aftereffect rather than adaptation to a different identity from the test stimuli (Vida and Mondloch, 2009). In addition, the identity aftereffect was not affected by whether the adaptation and test stimuli had the same or different facial expressions as 8 years-old children (Mian and Mondloch, 2012), showing an asymmetric representation between identity and facial expression.

Interestingly, an adult-like identity aftereffect was observed at age 4 years (Jeffery et al., 2013), but an adult-like facial expression aftereffect was observed over 7 years of age (Vida and Mondloch, 2009). However, it is premature to discuss this because there were different points in the stimuli used in these studies, such as whether they changed the model or if they used an anti-face. To achieve a deeper understanding of the development of facial expression and identity and to consider their relationships with each other, more research on the aftereffects of facial expressions is needed, such as experiments conducted under the same procedure for identity and facial expression, younger participants, or other expression categories.

3.2.2 Impairment face recognition and aftereffect

Many types of atypical social communication have been reported, among which autism spectrum disorder (ASD) is a major cause of deficits in facial recognition (Harms et al., 2010; Weigelt et al., 2012). Adaptation paradigms have been used to examine norm-based representations. For face identity, studies have shown that adapting to the anti-face leads to an aftereffect in both autism and typical development (TD) groups (Pellicano et al., 2007; Rhodes et al., 2014a; Walsh et al., 2015), and it becomes stronger as the intensity (i.e., the distance from the norm) of the face increases (Rhodes et al., 2014a; Walsh et al., 2015). This suggests that a norm-based facial recognition system can be implemented for people with ASD. Moreover, studies have also shown that the aftereffect size was smaller in children with ASD than in those with TD (Pellicano et al., 2007 on ages 8–13; Rhodes et al., 2014a on ages 9–14). This effect was comparable between healthy adults and adults with ASD (Cook et al., 2014; Walsh et al., 2015). In summary, these results indicate that there are no large qualitative differences in norm-based identity representation between adults with TD and ASD, suggesting that the deficits in identity recognition in adults with ASD are not due to perceptual representations. However, there are differences between children with TD and ASD, suggesting that children with ASD are slower to become proficient in norm-based representations than those with TD.

Likewise, for facial expressions, adaptation to facial expressions led to an aftereffect, in which recognition of the original facial expressions could be facilitated, even in people with ASD (Rhodes et al., 2018; Hudac et al., 2021). The stronger intensity of adaptation stimuli also led to a larger aftereffect (Rhodes et al., 2018), suggesting that ASD could represent facial expressions in a norm-based manner. In addition, a smaller aftereffect size was found in children with ASD (Rhodes et al., 2018), although no difference was found in adults (Rutherford et al., 2012; Cook et al., 2014). Rutherford et al. (2012) reported that the response patterns after facial expression adaptation differed between individuals with ASD and TD. After adapting to negative emotions (e.g., fear), ASD participants tended to choose a sad label for the test stimulus of neutral faces, although TD participants tended to choose a happy label for them. These results suggest that participants with ASD encode facial expressions in a different mental facial expression-space than those with TD, in which negative and positive expressions are not opposites on the same axis. As there is some debate on the relationship between the categories of facial expressions, this point needs to be further investigated in the future.

Prosopagnosia is another well-known neuropsychological disorder that causes deficits in face recognition but has normal intelligence, memory, and low-level vision. The adaptation paradigm has also been used to reveal face coding systems and representations in this group. It is known that there are different types of prosopagnosia depending on the different causes of symptoms and impaired cognitive processes. The former is the *congenital prosopagnosia* (CP, also known as developmental prosopagnosia), who had no known brain injury but had difficulty recognizing face by nature and the *acquired prosopagnosia* (AP), who impaired their ability of face recognition due to acquired brain damage. The latter is the *apperceptive* and the *associative prosopagnosia*, which are dysfunctions of face recognition processes (De Renzi et al., 1991). It is suggested that each of those types is associated with a different cognitive stage of Bruce and Young (1986)'s model, which describes from face perception to name identification and separates the distinctive face cognitive processes into multiple stages (Corrow et al., 2016 for a review). The *apperceptive prosopagnosia* is impaired the structural encoding, which is the first stage in Bruce and Young's model, resulting the failure of face perception. On the other hand, the *associative prosopagnosia* is impaired the face recognition units, the second stage in their model, resulting the impaired sense of familiarity and recall to the familiar faces though they can accurately perceive the facial structure.

Most adaptation studies for the prosopagnosia had been conducted for CP mainly. The pattern of identity aftereffects of CP was the same as that of the control participants, that is, adapting to anti-face enhanced identification of the original identity (Nishimura et al., 2010; Susilo et al., 2010; Palermo et al., 2011). However, Palermo et al. (2011) found the difference between CP and control participants for the response to the average face: After adapting to the anti-face, the controls regarded the average face as the original (opposite of anti-face) faces, but the response of CP was chance level. Moreover, the groups appeared to differ in discrimination precision, indicating that the controls had more precise discrimination. These results suggest that CP does not make identity judgments in the same way as the controls although CP they could discriminate between identities to some extent. Face space of them were more coarse (Nishimura et al., 2010), or based on high-level object coding mechanisms that are not specific to faces (Palermo et al., 2011). Nishimura et al. (2010) examined the identity aftereffect of not only CP but also AP, who was the only one of their participants, and reported that his/her performance was dissimilar from those of control and CP participants. Particularly, he/she showed no systematic response (i.e., responses did not fit to sigmoidal curves) according to identity intensity. So far, to our knowledge, there are few studies examining difference in face adaptation between AP and CP, and no studies between the *apperceptive* and the *associative prosopagnosia*. As we have seen, face adaptation is one of the useful paradigms for examining facial representations, and more research is needed in the future on people with atypical face recognition.

4 Beyond visually presented face

4.1 Representation of non-presented face

An adaptation study revealed that non-existent visual stimuli, such as imagery, cause aftereffects as well as the face in reality. This means that there is a common neural representation that is activated by both visual and mental (imagery) faces in high-level face-perception processes. For this topic, we focus on the study of mental imagery and the ensemble of facial expressions.

We can create vivid imagery even if the visual stimuli do not exist in front of us, and it has been reported that mental imagery and perception of visual stimuli activate the same brain regions (O'Craven and Kanwisher, 2000). Consistent with these results, adaptation to mental images of faces and facial expressions induced the same pattern of aftereffects as adaptation to real visual stimuli (Ryu et al., 2008; Zamuner et al., 2017). In these studies, participants were asked to associate the models' face identities with their names (Ryu et al., 2008) or to memorize pictures of a model expressing six basic emotions (Zamuner et al., 2017). Then, participants adapted to the realistic face or vividly visualized these faces (i.e., adapted to non-presented faces). Ryu et al. (2008) used matching anti-faces, mismatching anti-faces, and their imageries as adaptation stimuli. After adapting to matching anti-faces or vividly visualizing their faces, the intensity of the features needed to identify each person decreased compared to the control condition (in which they were not adapted to or visualized faces) and adaptation to mismatching anti-face conditions. Similarly, Zamuner et al. (2017) used a person with six basic facial expressions or imageries as adaptation stimuli, and the same facial expressions as adaptation stimuli were presented as test stimuli. After adapting to facial expressions or their images, recognition performance decreased compared with the control condition. These results indicate that the real face and imagined face shared a common representation. In addition, the results also showed that the size of the aftereffect by real faces was greater than that by imagery faces, except for surprised facial expressions. Taken together with the finding that the real face and imagery face activated the same brain region, but the real face was strongly activated (O'Craven and Kanwisher, 2000), it is suggested that the size of the aftereffect predicts the extent to which the face engages a particular neural region.

However, studies on adaptation to the sex of faces have shown inconsistent results (DeBruine et al., 2010; D'Ascenzo et al., 2014). D'Ascenzo et al. (2014) used three male and female faces as adaptation stimuli, and androgynous faces were created by morphing female and male faces as test stimuli. Participants were asked to rate masculinity or femininity by moving a slider on a scale that labeled masculinity or femininity at both ends. They reported that adapting to real faces resulted in a similar pattern of adaptation in a previous study: female judgment decreased more after adapting to female faces than to male faces. In contrast, adapting to the imagined face had the opposite pattern: female judgment increased more after adapting to female faces than male faces. They discussed the inconsistent results with Ryu et al. (2008) and suggested that different face properties in different processing evoked varied aftereffects. Based on this suggestion, it is possible that investigating the direction or strength of the aftereffect could reveal the different processing of various components of faces in imagery, and the distinction of the representation of real and imagery faces.

It is known that we can extract the average information from multiple visual stimuli automatically and rapidly, which is called the ensemble average, and it has been reported using faces (De Fockert and Wolfenstein, 2009) and facial expressions (Haberman and Whitney, 2007, 2009). This ensemble helps us understand the surrounding environment at a glance. There are two types of ensemble: the temporal statistical ensemble, in which the extracted visual stimulus is presented sequentially one by one at a time, and the spatial statistical ensemble, which involves extracting multiple visual stimuli presented simultaneously. It is important to note that the ensemble average extracted from the face groups is not necessarily the presented face. For facial expressions, adapting to sequentially or spatially presented multiple facial expressions showed the same pattern of aftereffects as when adapting to faces of the same intensity as the average of those stimuli (Ying and Xu, 2017; Minemoto et al., 2022b). Both studies have used different individuals with facial expressions as adaptation stimuli, so the average was a different individual from each model and looked more similar to morphed faces with 35 models used as test stimuli. As the size of the aftereffect was very small when the adaptation stimuli were an emotional voice or a dog's emotional posture, but not human facial expressions (Fox and Barton, 2007), the results suggest that the ensemble average can be represented visually, and it shares a common neural representation with a real face.

To compare the adaptation to ensemble average and real faces, Ying and Xu (2017), Minemoto et al. (2022a) used a morphed averaged face with adaptation stimuli and a model with the averaged intensity of facial expressions as adaptation stimuli and showed that the sizes of adaptations to the averaged face and ensemble average were comparable. Considering that imagery adaptation leads to weaker aftereffects (Ryu et al., 2008; Zamuner et al., 2017), these results suggest that imagery and ensemble representations may differ in intensity even though they share the same neural mechanism.

4.2 Social message and signals of face

A face carries rich information, and they are important to building social relationships because we can recognize the personality traits, inner states, or surroundings of others based on them (Zebrowitz, 1997; Winston et al., 2002; Ekman, 2012). Thus far, we have reviewed studies on the perception of face, and finally considered the social messages or signals that were conveyed.

Previous studies have reported that adaptation also occurs with social information such as trustworthiness (Engell et al., 2010; Wincenciak et al., 2013), friendliness (Prete et al., 2018), physical strength, dominance (Witham et al., 2021), and helping judgment (Minemoto et al., 2022a). Both trustworthiness and friendliness correlate with the perception of happy and angry facial expressions: happy facial expressions are associated with trustworthiness and high friendliness, and angry facial expressions are associated with untrustworthiness and threat (Oosterhof and Todorov, 2009; Ekman, 2012). Adaptation to angry faces increases trustworthiness and greater friendliness judgments to a subsequently neutral test face rather than adaptation to happy faces (Engell et al., 2010; Prete et al., 2018). Engell et al. (2010) also showed that this trustworthiness aftereffect remained when they used different sizes of test stimuli from the adaptation stimulus (i.e., 80%), and the strength of the aftereffect was influenced by the adaptation duration. Witham et al. (2021) showed the same patterns of aftereffects for perceived physical strength and dominance. After adapting to anti-angry facial expressions, the face with average expressions of six basic emotions and a neutral emotion appeared physically stronger and more dominant, although adaptation to anti-fearful facial expressions had opposite aftereffects (i.e., physically weaker and less dominant). These results suggest that trustworthiness, friendliness, physical strength, and dominance rely on the same or partially overlapping neural mechanisms involved in the perception of facial expressions. In addition, Witham et al. (2021) reported that adaptation to anti-happy faces showed a small but similar directed aftereffect to adaptation to anti-angry faces, suggesting that the function to enhance perceptual strength could not be specific to one facial expression category.

Wincenciak et al. (2013) investigated the effect of adaptation to more and less trustworthy neutral faces on trustworthiness judgments and showed that only female participants were affected, whereas males were not. The aftereffect observed in female participants had the typical characteristic of other face aftereffects: the neutral faces were rated untrustworthy after adapting to trustworthy faces, while they were rated trustworthy after adapting to untrustworthy faces. Wincenciak et al. (2013) noted that the previous study (Stirrat and Perrett, 2010) indicated that male observers were less influenced by visual information of the face, such as width-to-height ratio, which is linked with testosterone and is predictive of aggression, than female observers. Moreover, females with more subordinate traits are more influenced by the width-to-height ratio. These results suggest that different factors and processes are involved between men and women in the perception of trustworthiness, and that the adaptation paradigm may be useful for examining the process of social messages.

Interestingly, two different facial expressions may relate to the same social message. For example, it has been reported that the perception of sad and fearful facial expressions induces prosocial behaviors in helping judgment (Marsh et al., 2005, 2007; Ekman, 2012). Minemoto et al. (2022a) reported that adapting to persons with sad facial expressions reduced their perception of the need for help (e.g., how much participants thought the person needed help) for both those with sad and fearful facial expressions, whereas adapting to those with fearful facial expressions reduced it only for those with fearful facial expressions. Considering that adaptation to facial expressions consistently reduces the perception of the same expressions (Hsu and Young, 2004; Juricevic and Webster, 2012), these results indicate that adapting to sad facial expressions influences not only facial expression perception but also social signal processing. Given that adaptation and aftereffects occur automatically, people automatically perceive the need for help when they see sad facial expressions. By contrast, fearful facial expressions do not have this function, although both expressions give observers the impression of the need for help.

Currently, few adaptation studies have focused on social messages. However, as presented here, the adaptation paradigm can be a useful tool for investigating what signals we automatically process when we see faces and what is based on our social judgments.

5 Discussion: The future of the adaptation paradigm

Face identity and facial expressions are important social cues for communicating and establishing social relationships with others, and the adaptation paradigm is a good procedure to examine their process and representation. In this study, we reviewed behavioral studies on the aftereffects of face identity and facial expressions from the basic characteristics that have been observed in various studies to the still discussing topics.

Early studies using adaptation, with typical adult participants using realistic face stimuli, have provided widespread support for the idea that norm-based coding is used for face recognition processes (Leopold et al., 2001; Juricevic and Webster, 2012). Subsequent studies have provided a distinct or overlapping relationship between identity and facial expressions, which has already been suggested (Bruce and Young, 1986; Haxby et al., 2000; Calder and Young, 2005), by using the validity of facial models (Fox and Barton, 2007; Fox et al., 2008).

Studies on participants with immature or impaired face recognition (child, ASD, and prosopagnosia) reported that aftereffects were observed, suggesting that processes other than norm-based coding may be responsible for their atypical face processing (Burton et al., 2013; Jeffery et al., 2013). However, there are some problems with this topic. First, few studies have explored the aftereffects of facial expression in children. As Vida and Mondloch (2009) suggested, different developments were observed in different category boundaries, and other types of category boundaries of facial expressions need to be examined to understand the development of facial expressions. Second, ASD studies have proposed that ASD has an atypical relationship between the categories of facial expressions compared to TD, and the findings of the prosopagnosia study suggest that CP has an atypical face space. Further research is needed to explore the representation of facial expressions of people with deficits in facial recognition.

There has been a growing body of aftereffect research on information related to non-presented faces, such as imagery faces (Ryu et al., 2008; Zamuner et al., 2017), face ensemble averages (Ying and Xu, 2017; Minemoto et al., 2022b), and the social messages or perceived personality traits of facial expressions (Engell et al., 2010; Wincenciak et al., 2013; Prete et al., 2018; Witham et al., 2021; Minemoto et al., 2022a). The results indicated that non-presented faces also induced the aftereffect, and they showed basically the same pattern as a realistic face. In addition, social messages and perceived personality traits affected adaptation to facial expressions. This topic is one of the future directions for adaptation studies: investigating social messages or perceived personality through the face (hereinafter referred to as the "social roles of the face"). The social role of the face is essential for building relationships with others. Despite this, previous adaptation studies have mainly shed light on the characteristics and representations of faces and have not yet examined much of the representations and cognitive functions of the impressions we receive from faces. To examine its cognitive processes, an adaptation paradigm can reveal the function of various types of facial information (e.g., anger is related to trustworthiness). As there are so many different types of social roles of the face, related studies are still limited. As reported in Section 4.2, we would say that this topic has a wide range of unexamined aspects. If the recognition of the social roles of faces is not dependent on the perception of facial expressions, then there may be a group that is accurate in the perception of facial expressions but is unable to recognize the social roles of faces and struggle with it. Therefore, this topic will need to continue to be considered in the future.

Finally, we recapitulated the advantage of the method of adaptation compared with the direct response to stimuli. The adaptation paradigm can eliminate unexpected factors because the same test stimuli and tasks were used both before and after the adaptation phases. Specifically, task demands are often inferred when participants respond to stimuli, and this is more likely to occur in face research because faces are strongly social. The adaptation paradigm may avoid this serious issue by examining the shifted responses before and after adaptation; that is, we could investigate the difference regardless of the participants' attitude. This advantage is particularly useful when considering social messages in which the task demand is easily guessed.

Recently, we have been able to consider that face adaptation studies have approximately reached the stage of revealing the basic features of facial representations. However, as discussed in this review, the potential for new applications of the face aftereffect remains open.

Author contributions

KM and YU conceptualized the review. KM was primarily responsible for the article research and drafted the original manuscript. YU revised and supervised the manuscript. Both authors approved the final version of the manuscript.

Funding

This work was supported by the JSPS KAKENHI Grant-in-Aid for Scientific Research (A) Grant Number 19H00628 and Scientific Research on Innovative Areas Grant Number 20H04577 *via* YU.

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

Afraz, S. R., and Cavanagh, P. (2008). Retinotopy of the face aftereffect. *Vision Res.* 48, 42–54. doi: 10.1016/j.visres.2007.10.028

Bruce, V., and Young, A. (1986). Understanding face recognition. *Br. J. Psychol.* 77, 305–327. doi: 10.1111/j.2044-8295.1986.tb02199.x

Burton, N., Jeffery, L., Bonner, J., and Rhodes, G. (2016). The timecourse of expression aftereffects. *J. Vis.* 16:1. doi: 10.1167/16.15.1

Burton, N., Jeffery, L., Calder, A. J., and Rhodes, G. (2015). How is facial expression coded? *J. Vis.* 15, 1–13. doi: 10.1167/15.1.1

Burton, N., Jeffery, L., Skinner, A. L., Benton, C. P., and Rhodes, G. (2013). Nine-year-old children use norm-based coding to visually represent facial expression. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 1261–1269. doi: 10.1037/a0031117

Butler, A., Oruç, I., Fox, C. J., and Barton, J. J. S. (2008). Factors contributing to the adaptation aftereffects of facial expression. *Brain Res.* 1191, 116–126. doi: 10.1016/j.brainres.2007.10.101

Calder, A. J., and Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nat. Rev. Neurosci.* 6, 641–651. doi: 10.1038/nrn1724

Calder, A. J., Keane, J., Young, A. W., and Dean, M. (2000). Configural information in facial expression perception. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 527–551. doi: 10.1037/0096-1523.26.2.527

Campbell, J., and Burke, D. (2009). Evidence that identity-dependent and identity-independent neural populations are recruited in the perception of five basic emotional facial expressions. *Vision Res.* 49, 1532–1540. doi: 10.1016/j.visres.2009.03.009

Carroll, J. M., and Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *J. Pers. Soc. Psychol.* 70, 205–218. doi: 10.1037/0022-3514.70.2.205

Clifford, C. W. G., and Palmer, C. J. (2018). Adaptation to the direction of others' gaze: A review. *Front. Psychol.* 9:2165. doi: 10.3389/fpsyg.2018.02165

Cook, R., Brewer, R., Shah, P., and Bird, G. (2014). Intact facial adaptation in autistic adults. *Autism Res.* 7, 481–490. doi: 10.1002/aur.1381

Cook, R., Matei, M., and Johnston, A. (2011). Exploring expression space: Adaptation to orthogonal and anti-expressions. *J. Vis.* 11, 1–9. doi: 10.1167/11.4.2

Corrow, S. L., Dalrymple, K. A., and Barton, J. J. S. (2016). Prosopagnosia: Current perspectives. *Eye Brain* 8, 165–175. doi: 10.2147/EB.S92838

D'Ascenzo, S., Tommasi, L., and Laeng, B. (2014). Imagining sex and adapting to it: Different aftereffects after perceiving versus imagining faces. *Vision Res.* 96, 45–52. doi: 10.1016/j.visres.2014.01.002

De Fockert, J., and Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Q. J. Exp. Psychol.* 62, 1716–1722. doi: 10.1080/17470210902811249

De Renzi, E., Faglioni, P., Grossi, D., and Nichelli, P. (1991). Apperceptive and associative forms of prosopagnosia. *Cortex* 27, 213–221. doi: 10.1016/S0010-9452(13)80125-6

DeBruine, L. M., Welling, L. L. M., Jones, B. C., and Little, A. C. (2010). Opposite effects of visual versus imagined presentation of faces on subsequent sex perception. *Vis. cogn.* 18, 816–828. doi: 10.1080/13506281003691357

Eimer, M., Kiss, M., and Nicholas, S. (2010). Response profile of the face-sensitive N170 component: A rapid adaptation study. *Cereb. Cortex* 20, 2442–2452. doi: 10.1093/cercor/bhp312

Ekman, P. (2012). *Emotions revealed: Understanding faces and feelings*. London: Weidenfeld & Nicolson.

Ekman, P., and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *J. Pers. Soc. Psychol.* 17, 124–129. doi: 10.1037/h0030377

Ekman, P., Levenson, R. W., and Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science* 221, 1208–1210. doi: 10.1126/science.6612338

Ellamil, M., Susskind, J. M., and Anderson, A. K. (2008). Examinations of identity invariance in facial expression adaptation. *Cogn. Affect. Behav. Neurosci.* 8, 273–281. doi: 10.3758/CABN.8.3.273

Engell, A. D., Todorov, A., and Haxby, J. V. (2010). Common neural mechanisms for the evaluation of facial trustworthiness and emotional expressions as revealed by behavioral adaptation. *Perception* 39, 931–941. doi: 10.1068/p6633

Fox, C. J., and Barton, J. J. S. (2007). What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain Res.* 1127, 80–89. doi: 10.1016/j.brainres.2006.09.104

Fox, C. J., Oruç, I., and Barton, J. J. S. (2008). It doesn't matter how you feel. The facial identity aftereffect is invariant to changes in facial expression. *J. Vis.* 8, 11.1–13. doi: 10.1167/8.3.11

Frisby, J. P. (1979). *Seeing: Illusion, brain, and mind*. Oxford: Oxford University Press.

Gibson, J. J., and Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. *J. Exp. Psychol.* 20, 453–467. doi: 10.1037/h0059826

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10, 14–23. doi: 10.1016/j.tics.2005.11.006

Haberman, J., and Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* 17, R751–R753. doi: 10.1016/j.cub.2007.06.039

Haberman, J., and Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 718–734. doi: 10.1037/a0013899

Harms, M. B., Martin, A., and Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychol. Rev.* 20, 290–322. doi: 10.1007/s11065-010-9138-6

Harris, A., and Nakayama, K. (2007). Rapid face-selective adaptation of an early extrastriate component in MEG. *Cereb. Cortex* 17, 63–70. doi: 10.1093/cercor/bhj124

Harris, A., and Nakayama, K. (2008). Rapid adaptation of the M170 response: Importance of face parts. *Cereb. Cortex* 18, 467–476. doi: 10.1093/cercor/bhm078

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

Hong, S. W., and Yoon, K. L. (2018). Intensity dependence in high-level facial expression adaptation aftereffect. *Psychon. Bull. Rev.* 25, 1035–1042. doi: 10.3758/s13423-017-1336-2

Hsu, S., and Young, A. (2004). Adaptation effects in facial expression recognition. *Vis. cogn.* 11, 871–899. doi: 10.1080/13506280444000030

Hudac, C. M., Santhosh, M., Celerian, C., Chung, K. M., Jung, W., and Webb, S. J. (2021). The role of racial and developmental experience on emotional adaptive coding in autism spectrum disorder. *Dev. Neuropsychol.* 46, 93–108. doi: 10.1080/87565641.2021.1900192

Jaquet, E., and Rhodes, G. (2008). Face aftereffects indicate dissociable, but not distinct, coding of male and female faces. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 101–112. doi: 10.1037/0096-1523.34.1.101

Jeffery, L., and Rhodes, G. (2011). Insights into the development of face recognition mechanisms revealed by face aftereffects. *Br. J. Psychol.* 102, 799–815. doi: 10.1111/j.2044-8295.2011.02066.x

Jeffery, L., McKone, E., Haynes, R., Firth, E., Pellicano, E., and Rhodes, G. (2010). Four-to-six-year-old children use norm-based coding in face-space. *J. Vis.* 10:18. doi: 10.1167/10.5.18

Jeffery, L., Read, A., and Rhodes, G. (2013). Four year-olds use norm-based coding for face identity. *Cognition* 127, 258–263. doi: 10.1016/j.cognition.2013.01.008

Jeffery L., Rhodes, G., McKone, E., Pellicano, E., Crookes, K., and Taylor, E. (2011). Distinguishing norm-based from exemplar-based coding of identity in children: Evidence from face identity aftereffects. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1824–1840. doi: 10.1037/a0025643

Jenkins, R., Beaver, J. D., and Calder, A. J. (2006). I thought you were looking at me: Direction-specific aftereffects in gaze perception. *Psychol. Sci.* 17, 506–513. doi: 10.1111/j.1467-9280.2006.01736.x

Juricevic, I., and Webster, M. A. (2012). Selectivity of face aftereffects for expressions and anti-expressions. *Front. Psychol.* 3:1–10. doi: 10.3389/fpsyg.2012.00004

Kaufmann, J. M., and Schweinberger, S. R. (2004). Expression influences the recognition of familiar faces. *Perception* 33, 399–408. doi: 10.1068/p5083

Kovács, G., Cziraki, C., Vidnyánszky, Z., Schweinberger, S. R., and Greenlee, M. W. (2008). Position-specific and position-invariant face aftereffects reflect the adaptation of different cortical areas. *Neuroimage* 43, 154–164. doi: 10.1016/j.neuroimage.2008.06.042

Kovács, G., Zimmeer, M., Harza, I., Antal, A., and Vidnyánszky, Z. (2005). Position-specificity of facial adaptation. *Neuroreport* 16, 1945–1949. doi: 10.1097/01.wnr.0000187635.76127.bc

Kovács, G., Zimmer, M., Harza, I., and Vidnyánszky, Z. (2007). Adaptation duration affects the spatial selectivity of facial aftereffects. *Vision Res.* 47, 3141–3149. doi: 10.1016/j.visres.2007.08.019

Laurence, S., and Hole, G. (2012). Identity specific adaptation with composite faces. *Vis. cogn.* 20, 109–120. doi: 10.1080/13506285.2012.655805

Leopold, D. A., O'Toole, A. J., Vetter, T., and Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat. Neurosci.* 4, 89–94. doi: 10.1038/82947

Leopold, D. A., Rhodes, G., Müller, K.-M., and Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proc. R. Soc. B Biol. Sci.* 272, 897–904. doi: 10.1098/rspb.2004.3022

Little, A. C., DeBruine, L. M., and Jones, B. C. (2005). Sex-contingent face after-effects suggest distinct neural populations code male and female faces. *Proc. R. Soc. B Biol. Sci.* 272, 2283–2287. doi: 10.1098/rspb.2005.3220

Marsh, A. A., Ambady, N., and Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach- and avoidance-related behaviors. *Emotion* 5, 119–124. doi: 10.1037/1528-3542.5.1.119

Marsh, A. A., Kozak, M. N., and Ambady, N. (2007). Accurate identification of fear facial expressions predicts prosocial behavior. *Emotion* 7, 239–251. doi: 10.1037/1528-3542.7.2.239

Mian, J. F., and Mondloch, C. J. (2012). Recognizing identity in the face of change: The development of an expression-independent representation of facial identity. *J. Vis.* 12, 17–17. doi: 10.1167/12.7.17

Minemoto, K., Ueda, Y., and Yoshikawa, S. (2022a). "New use of the facial adaptation method: Understanding what facial expressions evoke social signals," in *Proceedings of the vision science society (VSS)*, (St.Petersburg, FL).

Minemoto, K., Ueda, Y., and Yoshikawa, S. (2022b). The aftereffect of the ensemble average of facial expressions on subsequent facial expression recognition. *Atten. Percept. Psychophys.* 84, 815–828. doi: 10.3758/s13414-021-02407-w

Mondloch, C. J., Geldart, S., Maurer, D., and Le Grand, R. (2003). Developmental changes in face processing skills. *J. Exp. Child Psychol.* 86, 67–84. doi: 10.1016/S0022-0965(03)00102-4

Moriya, J., Tanno, Y., and Sugiura, Y. (2013). Repeated short presentations of morphed facial expressions change recognition and evaluation of facial expressions. *Psychol. Res.* 77, 698–707. doi: 10.1007/s00426-012-0463-7

Mueller, R., Utz, S., Carbon, C. C., and Strobach, T. (2020). Face adaptation and face priming as tools for getting insights into the quality of face space. *Front. Psychol.* 11:166. doi: 10.3389/fpsyg.2020.00166

Nemrodov, D., and Itier, R. J. (2011). The role of eyes in early face processing: A rapid adaptation study of the inversion effect. *Br. J. Psychol.* 102, 783–798. doi: 10.1111/j.2044-8295.2011.02033.x

Nishimura, M., Doyle, J., Humphreys, K., and Behrmann, M. (2010). Probing the face-space of individuals with prosopagnosia. *Neuropsychologia* 48, 1828–1841. doi: 10.1016/j.neuropsychologia.2010.03.007

Nishimura, M., Maurer, D., Jeffery, L., Pellicano, E., and Rhodes, G. (2008). Fitting the child's mind to the world: Adaptive norm-based coding of facial identity in 8-year-olds. *Dev. Sci.* 11, 620–627. doi: 10.1111/j.1467-7687.2008.00706.x

O'Craven, K. M., and Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J. Cogn. Neurosci.* 12, 1013–1023. doi: 10.1162/08989290051137549

Oosterhof, N. N., and Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion* 9, 128–133. doi: 10.1037/a0014520

Palermo, R., Jeffery, L., Lewandowsky, J., Fiorentini, C., Irons, J. L., Dawel, A., et al. (2018). Adaptive face coding contributes to individual differences in facial expression recognition independently of affective factors. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 503–517. doi: 10.1037/xhp0000463

Palermo, R., Rivolta, D., Wilson, C. E., and Jeffery, L. (2011). Adaptive face space coding in congenital prosopagnosia: Typical figural aftereffects but abnormal identity aftereffects. *Neuropsychologia* 49, 3801–3812. doi: 10.1016/j.neuropsychologia.2011.09.039

Pell, P. J., and Richards, A. (2011). Cross-emotion facial expression aftereffects. *Vision Res.* 51, 1889–1896. doi: 10.1016/j.visres.2011.06.017

Pell, P. J., and Richards, A. (2013). Overlapping facial expression representations are identity-dependent. *Vision Res.* 79, 1–7. doi: 10.1016/j.visres.2012.12.009

Pellicano, E., Jeffery, L., Burr, D., and Rhodes, G. (2007). Abnormal adaptive face-coding mechanisms in children with autism spectrum disorder. *Curr. Biol.* 17, 1508–1512. doi: 10.1016/j.cub.2007.07.065

Petrovski, S., Rhodes, G., and Jeffery, L. (2018). Adaptation to dynamic faces produces face identity aftereffects. *J. Vis.* 18, 1–11. doi: 10.1167/18.13.13

Prete, G., Laeng, B., and Tommasi, L. (2018). Modulating adaptation to emotional faces by spatial frequency filtering. *Psychol. Res.* 82, 310–323. doi: 10.1007/s00426-016-0830-x

Rhodes, G., and Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Res.* 46, 2977–2987. doi: 10.1016/j.visres.2006.03.002

Rhodes, G., and Leopold, D. A. (2012). "Adaptive norm-based coding of face identity," in *Oxford handbook of face perception*, eds A. W. Calder, G. Rhodes, M. H. Johnston, and J. V. Haxby (Oxford: Oxford University Press), 263–286. doi: 10.1093/oxfordhb/9780199559053.013.0014

Rhodes, G., Brennan, S., and Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cogn. Psychol.* 19, 473–497. doi: 10.1016/0010-0285(87)90016-8

Rhodes, G., Burton, N., Jeffery, L., Read, A., Taylor, L., and Ewing, L. (2018). Facial expression coding in children and adolescents with autism: Reduced adaptability but intact norm-based coding. *Br. J. Psychol.* 109, 204–218. doi: 10.1111/bjop.12257

Rhodes, G., Ewing, L., Jeffery, L., Avard, E., and Taylor, L. (2014a). Reduced adaptability, but no fundamental disruption, of norm-based face-coding mechanisms in cognitively able children and adolescents with autism. *Neuropsychologia* 62, 262–268. doi: 10.1016/j.neuropsychologia.2014.07.030

Rhodes, G., Jeffery, L., Clifford, C. W. G., and Leopold, D. A. (2007). The timecourse of higher-level face aftereffects. *Vision Res.* 47, 2291–2296. doi: 10.1016/j.visres.2007.05.012

Rhodes, G., Jeffery, L., Taylor, L., Hayward, W. G., and Ewing, L. (2014b). Individual differences in adaptive coding of face identity are linked to individual differences in face recognition ability. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 897–903. doi: 10.1037/a0035939

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W. G., and Nakayama, K. (2003). Fitting the mind to the world: Face adaptation and attractiveness aftereffects. *Psychol. Sci.* 14, 558–566. doi: 10.1046/j.0956-7976.2003.psci_1465.x

Rhodes, G., Pond, S., Burton, N., Kloth, N., Jeffery, L., Bell, J., et al. (2015). How distinct is the coding of face identity and expression? Evidence for some common dimensions in face space. *Cognition* 142, 123–137. doi: 10.1016/j.cognition.2015.05.012

Rhodes, G., Pond, S., Jeffery, L., Benton, C. P., Skinner, A. L., and Burton, N. (2017). Aftereffects support opponent coding of expression. *J. Exp. Psychol. Hum. Percept. Perform.* 43, 619–628. doi: 10.1037/xhp0000322

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714

Russell, J. A., and Bullock, M. (1985). Multidimensional scaling of emotional facial expressions. Similarity from preschoolers to adults. *J. Pers. Soc. Psychol.* 48, 1290–1298. doi: 10.1037/0022-3514.48.5.1290

Rutherford, M. D., Chattha, H. M., and Krysko, K. M. (2008). The use of aftereffects in the study of relationships among emotion categories. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 27–40. doi: 10.1037/0096-1523.34.1.27

Rutherford, M. D., Troubridge, E. K., and Walsh, J. (2012). Visual afterimages of emotional faces in high functioning autism. *J. Autism Dev. Disord.* 42, 221–229. doi: 10.1007/s10803-011-1233-x

Ryu, J. J., Borrmann, K., and Chaudhuri, A. (2008). Imagine jane and identify John: Face identity after effects induced by imagined faces. *PLoS One* 3:e2195. doi: 10.1371/journal.pone.0002195

Schweinberger, S. R., and Neumann, M. F. (2016). Repetition effects in human ERPs to faces. *Cortex* 80, 141–153. doi: 10.1016/j.cortex.2015.11.001

Seyama, J., and Nagayama, R. S. (2006). Eye direction aftereffect. *Psychol. Res.* 70, 59–67. doi: 10.1007/s00426-004-0188-3

Skinner, A. L., and Benton, C. P. (2010). Anti-expression aftereffects reveal prototype-referenced coding of facial expressions. *Psychol. Sci. J. Am. Psychol. Soc.* 21, 1248–1253. doi: 10.1177/0956797610380702

Skinner, A. L., and Benton, C. P. (2012). The expressions of strangers: Our identity-independent representation of facial expression. *J. Vis.* 12, 1–13. doi: 10.1167/12.2.12

Song, M., Shinomori, K., Qian, Q., Yin, J., and Zeng, W. (2015). The change of expression configuration affects identity-dependent expression aftereffect but not identity-independent expression aftereffect. *Front. Psychol.* 6:1937. doi: 10.3389/fpsyg.2015.01937

Sou, K. L., and Xu, H. (2019). Brief facial emotion aftereffect occurs earlier for angry than happy adaptation. *Vision Res.* 162, 35–42. doi: 10.1016/j.visres.2019.07.002

Stirrat, M., and Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychol. Sci.* 21, 349–354. doi: 10.1177/0956797610362647

Strobach, T., and Carbon, C. C. (2013). Face adaptation effects: Reviewing the impact of adapting information, time, and transfer. *Front. Psychol.* 4:1–12. doi: 10.3389/fpsyg.2013.00318

Susilo, T., McKone, E., Dennett, H., Darke, H., Palermo, R., Hall, A., et al. (2010). Face recognition impairments despite normal holistic processing and face space coding: Evidence from a case of developmental prosopagnosia. *Cogn. Neuropsychol.* 27, 636–664. doi: 10.1080/02643294.2011.613372

Swe, D., Burton, N., and Rhodes, G. (2019). Can expression aftereffects be explained by tilt adaptation? *J. Vision* 19:79. doi: 10.1167/19.8.79

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol. Sect. A* 43, 161–204. doi: 10.1080/14640749108400966

Vicari, S., Reilly, J. S., Pasqualetti, P., Vizzotto, A., and Caltagirone, C. (2000). Recognition of facial expressions of emotions in school-age children: The intersection of perceptual and semantic categories. *Acta Paediatr. Int. J. Paediatr.* 89, 836–845. doi: 10.1111/j.1651-2227.2000.tb00392.x

Vida, M. D., and Mondloch, C. J. (2009). Children's representations of facial expression and identity: Identity-contingent expression aftereffects. *J. Exp. Child Psychol.* 104, 326–345. doi: 10.1016/j.jecp.2009.06.003

Walsh, J. A., Maurer, D., Vida, M. D., Rhodes, G., Jeffery, L., and Rutherford, M. D. (2015). Norm-based coding of facial identity in adults with autism spectrum disorder. *Vision Res.* 108, 33–40. doi: 10.1016/j.visres.2014.11.018

Walther, C., Schweinberger, S. R., Kaiser, D., and Kovács, G. (2013). Neural correlates of priming and adaptation in familiar face perception. *Cortex* 49, 1963–1977. doi: 10.1016/j.cortex.2012.08.012

Webster, M. a, and MacLeod, D. I. a (2011). Visual adaptation and face perception. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 366, 1702–1725. doi: 10.1098/rstb.2010.0360

Webster, M. A. (1996). Human colour perception and its adaptation. *Netw. Comput. Neural Syst.* 7, 587–634. doi: 10.1088/0954-898X_7_4_002

Webster, M. A., and MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychon. Bull. Rev.* 6, 647–653. doi: 10.3758/BF03212974

Webster, M. A., Kaping, D., Mizokami, Y., and Duhamel, P. (2004). Adaptation to natural facial categories. *Nature* 428, 557–561. doi: 10.1038/nature02420

Weigelt, S., Koldewyn, K., and Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neurosci. Biobehav. Rev.* 36, 1060–1084. doi: 10.1016/j.neubiorev.2011.12.008

Wincenciak, J., Dzhelyova, M., Perrett, D. I., and Barraclough, N. E. (2013). Adaptation to facial trustworthiness is different in female and male observers. *Vision Res.* 87, 30–34. doi: 10.1016/j.visres.2013.05.007

Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R., and Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *J. Neurophysiol.* 92, 1830–1839. doi: 10.1152/jn.00155.2004

Winston, J. S., Strange, B. A., O'Doherty, J., and Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nat. Neurosci.* 5, 277–283. doi: 10.1038/nn816

Witham, C., Foo, Y. Z., Jeffery, L., Burton, N. S., and Rhodes, G. (2021). Anger and fearful expressions influence perceptions of physical strength: Testing the signalling functions of emotional facial expressions with a visual aftereffects paradigm. *Evol. Hum. Behav* 42, 547–555. doi: 10.1016/j.evolhumbehav.2021.05.005

Xu, H., Dayan, P., Lipkin, R. M., and Qian, N. (2008). Adaptation across the cortical hierarchy: Low-level curve adaptation affects high-level facial-expression judgments. *J. Neurosci.* 28, 3374–3383. doi: 10.1523/JNEUROSCI.0182-08.2008

Ying, H., and Xu, H. (2017). Adaptation reveals that facial expression averaging occurs during rapid serial presentation. *J. Vision* 17, 1–19. doi: 10.1167/17.1.15

Young, A. W., Hellawell, D., and Hay, D. C. (1987). Configurational information in face perception. *Perception* 16, 747–759. doi: 10.1068/p160747

Zamuner, E., Oxner, M., and Hayward, W. G. (2017). Visual perception and visual mental imagery of emotional faces generate similar expression aftereffects. *Conscious. Cogn.* 48, 171–179. doi: 10.1016/j.concog.2016.11.010

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?*. Boulder, CO: Westview Press.

Zimmer, M., and Kovács, G. (2011). Position specificity of adaptation-related face aftereffects. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 586–595. doi: 10.1098/rstb.2010.0265

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 12 January 2023
DOI 10.3389/fpsyg.2022.1035524

Check for updates

OPEN ACCESS

EDITED BY
Chihiro Saegusa,
Kao Corporation,
Japan

REVIEWED BY
Tomoyo Morita,
National Institute of Information and Communications Technology,
Japan
Rufin Vogels,
KU Leuven,
Belgium
Jessica Taubert,
The University of Queensland,
Australia

*CORRESPONDENCE
Ryan Elson
✉ ryan.elson@nottingham.ac.uk

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

RECEIVED 02 September 2022
ACCEPTED 20 December 2022
PUBLISHED 12 January 2023

CITATION
Elson R, Schluppeck D and Johnston A (2023) fMRI evidence that hyper-caricatured faces activate object-selective cortex.
Front. Psychol. 13:1035524.
doi: 10.3389/fpsyg.2022.1035524

fMRI evidence that hyper-caricatured faces activate object-selective cortex

Ryan Elson*, Denis Schluppeck and Alan Johnston

School of Psychology, University of Nottingham, Nottingham, United Kingdom

Many brain imaging studies have looked at the cortical responses to object categories and faces. A popular way to manipulate face stimuli is by using a "face space," a high dimensional representation of individual face images, with the average face located at the origin. However, how the brain responds to faces that deviate substantially from average has not been much explored. Increasing the distance from the average (leading to increased caricaturing) could increase neural responses in face-selective regions, an idea supported by results from non-human primates. Here, we used a face space based on principal component analysis (PCA) to generate faces ranging from average to heavily caricatured. Using functional magnetic resonance imaging (fMRI), we first independently defined face-, object- and scene-selective areas with a localiser scan and then measured responses to parametrically caricatured faces. We also included conditions in which the images of faces were inverted. Interestingly in the right fusiform face area (FFA), we found that the patterns of fMRI response were more consistent as caricaturing increased. However, we found no consistent effect of either caricature level or facial inversion on the average fMRI response in the FFA or face-selective regions more broadly. In contrast, object-selective regions showed an increase in both the consistency of response pattern and the average fMRI response with increasing caricature level. This shows that caricatured faces recruit processing from regions typically defined as object-selective, possibly through enhancing low-level properties that are characteristic of objects.

KEYWORDS

faces, object-selective cortex, face space, fMRI, caricatured faces, PCA

1. Introduction

In regular social interactions, we may encounter hundreds of faces every day. Most human observers can rapidly recognise the identity (Ramon et al., 2011), process the emotion (Leppänen and Hietanen, 2004), or form an impression of a person or their intentions (Bar et al., 2006; Willis and Todorov, 2006; Sutherland et al., 2013) from visual information alone.

It is estimated that humans know on average around 5,000 faces (Jenkins et al., 2018) but despite much research, it is largely unknown how we encode all those familiar faces, in addition to all unfamiliar ones. Face space (Valentine, 1991; Valentine et al., 2016), an influential account of face representation, has been widely used to study the neural representation of faces in humans (Loffler et al., 2005; Carlin and Kriegeskorte, 2017) and non-human primates (Leopold et al., 2006; Chang and Tsao, 2017). The idea has also found application in automatic face recognition systems (Sirovich and Kirby, 1987; Turk and Pentland, 1991; Zhu et al., 2013; Deng et al., 2014). Caricatured face images that deviate substantially from average, including artistic caricatures, evidently amplify characteristic features of faces. But it is unclear if and how face space is represented in the brain and what the exact neural representation of faces distant from the average face might be.

In *face space*, a multidimensional space with a representation of an average face at the origin, individual face *exemplars* are thought of as points (at a certain distance and direction with respect to the origin). The dimensions of this space could be derived from discrete, descriptive changes in the shape or position of features (e.g., the distance between the eyes or the width of the mouth). Alternatively, the dimensions may reflect more abstract and global descriptors of shape and texture.

In a face space representation, individual *identities* correspond to a given direction relative to the origin. The distance from the origin indicates how different a particular face is from the average. Faces whose representation is located a greater distance from the average are expected to generate stronger responses from the population of neurons sensitive to the given identity's facial properties. This idea of "norm-based" coding, coding relative to the average or norm, has received strong supporting evidence (e.g., see Leopold et al., 2001; Anderson and Wilson, 2005; Jiang et al., 2006, 2007; Rhodes and Jeffery, 2006; Webster and MacLeod, 2011; Little, 2012; Chang and Tsao, 2017). Neurons representing facial information could form a basis to span this space, rather than being tuned to a particular identity. The projections of a face onto a set of basis neurons may code the different identities (Chang and Tsao, 2017) in terms of the relative firing rates of this population of neurons.

The neural basis of norm-based coding has recently been clarified by new research in macaques (Koyano et al., 2021). Rhodes and Jeffery (2006) proposed that norm-based coding was based on two opponent channels with the average face activating each equally. The opponent channels can be associated with 'axis coding' that shows monotonic ramp-tuning through the norm in single-cell recordings (Chang and Tsao, 2017). Norm-based coding also gives rise to V-shaped coding (e.g., Freiwald and Hosoya, 2021) whereby there is a minimum response to the norm relative to more peripheral faces, regardless of direction. V-shaped coding was first demonstrated at a single cell level and at a population level by Leopold et al. (2006). Recently, evidence has shown that both mechanisms are present in the same set of individual neurons, with axis coding occurring approximately 100 ms before V-shaped coding (Koyano et al., 2021). However, the V-shape was driven by a decrease in the firing rate to average faces, likely from lateral inhibition resulting from synchronous firing across the population to the average face (Koyano et al., 2021). Whilst axis coding supports the initial coding of the neuron, V-shaped responses reflect a consequence of many neurons firing to the average face in synchrony.

Chang and Tsao (2017) show that rather than responding to specific identities, neurons in areas ML/MF of the macaque temporal lobe (middle lateral/middle fundus) and the more anterior AM (anterior medial) responded to combinations of shape and texture information. Firing rates increased linearly with the magnitude of a face's projection onto the neuron's preferred dimension or 'axis' of change, but only in the preferred direction of change; face stimuli along the same axis but on the opposite side of the mean decreased the neuron's firing rate. Variations in facial appearance orthogonal to a neuron's preferred dimension, however, did not change its firing rates. This invariance to changes along orthogonal axes may explain the lack of an aftereffect to faces that lie on a different trajectory from the adapting stimulus (Leopold et al., 2001; Anderson and Wilson, 2005; Rhodes and Jeffery, 2006). From a theoretical standpoint, it allows face processing to be based on a highly efficient calculation (linear projection), requiring relatively few neurons to encode a very high-dimensional face space. Interestingly, it has also been found that faces activate a more broadly-based representation within an object space. Recent work has shown that faces may be situated in the animate, 'stubby' quadrant of the identified 2D (animate/inanimate and stubby/spiky) space, although many aspects of this representation remain unknown (Bao et al., 2020).

Because faces are more densely clustered around the mean, those further from average should appear more distinctive (Valentine et al., 2016). This idea is supported by evidence showing that caricatures are rated as more distinctive than their veridical face or anti-caricature (Lee et al., 2000). If the dimensions are ordered in terms of the amount of facial variance they encode, then more distinctive faces may also load more onto less prevalent dimensions of variation, in which case direction in the space may also reflect distinctiveness (Hancock et al., 1996). The direction and distinctiveness in face space not only impacts recognition, but also the first impression that is attributed to that face (Olivola et al., 2014; Over and Cook, 2018), and can indicate poor childhood health or genetic disorders (Rhodes et al., 2001; Gad et al., 2008; Babovic-Vuksanovic et al., 2012; Dolci et al., 2021).

Faces can also be made artificially more distinctive through caricaturing. Caricatures, versions of 'veridical' face images that can be derived from extrapolations in face space, enhance behavioural performance over veridical faces, suggesting they may elicit stronger responses in the brain. Caricaturing line drawings and photographs enhances recognition (Rhodes et al., 1987; Mauro and Kubovy, 1992; Lee et al., 2000;Kaufmann and Schweinberger, 2012; Schulz et al., 2012), whilst anti-caricaturing (making the stimuli more average) leads to longer reaction times (Rhodes et al., 1987; Schulz et al., 2012) and reduced identification accuracy (Lee et al., 2000). Interestingly, caricaturing even improves recognition accuracy in deep convolutional neural networks (Hill et al., 2019). Subsequent recognition of veridical faces is enhanced by caricaturing during encoding (Rodríguez et al., 2009), suggesting that exaggerating the features or configuration can help create representations for new faces. Furthermore, adapting to caricatures makes veridical images appear more average (Carbon and Leder, 2005), consistent with the idea that the subset of neurons processing caricatured faces are the same as for their veridical versions. Caricaturing exemplars from the norm also increases the EEG amplitude of the face-selective N170 and N250 ERP responses (Kaufmann and Schweinberger, 2012; Schulz et al., 2012), although other neural responses such as the P200, decreased with distance from average (Schulz et al., 2012), suggesting that some neural processes may encode averageness and typicality.

Studies investigating distance from average on the neural response have adopted a variety of methods making direct comparison difficult (Loffler et al., 2005; Leopold et al., 2006; Susilo et al., 2010; Davidenko et al., 2012; McKone et al., 2014; Carlin and Kriegeskorte, 2017; Chang and Tsao, 2017). Chang and Tsao (2017) found near-linear increases with increasing distance through the average in macaques using single unit recordings, as has prior research (Leopold et al., 2006; which included moderately caricatured faces). Likewise, some behavioural work in humans using adaptation has found that the strength of the aftereffect caused by adapting to faces with varying eye and mouth height increased linearly, even outside the range of natural variability (Susilo et al., 2010). Other research suggests that the strength of identity aftereffects following adaptation increases linearly, but then is slightly reduced but constant past the 'naturalness boundary' (McKone et al., 2014). Results from functional magnetic resonance imaging (fMRI) studies have found saturating responses to stimuli at a certain distance from average (Loffler et al., 2005; Carlin and Kriegeskorte, 2017; see pg. 1,387). The faces in these studies did not extend far past the range of natural plausibility.

Electrical brain stimulation of the fusiform face area (FFA; Kanwisher et al., 1997) produces metamorphosis of viewed faces (Parvizi et al., 2012), suggesting that hyperactivity in the FFA delivers the perception of a caricatured face and thus may represent distance in face space. The perceived change in shape is consistent with suggestions that the FFA is homologous to the area ML in macaques (Tsao et al., 2003, 2008; note the 2003 paper refers to area ML as macaque area pSTS) given that this region shows greater sensitivity to shape over texture (Chang and Tsao, 2017). There is debate, however, over exact homology between human and macaque face processing systems (Yovel and Freiwald, 2013; Rossion and Taubert, 2019).

Hyper-caricatures, images that appear distorted beyond the range of natural appearance, can be generated by extrapolating in face space. In a face space constructed by principal component analysis (PCA), using weights much larger than those corresponding to typical faces shifts the representation further from the mean (see Figure 1). This allows the generation of a parametrically controlled set of realistic and hyper-caricatured faces that can be used as stimuli for brain imaging. Specifically, we wanted to explore how the blood-oxygen dependent (BOLD) fMRI signal changes in face-selective cortex, including the FFA, with stimuli at various distances from average in face space and with concomitant changes in perceived naturalness.

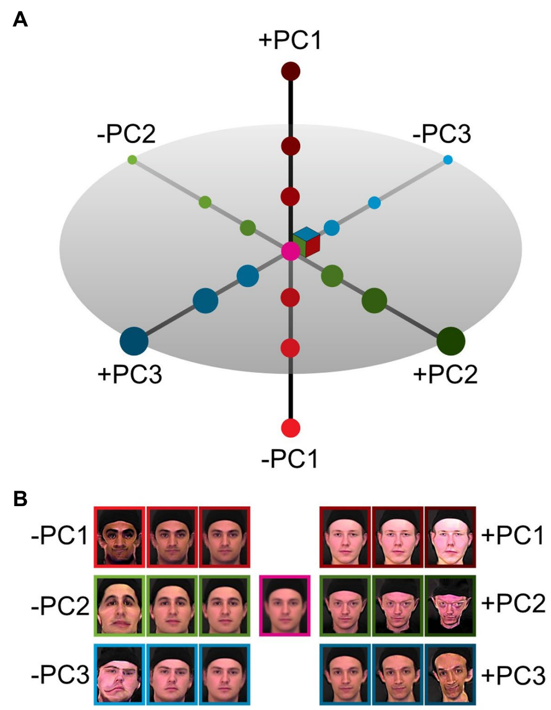


FIGURE 1
Caricaturing in face space. **(A)** An illustration of the three major axes of a principal component face space constructed from images of male faces. The origin in this space corresponds to the average face. Principal components (PC1 red, PC2 green, PC3 blue) are ordered by variance explained in the underlying data. **(B)** Example images created by modulating each of the principal components independently. Positive and negative deviations from the origin result in opposing changes in reconstructed images, increasingly caricatured with larger distances from the origin (average face, centre).

We hypothesised that there would be an increase in the BOLD response amplitude in the FFA and other face-selective areas with increases in caricature level. To summarise the experimental design, participants first undertook a behavioural session in which they identified the point along different directions in the PCA space where the face stimuli switch from appearing natural to caricatured. The caricature level of stimuli for the fMRI session were then chosen to straddle those perceptual boundaries: some stimuli appeared closer to average and natural, whilst others appeared hyper-caricatured. Stimuli were presented in an event-related design to avoid adaptation to a specific axis (Loffler et al., 2005; Davidenko et al., 2012). Inverted (upside down) stimuli were also presented to identify low-level effects of increased caricaturing (Davidenko et al., 2012). Inverted faces contain the same low-level properties as their upright counterparts, but have been shown to decrease the fMRI response in face-selective areas (Yovel and Kanwisher, 2004, 2005; Nasr and Tootell, 2012; James et al., 2013). We therefore considered that the effect of caricature level might be greater for upright faces than inverted faces.

Our results show that in the right fusiform face area (FFA), *the patterns of fMRI response* were more consistent as caricaturing increased. However, we found no consistent effect of either caricature level or facial inversion on the average fMRI response in the FFA or face-selective regions more widely. Therefore, we also explored the response in object and scene-selective areas. In contrast to face-selective regions, object-selective regions showed an increase in both the consistency of response pattern as well as average fMRI response with increasing caricature level.

2. Materials and methods

2.1. Participants

Nine healthy, neurologically intact volunteers with normal or corrected-to-normal vision were recruited for this study. Participants were aged between 22 and 36 years old (mean = 27 years, 6 months, SD = 4 years, 1 month). Three were female, six were male. No other demographic details were collected. The sample was a mix of postgraduate research students and staff from the School of Psychology at the University of Nottingham, recruited through a mix of convenience and snowball sampling. All participants gave fully informed consent and were screened for any MRI contraindicators before taking part in the experiment. The study was approved by the School's ethics committee.

2.2. Apparatus

The experiment was built in MATLAB version 9.5 (R2018b) using the Psychophysics Toolbox extensions (Psychtoolbox-3 version 3.0.17; Brainard, 1997; Pelli, 1997; Kleiner, 2007). The behavioural experiment was run on a 13" MacBook Pro (1,280 × 800 pixels). Participants responded solely through moving and clicking the mouse. Viewing distance was approximately 60 cm. For the MRI experiment, stimuli were presented on a 32″, 1,920 × 1,080 pixels BOLDscreen32 (CRS Ltd., Rochester, Kent) with a refresh rate of 120 Hz at the back of the bore through a mirror mounted on the head coil. Viewing distance was approximately 120 cm.

2.3. Stimuli

Stimuli were made using two separate PCA spaces, one derived from 50 images of female faces and another from 50 male faces. The input images were all aligned using the positions of the eyes and then warped to the average of the faces using the Multi-channel Gradient Model (Johnston et al., 1992, 1999), providing shape-free textures as well as the x and y warp information to convert the texture of the face back to the individual's facial shape. The x-y warp fields were appended to the shape-free textures and PCA was performed on these full warp-texture vectors using a procedure described by Nagle et al. (2013). The PCA extracts texture and shape covariations and maps these commonalities into an orthogonal space. Face images can be reconstructed by taking the texture for a given position of the PCA space, and spatially displacing the pixels by the distances contained in the corresponding x-y warp fields (see Supplementary Figure S1). Reconstructed stimuli were 100 pixels wide by 120 pixels high. In the MRI experiment, the stimuli were feathered into the RGB background around the edges.

To create the stimulus set for the experiment, the first 5 components in each of the PCA spaces were manipulated. The PCA returns eigenvectors of unit length. It also returns values of how the input images load onto each of the components. The components in our space were scaled by 1 standard deviation (SD) of the loadings, such that moving 1 'unit' along a given component reflected a change of 1 standard deviation of the loadings of the input set on that component.

2.4. Behavioural task

To establish the caricature levels at which faces turned from natural (physically plausible) to unnatural (physically implausible), we performed a behavioural experiment outside the scanner. This also helped to familiarise participants with the stimuli.

Stimuli scaled the first five components of each gender's PCA space in both the positive and negative directions (20 possible stimulus directions: 2 gender *5 PCs *2 directions), with each unique trial type presented 6 times in a random order – 120 trials in total. Stimuli were presented centrally on a grey background at half the screen height (approximately 11.2° of visual angle).

Using a method of adjustment, participants identified the transition points to unnatural stimuli by moving a mouse. Stimuli were dynamically updated at a caricature level controlled by the horizontal position of the mouse. A red dot on a scale bar served as a visual cue. Before each trial, an animation showed the full range of possible caricaturing for that trial (see Supplementary Figure S2A, for demonstration videos see Supplementary materials). Participants confirmed their choice with a mouse click and the next trials started after a 1,000 ms inter-stimulus interval. Because some components lead to distortions faster than others, the caricaturing applied to the stimuli was based on some pilot results from 5 independent participants (see Supplementary Table S1). Randomly varying the maximal amount of caricaturing on each trial prevented the slider's position being used to indicate the boundary for the given component.

No fixation cross was presented so participants could freely explore the faces, and there was no time limit. Breaks were provided every 40 trials. On average participants took approximately 30 minutes to complete the experiment.

For each participant, the average naturalness boundary for each component was calculated by taking the mean transition point across the 6 repetitions. The value of this position on the scale translated to the number of standard deviations (in terms of the loadings of the input set onto the PCA space) from the origin of the space. The results of the first 7 participants were used to scale the stimuli for the MRI experiment (see Supplementary Figure S3; Supplementary Table S2). Results of all participants can be seen in Supplementary Table S3.

2.5. MRI study

2.5.1. Localiser and caricature scans

The fMRI study consisted of two sets of scans. To find cortical regions responding to various categories of stimuli, we ran a standard functional localiser experiment using a randomised block design. We also ran a set of event-related scans in which individual images of test stimuli were presented ("caricature scans").

In the functional localiser, images of faces, scenes and objects were presented in a block design. Each block consisted of 8 images from one category. Face stimuli included photographs of 24 different identities (12 male, 12 female) taken at frontal pose, and 45° rotated in yaw in either direction. Not all views of each identity were presented. Images of scenes included both natural and manmade scenes, including pictures of buildings, both from the inside and outside. Objects included both manmade and natural objects. Faces and objects were presented on greyscale masks to occupy the same space as the scene stimuli (see Supplementary Figure S4). All stimuli were presented centrally and extended to approximately ± 8° of visual angle. Each stimulus was presented for 1 s with no ISI, with 8 s between blocks giving an 8 s ON, 8 s OFF sequence. The experiment began and ended with 8 s OFF. During the localiser a simple attention task was used: a black fixation cross was presented centrally throughout which 130 times within a scan turned red for 50 ms and participants had to respond by pressing any button on the button box. Any response within 1.5 s was classed as a hit. Each run of the functional localiser took 6 min and 32 s.

During the caricature scan participants were presented with stimuli created by modulating the first three components of the male PCA space from the behavioural study. Using the averaged naturalness boundaries from the behavioural experiment, participants were shown faces that corresponded to the mean (across participants) naturalness boundary (0SDs), one SD (across participant responses) closer to the average face (−1SD), or one, three or six SDs further away from the average (see Figure 2A).

In each run, participants were presented with 5 caricature levels for each component (the average naturalness boundary plus −1, 0, +1, +3, and + 6SDs). Picture plane inverted images of the most (+6SDs) and least caricatured (−1SD) face stimuli were also presented. This provided 21 unique stimuli per run

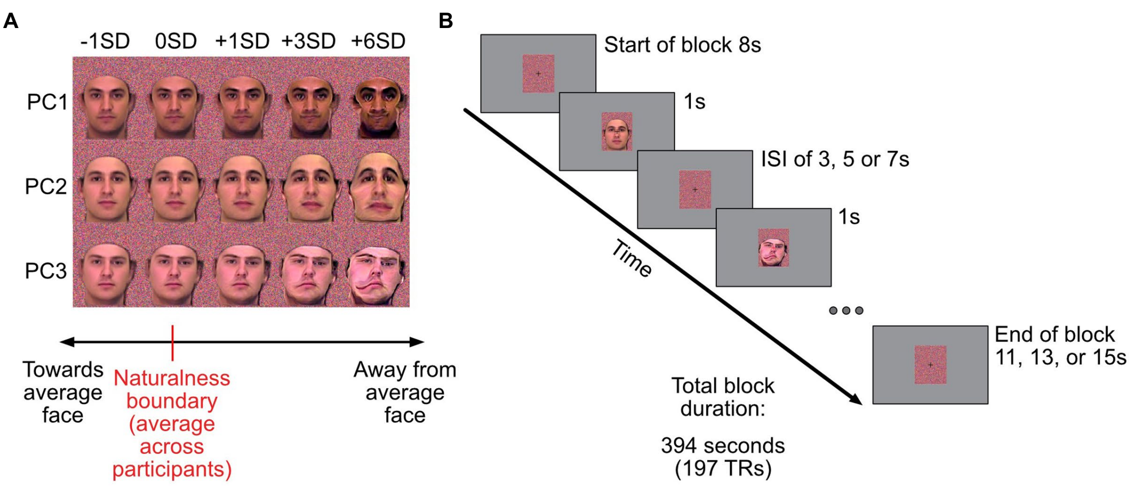


FIGURE 2
Caricatured stimuli used and outline of the fMRI experiment. **(A)** Example images corresponding to the five scaling levels used in the MRI experiment. First column of images: stimuli that are 1 standard deviation (SD) closer to the average face. Second column: group-averaged naturalness boundary for a given component (0SD). Other columns: images corresponding to +1, + 3 and+6SD away from average. Stimuli were presented on colour masks (Gaussian noise on each R, G, B channel, with mean and standard deviation derived from the face stimuli). **(B)** Timings for trials in the event-related caricature scan. The experiment started and ended with 8s of fixation (dynamically changing coloured masks only). Stimuli were presented for 1s, followed by a variable inter-stimulus interval of 3, 5 or 7s. To control for attention, participants had to report colour changes of the fixation cross (black to red), which occurred randomly 42 times within each run. The background colour mask changed dynamically every second.

(15 upright and 6 inverted) which were repeated 3 times each in a run. The experiment started with 8 s of rest. Subsequently stimuli were presented for 1 s with a variable ISI of either 3, 5, or 7 s, with equal numbers of each ISI duration across each run. Trial timings can be seen in Figure 2B. The order of stimuli and ISI durations was pseudorandomised across runs. To ensure all runs were the same duration, the final stimulus of each run was always followed by the remaining ISI and a further 8 s of rest (minimum of 11 s in total) to allow for the lag in the haemodynamic response. Each run lasted 6 min and 34 s. As in the localiser scan, participants responded when the centrally presented fixation cross turned red. This occurred 42 times during the run.

2.5.2. Data acquisition

For the localiser, functional data were acquired across 2 block-design runs, each lasting 392 s (196 volumes), one at the start of the scanning session and one at the end. Caricature scans were acquired across 3 event-related runs (4 for one participant), each lasting 394 s (197 volumes).

Data were acquired on a 3 T MRI scanner (Phillips Achieva) at the Sir Peter Mansfield Imaging Centre at the University of Nottingham using a standard 32 channel head coil. Functional (BOLD) images were acquired with 2D gradient echo EPI sequence (multiband 2, SENSE $r = 1$). Parameters were TR/TE 2000 ms/32 ms, FA 77°. There were 34 axial slices; voxel size was $2.4 \times 2.4 \times 3$ mm, 80×80 voxels per slice. High-resolution T1 MPRAGE structural images were obtained with the following parameters: TR/TE 8.1 ms/3.7 ms, 1 mm isotropic voxels, 256×256 voxels, FOV = 256×256 mm, 160 sagittal slices.

2.5.3. Data analysis

We used a combination of tools to analyse fMRI data: mrTools (Gardner et al., 2018) and custom MATLAB code, as well as FreeSurfer (Fischl, 2012) for cortical segmentation and anatomically defined regions of interest and FSL (Jenkinson et al., 2012) for spatial smoothing and mask dilation. Analyses were performed in individual participant space.

2.5.4. Anatomically restricting the analyses

We focused our analysis on the occipito-temporal cortex, bilaterally, including the FFA, the OFA (occipital face area; Puce et al., 1996; Halgren et al., 1999) and pSTS (posterior superior temporal sulcus; Morris et al., 1996, 1998). We defined larger anatomical ROIs from FreeSurfer parcellations to span the majority of the occipito-temporal cortex, spanning both hemispheres (combining 'lateraloccipital', 'fusiform', 'inferiortemporal', 'middletemporal', 'superiortemporal', 'bankssts', 'supramarginal' and 'inferiorparietal' ROIs from the Deskian/Killiany atlas). ROIs were created by converting the parcellation labels into volumetric masks (FreeSurfer: *mri_annotaion2label* and *mri_label2vol*) and dilated using a single pass of a 3-voxel box kernel to fill any holes (*fslmaths*).

2.5.5. Pre-processing

The caricature and localiser scans were first motion corrected within and between scans in mrTools (Gardner et al., 2018) using the mean volume of the second caricature scan (mid-point of the scanning session) as a reference frame. Motion correction used linear interpolation and drift correction was applied. The motion corrected functional runs were then spatially aligned to the participants' anatomical scans. The localiser data was spatially smoothed (3D Gaussian, FWHM 5 mm). For the caricature data, voxelwise data was extracted from the face, object, and scene-selective ROIs. For the univariate analysis and multivariate pattern analysis on the data no spatial smoothing was applied.

For both the localiser and caricature scans, data were converted to percentage signal change by subtracting the mean intensity for each voxel across the scan, and dividing by the mean [(x-mean)/mean], temporally high-pass filtered (cut-off 0.01 Hz) and, for the univariate analysis, concatenated over scans, taking care to keep track to the transition points between scans. This allowed for the GLM analysis to be reframed in block matrices, requiring only one GLM per set of localiser and caricature scans.

2.5.6. Defining the FFA and face-selective, object-selective, and scene-selective voxels

To define participant-specific functional ROIs, we used a GLM approach and restricted the analysis to the anatomical ROI described above. Analyses were performed in individual scan space. The 3 explanatory variables (EVs) were faces, objects and scenes, specified by 8 s ON boxcar regressors convolved with a double gamma haemodynamic response function (HRF). To define face-selective areas responses to face blocks were compared to blocks of objects and scenes (faces > objects + scenes). Voxels that responded significantly more to faces over objects and scenes were defined as face-selective. Corresponding contrasts then defined object-selective (objects > faces + scenes) and scene-selective areas (scenes > faces + objects). We used family-wise error (FWE) correction to account for multiple comparisons.

The functional ROIs were then defined on flat map representations of the corresponding statistical maps. A cluster corresponding to the FFA was present in each participant bilaterally (for details see Supplementary Table S4), however, in some participants, the boundaries were less clear, and even with family-wise error correction extended further along the fusiform gyrus and even into the neighbouring sulcus. In these cases, the FFA was defined as one contiguous cluster within a region restricted anatomically to the fusiform gyrus (from a FreeSurfer parcellation, FFA definition in each participant can be seen in Supplementary Figure S5). The pattern of response elsewhere however was more variable. Therefore, rather than trying to identify spatially consistent ROIs across participants, we simply allocated voxels to the 3 categories 'face-selective', 'object-selective' and 'scene-selective' based on the contrasts above. Face-selective voxels included the FFA. Voxels that responded significantly to more than one contrast were removed, such that each ROI only contained voxels that exclusively appeared for that contrast. Functional ROIs from one participant can be seen in Figures 3B–D.

2.5.7. Univariate analysis

To assess the effect of caricature level in the FFA, and face-, object- and scene-selective areas, we first used a deconvolution analysis (e.g., see Gardner et al., 2005; Besle et al., 2013). This provided an estimate of the event-related BOLD response for each of the 7 stimulus types (5 caricature levels, upright images; 2 inverted images). From these event-related responses (see Figure 3E), we calculated an index of the response amplitude of the first 5 TRs after stimulus onset, by first normalising to the level at stimulus onset (the first TR) and then obtaining the mean signed deviation (MSD) across the subsequent four TRs.

2.5.8. Multivariate pattern analysis

To look at patterns of response across the regions of interest, we also performed a correlation-based multivariate pattern analysis (MVPA). We compared the correlations in response patterns (beta values) between all 5 caricature levels of upright stimuli. The analysis was performed on the left and right FFA, left and right face-selective cortex, and left and right object-selective cortex.

The β values were obtained for each caricature scan repeat separately using a GLM similar to that described above, but assuming a canonical haemodynamic response function (double gamma). There were 5 explanatory variables (one for each upright caricature condition). The two additional conditions (inverted stimuli) were included as nuisance regressors. For each region of interest in the analysis, we then calculated the correlations of the β coefficient maps across regressors (avoiding within-scan comparisons). We then applied Fisher's transform to convert from correlation, r, to Z and averaged these Z-values across scans for each participant separately.

3. Results

The average fMRI response in the FFA, as well as face-selective voxels overall, did not show a consistent change with either caricature level or inversion, as assessed by univariate analysis and ANOVA. Interestingly, however, in the right fusiform face area (FFA), we found that the patterns of fMRI response were more consistent as caricaturing increased as assessed by multivariate pattern analysis (MVPA). In contrast, object-selective regions showed an increase in both average fMRI response with increasing caricature level (univariate analysis), and the consistency of response pattern (MVPA).

3.1. Univariate analyses

To assess the effects of caricature level and inversion in the FFA we performed two separate within-subjects ANOVAs, one to assess the effect of caricature level and orientation using the least and most caricatured faces, and one to assess the effect of caricature level using all 5 caricature levels of upright stimuli. The first was a $2 \times 2 \times 2$ ANOVA with hemisphere (left, right), stimulus orientation (upright, inverted) and caricature level (−1SD, +6SD), the second a 2×5 ANOVA with hemisphere (left, right) and caricature level (all 5 levels of upright caricature). ANOVAs were performed using IBM SPSS Statistics version 25.

To investigate the response amplitudes in the face-, object- and scene-selective voxels we performed the same two ANOVAs as for the FFA but including ROI as an additional independent variable with 3 levels (face-selective, object-selective, and scene-selective).

3.1.1. Caricature level in the FFA

The event-related response profiles showed a clear trial-locked response to the 5 caricature levels across all regions. Figure 3E shows the average deconvolution timeseries for the right FFA across subjects (thick lines), as well as traces for individual participants (thin lines).

When assessing caricature level (−1SD, +6SD), including both upright and inverted stimuli, there was no main effect of caricature level [$F(1,8) = 3.08$, $p = 0.117$, $\eta_p^2 = 0.28$], but there was a significant interaction between hemisphere and caricature level [$F(1,8) = 5.86$, $p = 0.042$, $\eta_p^2 = 0.42$]. The interaction was driven by a stronger increase in the response amplitude in the right FFA than the left FFA [$t(8) = 2.42$, $p = 0.042$] to an increase in caricature level (Figure 4A), although the effect of caricature level in the right FFA was marginal [$F(1,8) = 5.18$, $p = 0.052$, $\eta_p^2 = 0.39$].

We found no interaction between hemisphere and caricature level when we assessed all 5 levels of the upright stimuli [$F(4,32) = 1.59$, $p = 0.200$, $\eta_p^2 = 0.17$, Figure 4B], nor a main effect of caricature level [$F(4,32) = 1.99$, $p = 0.119$, $\eta_p^2 = 0.20$].

3.1.2. Caricature level in face, object and scene-selective regions

We also compared responses across face-selective regions more generally, as well as in object- and scene-selective regions (see Supplementary Table S5 for details). The data are shown in Figure 5.

We found no significant main effect of caricature level when assessing the effect of caricature level (extremes) and inversion (Figure 5A), but there was a significant interaction with ROI [$F(2,16) = 6.08$, $p = 0.011$, $\eta_p^2 = 0.43$]. This interaction showed the effect of caricature level was only present in the object-selective cortex, with the object-selective cortex increasing in response amplitude with an increase in caricature level [$t(8) = 2.49$, $p = 0.038$].

When assessing all 5 levels of upright caricature, there was a main effect of caricature level [$F(4,32) = 3.17$, $p = 0.027$, $\eta_p^2 = 0.28$] driven by a general increase in response amplitude as a function of caricature level, which was particularly prominent for highly caricatured (+6SD) faces. The ANOVA showed there to be a positive linear trend between response amplitude and caricature level [$F(1,8) = 16.83$, $p = 0.003$, $\eta_p^2 = 0.68$]. Highly caricatured faces (+6SD) elicited a stronger response than -1SD [$t(8) = 4.83$, $p = 0.001$], 0SD [$t(8) = 2.47$, $p = 0.039$] and + 1SD [$t(8) = 3.22$, $p = 0.012$] caricatures, although only the first of these survived Bonferroni-correction ($a = 0.005$).

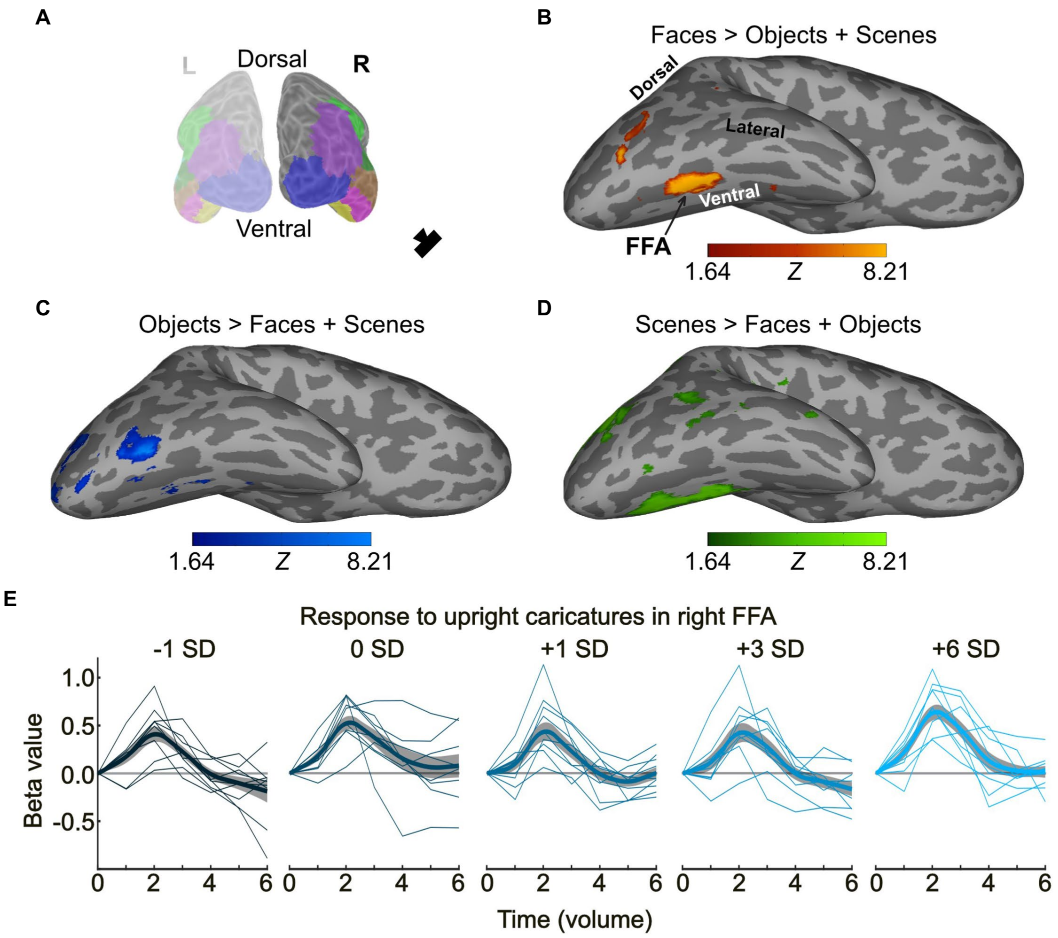


FIGURE 3
Defining face, object, and scene-selective regions and event-related fMRI responses in right FFA. **(A)** Posterior view of left and right inflated cortical hemispheres. Camera symbol indicates the view in panels B-D. Light grey, gyri; dark grey sulci. Regions in colour, Freesurfer parcellations used to form the bilateral occipitotemporal ROI, including lateral occipital (blue), fusiform gyrus (yellow), inferior temporal (pink), middle temporal (brown), bank of the superior temporal sulcus (dark green), inferior parietal (purple) and supramarginal gyrus (light green). **(B)** Face-selective regions in one participant based on the contrast faces > objects + scenes from the localiser scans (FFA, fusiform face area). Object-selective **(C)** and scene-selective **(D)** voxels defined using the contrasts objects > faces + scenes, and scenes > faces + objects, respectively. The colour bars in B-D show the Z-statistic for the contrast, thresholded at Z>1.64 (corresponding to *p*<0.05, with FWE correction). Maps show the voxels exclusively defined by these contrasts, with any overlap removed. **(E)** Response amplitude in the right FFA across participants from stimulus onset as a function of time for the five different levels of caricaturing (upright only) from the deconvolution analysis. Y values show the beta-coefficients from the deconvolution, normalised to *t*=0. Thin lines show the average timeseries for each participant. Thick lines show the group average, smoothed over time (for display purposes only). Shaded areas show±1 SEM across participants. Solid grey line shows *Y*=0. Colour represents the caricature level.

Although the interaction did not reach significance [$F(8,64) = 1.96$, $p = 0.066$, $\eta_p^2 = 0.20$], the overall effect was primarily driven by object-selective regions. The data in Figure 5B shows a constant response across caricature level in face-selective and scene-selective areas, but an increase in the response amplitude with increasing caricature level in the object-selective areas. To support this, separate ANOVAs for each ROI revealed that in the face-selective and scene-selective regions there was no effect of hemisphere, caricature, nor any interaction. In the object-selective regions there was no main effect of hemisphere nor interaction, but there was a significant effect of the caricature condition [$F(4,32) = 4.76$, $p = 0.004$, $\eta_p^2 = 0.37$] paired with a positive linear effect [$F(1,8) = 26.69$, $p = 0.001$, $\eta_p^2 = 0.77$]. Highly caricatured faces (+6SD) again elicited a stronger response over −1SD [$t(8) = 5.31$, $p = 0.001$], 0SD [$t(8) = 2.82$, $p = 0.023$] and + 1SD faces [$t(8) = 3.97$, $p = 0.004$]. The difference between +6SD and 0SD was not significant when correcting for multiple comparisons ($a = 0.005$). Interestingly faces on the naturalness boundary (0SD) also elicited a greater response than the most average (-1SD) faces [$t(8) = 2.58$, $p = 0.032$].

3.1.3. Effects of ROI, orientation, and hemisphere

We found that there was a decrease in response amplitude from face, to object, to scene-selective cortex in response to our face stimuli. Main effects of ROI were significant when assessing

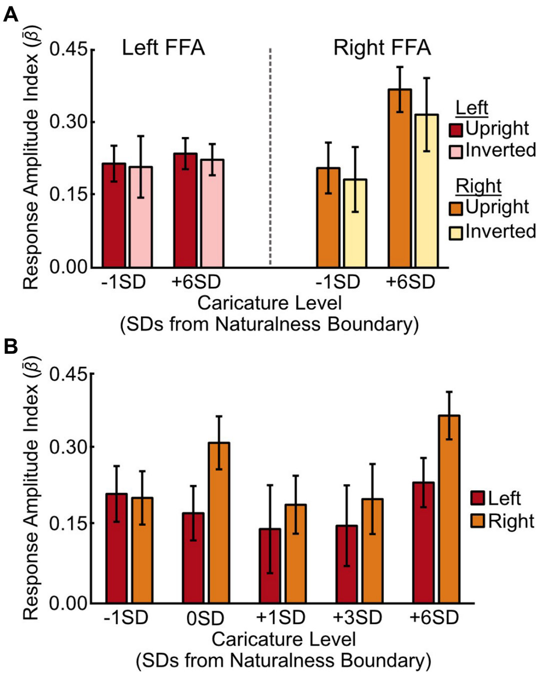


FIGURE 4
Response amplitudes in the FFA for different stimulus conditions. **(A)** Average response amplitude in the left (red/pink) and right (orange/yellow) FFA to the most (+6SD) and least (-1SD) caricatured faces in both the upright and inverted conditions. **(B)** Average response amplitudes in the left (red) and right (orange) FFA to each of the five caricature conditions for the upright stimuli only. Bars are grouped according to deviations from the average naturalness boundary (-1SD, closer to the average face; +6SD is highly caricatured). Y axes show the response amplitude index, measured by offsetting the β coefficients from the deconvolution analysis by t_0, and averaging $t_{1\text{-}4}$. Error bars show ± 1 SEM across participants.

the response to upright and inverted, −1SD and + 6SD caricatured stimuli [$F(1.27,10.12) = 9.63$, $p = 0.008$, $\eta_p^2 = 0.55$, Greenhouse–Geisser correction applied] and when assessing all 5 levels of upright stimuli [$F(2,16) = 9.95$, $p = 0.002$, $\eta_p^2 = 0.55$]. All pairwise comparisons were significant prior to correction (all $p < 0.046$) with the difference between face and scene-selective regions surviving correction ($a = 0.017$) in both analyses (both $p < 0.012$).

We found no main effects of, nor interactions with, orientation in any of our ROIs, and there were also no significant main effects of hemisphere. Generally, there was a greater response amplitude in the right hemisphere ROIs, which was most notable in the FFA when assessing the response to all five upright caricature levels [$F(1,8) = 4.96$, $p = 0.057$, $\eta_p^2 = 0.38$].

3.2. Multivariate pattern analysis

The results of the correlation analysis can be seen in Figure 6. In each ROI, we tested whether there was a significant positive correlation in the response patterns for each pair of caricature levels. Significance was assessed using one-sample t-tests to test if the group-level Z-value was significantly greater than 0 (Bonferroni-corrected $a = .003$) and is indicated by bold, underlined values in Figure 6. We then assessed how the response patterns varied as a function of caricature level using a one-way within-subjects ANOVA with the 5 levels of 'same' caricature correlations (i.e., the diagonals in Figure 6) as the independent variable.

3.2.1. Caricature level in the FFA

In the right FFA, the correlation coefficient (converted to Fisher's Z) increased as a function of caricature level (Figure 6D), supported by a significant positive linear trend [$F(1,8) = 7.60$, $p = 0.025$, $\eta_p^2 = 0.49$], indicating increasing consistency in the patterns of responses between stimulus categories including highly caricatured faces. In the left FFA, many of the correlations were significant at a group level, but the overall increase with caricature level, as seen in the right FFA, was not.

3.2.2. Caricature level in face and object-selective cortex

When looking at the consistency of response patterns in face and object-selective regions more broadly, we found that only object-selective regions bilaterally showed an increase in consistency with caricature level. Right face-selective regions were sensitive to caricature level, but the change in response pattern was less clear.

In the left face-selective regions there was no significant effect of caricature level on the correlations. In the right face-selective cortex there was a significant main effect of caricature level [$F(4,32) = 5.06$, $p = 0.003$, $\eta_p^2 = 0.39$] however the response profile was less clear than in the right FFA.

In both the left and the right object-selective regions there was a positive linear trend in the correlation to same caricature level trials as a function of caricature level [left: $F(1,8) = 21.91$, $p = 0.002$, $\eta_p^2 = 0.73$; right: $F(1,8) = 10.09$, $p = 0.013$, $\eta_p^2 = 0.56$]. At the group level however, no Z-values were greater than 0 in the left object-selective cortex. In the right hemisphere only correlations between +3SD and + 3SD stimuli and + 3SD and + 6SD stimuli were significant.

4. Discussion

We investigated the effect of caricaturing on the fMRI response in visual areas defined by preference to faces, objects, and scenes. Based on evidence of ramp coding in single cell recordings in macaques (Leopold et al., 2006; Chang and Tsao, 2017) to face stimuli of increasing distance from the mean face in the neuron's preferred direction of change, we reasoned that there may be an increase in the response amplitude of the FFA with increasing caricature level, even when faces appeared heavily distorted.

Surprisingly, we found no clear change in the average response amplitude in the FFA, or face-selective cortex more broadly, with increasing caricature level. In contrast, we found an increase in response in object-selective cortex, particularly for highly

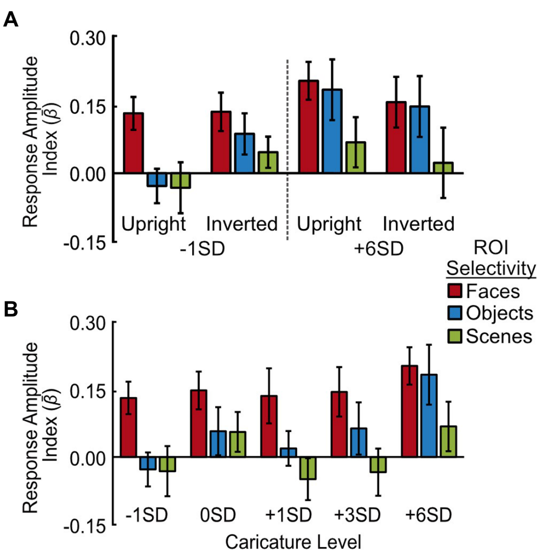


FIGURE 5
Response amplitudes to caricatured faces in face-, object- and scene-selective voxels. **(A)** Average response amplitudes for least (-1SD) and most (+6SD) caricatured stimuli. Data are grouped by image orientation (upright, inverted) and region of interest (face-, object-, and scene-selective voxels across hemispheres). **(B)** Average response amplitude to the 5 levels of caricatured, upright faces in face-selective (red), object-selective (blue) and scene-selective (green) voxels. As there was no interaction with hemisphere the response amplitudes are averaged across hemispheres. Y axes show the response amplitude index, measured by offsetting the β coefficients from the deconvolution analysis by t_0, and averaging $t_{1\text{-}4}$. Error bars show $\pm$1 SEM of the between-subjects variance.

caricatured faces. There was no significant change in response in scene-selective areas.

An increase in the consistency of the response pattern in object-selective cortex was also observed with increasing caricature level, measured using MVPA. Caricaturing therefore both enhanced the average response, and the consistency in which the stimuli were processed within object-selective cortex. How or why caricatured faces activate object-selective cortex is unclear.

The results seen in object-selective regions may result from changes to low-level or even mid-level properties that vary with caricature level, rather than a response to caricatured faces *per se* or the assignment of hyper-caricatured faces to a separate object class other than faces. Higher-level visual regions, including the FFA (Weibert et al., 2018), are sensitive to the lower-level image properties that are characteristic of different categories of objects (see Andrews et al., 2015). Caricaturing may have therefore emphasised particular low or mid-level properties that object-selective neurons are tuned to, such as certain shapes or curvatures that distinguish animate faces, bodies and animals from inanimate objects (Zachariou et al., 2018; Yue et al., 2020; Yetter et al., 2021) or changes in bilateral symmetry (Bona et al., 2015). The areas defined as object-selective responded *more* to objects than faces and scenes despite many voxels responding to all three categories (see Supplementary Figure S6) so the changes with caricature level may have generated stimulus properties that are more characteristic of generic objects than faces. The changes in our stimuli, including changes in texture, colour (Lafer-Sousa et al., 2016), shape, curvature (Yue et al., 2020; Yetter et al., 2021) and external contours, as well as higher-level changes, may have caused a shift in object-space (Bao et al., 2020).

Regardless of the exact mechanism for why or how caricatures activate object-selective cortex, it is evident that object-selective cortex is sensitive to caricature level, raising the possibility of its involvement in the perceptual evaluation of faces. To our knowledge, these are the first findings that show that caricatured faces elicit increased responses in regions typically involved in processing objects. These findings can potentially have important implications for understanding how we might form impressions from or recognise more distinctive faces. Although the faces in our experiment were artificially caricatured, faces in the real world can be naturally distinctive too, for example a number of (often genetic) disorders give rise to naturally distinctive faces (Gad et al., 2008; Babovic-Vuksanovic et al., 2012; Dolci et al., 2021). Our findings therefore raise a number of questions as to whether, and if so how, object-selective cortex contributes to our social evaluation of faces.

Returning to face-selective cortex, we initially found evidence that our most caricatured faces elicited a stronger response in the right FFA (but not left) compared to our least caricatured faces when we included both upright and inverted stimuli. This interaction between hemispheres is potentially consistent with the idea of a greater involvement of the right FFA in face perception, for example, electrical brain stimulation only impacts perception of faces when applied to the right FFA and not the left (Rangarajan et al., 2014).

We found no evidence of an effect of caricature level in face-selective cortex however when we assessed for a graded change in response amplitude across the complete range of caricature levels; there was no effect of hemisphere, caricature level nor an interaction, in either the FFA, specifically, or face-selective regions more widely. This may reflect a plateau in the BOLD response (Loffler et al., 2005; McKone et al., 2014; Carlin and Kriegeskorte, 2017). Since most of the stimuli were 'caricatured' to some degree the results could reflect response saturation; even the least caricatured stimuli could be identified as a particular individual. Alternatively, since Chang and Tsao (2017) report ramp-like tuned cells which increase their firing along an axis passing through the mean face, increasing caricature level may increase the firing of some cells whilst reducing the firing of others, leading to no net increase in the response across the population within a voxel.

Interestingly, the multivariate pattern of response in the right FFA became more consistent with increasing caricature level. This suggests a pattern of systematic increases and decreases in response rate across a population of cells. For the right face-selective regions more broadly, the change in spatial consistency was less clear, with slightly increased consistency for more caricatured faces, but decreased spatial consistency for intermediate caricatures. In the left hemisphere there was no

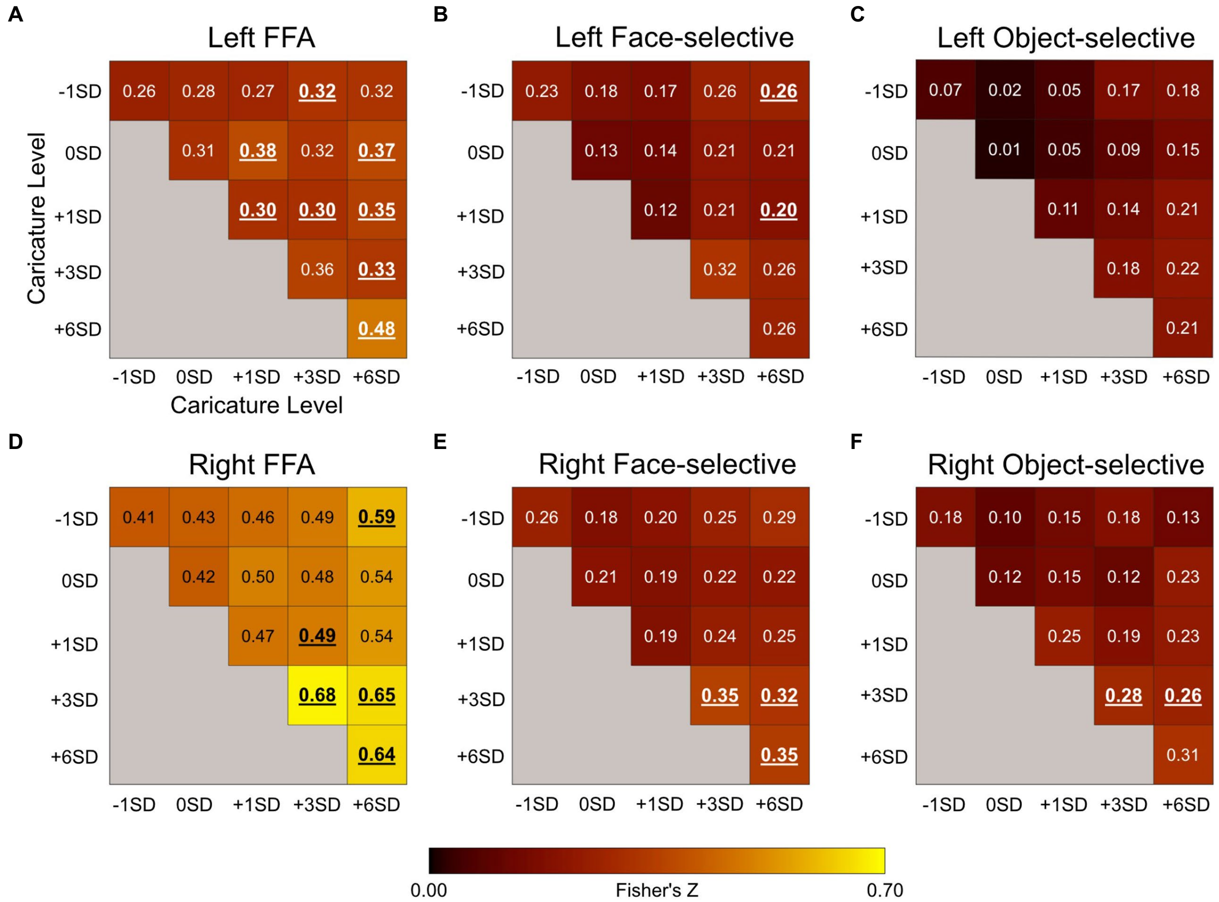


FIGURE 6
Plots showing the Fisher's Z for the correlation coefficients between β maps corresponding to the 5 upright levels of caricature from the MVPA analysis in the left FFA **(A)**, the left face-selective voxels **(B)**, left object-selective voxels **(C)**, the right FFA **(D)**, right face-selective voxels **(E)**, and the right object-selective voxels **(F)**. The diagonal reflects the average correlations between the response patterns to stimuli of the same caricature level, whilst the off-diagonal reflects correlations between different caricature levels. Only between-scan correlations were assessed. Values in bold and underlined were significantly greater than 0 at a group level, measured using one-sample *t*-tests (Bonferroni-corrected *a*=.003). Font colour for display purposes only.

effect of caricature level, consistent with the functional differences in the left and right FFA. The increase in spatial consistency in the absence of an average increase in the fMRI response in the right FFA is particularly interesting since stimuli at the same caricature level could vary substantially in terms of their low-level properties, given that different PCA components were modulated. Despite these low-level differences, which became more pronounced as caricaturing increased, the response pattern became significantly more consistent. This indicates that the increase in the overall consistency of the pattern was maintained regardless of any variation in the individual patterns themselves. Likewise, the general increase in consistency appeared to hold regardless of whether we compared the response patterns between the same caricature level, or different caricature levels. This suggests that different levels of caricature are processed by the same set of voxels, and is consistent with the idea that voxel-wise responses scale with varying distances from the average in a face space.

Our analysis of inversion showed no effect of orientation nor interaction with it. The lack of effect of orientation is at odds with some prior research showing an inversion effect in FFA (Yovel and Kanwisher, 2004, 2005; Nasr and Tootell, 2012; James et al., 2013), but is in line with other findings in literature suggesting that the inversion effect is weak (Gilaie-Dotan et al., 2010) or even absent (e.g., see Aguirre et al., 1999; Haxby et al., 1999; Epstein et al., 2006). The initial evidence of an effect of caricature level alongside a lack of inversion effect potentially suggests that the FFA is sensitive to changes in low-level properties, consistent with prior evidence (Weibert et al., 2018) and that this may not be specific to upright faces.

To conclude, in the FFA and face-selective areas more generally, we found no substantive effect of caricature level on the average response amplitude, although we did find evidence that the right FFA is sensitive to caricature level using MVPA, with the consistency of the response pattern increasing with caricature level. In contrast, we found a significant increase in both the response pattern consistency and the average response amplitude in object-selective cortex to increasing caricature level. This suggests that caricatured faces might recruit cortex typically defined as object-selective, potentially because they share more low-level features with objects. This may have implications for understanding how distinctive faces might be processed, both in terms of recognition and forming impressions.

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 at: "Investigating the fMRI response to highly caricatured faces," doi:10.17605/OSF.IO/JEC6U.

Ethics statement

The studies involving human participants were reviewed and approved by School of Psychology Ethics Committee, University of Nottingham. 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

RE, AJ, and DS contributed to the conception and design of the study. RE built the experimental scripts, wrote the manuscript. RE and DS collected the data. RE analysed the data with support from DS and AJ. AJ and DS supervised the study. All authors contributed to manuscript revision, read, and approved the final version for submission.

Funding

The scan time was awarded by the Sir Peter Mansfield Imaging Centre through a pump-priming scheme. This work was also supported by the Economic and Social Research Council [grant number: ES/P000711/1] for providing inconvenience allowances for participants. Open access fees were supported by the UKRI block grant for the University of Nottingham.

Acknowledgments

We would like to thank Michel Valstar from the School of Computer Science, University of Nottingham, for his input into the experimental design, as well as to Chris Scholes (School of Psychology, University of Nottingham) and David Watson (School of Psychology, University of York) for their general assistance. We would like to thank Fintan Nagle for constructing the PCA spaces and for allowing us to share the spaces through the Open Science Framework. Initial findings in the FFA from the first 7 participants were presented at the Applied Vision Association Christmas Meeting 2021 and the abstract published in i-Perception (Elson et al., 2022).

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/fpsyg.2022.1035524/full#supplementary-material

References

Aguirre, G. K., Singh, R., and D'Esposito, M. (1999). Stimulus inversion and the responses of face and object-sensitive cortical areas. *Neuroreport* 10, 189–194. doi: 10.1097/00001756-199901180-00036

Anderson, N. D., and Wilson, H. R. (2005). The nature of synthetic face adaptation. *Vis. Res.* 45, 1815–1828. doi: 10.1016/j.visres.2005.01.012

Andrews, T. J., Watson, D. M., Rice, G. E., and Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *J. Vis.* 15:3. doi: 10.1167/15.7.3

Babovic-Vuksanovic, D., Messiaen, L., Nagel, C., Brems, H., Scheithauer, B., Denayer, E., et al. (2012). Multiple orbital neurofibromas, painful peripheral nerve tumors, distinctive face and marfanoid habitus: a new syndrome. *Eur. J. Hum. Genet.* 20, 618–625. doi: 10.1038/ejhg.2011.275

Bao, P., She, L., McGill, M., and Tsao, D. Y. (2020). A map of object space in primate inferotemporal cortex. *Nature* 583, 103–108. doi: 10.1038/s41586-020-2350-5

Bar, M., Neta, M., and Linz, H. (2006). Very first impressions. *Emotion* 6, 269–278. doi: 10.1037/1528-3542.6.2.269

Besle, J., Sánchez-Panchuelo, R., Bowtell, R., Francis, S., and Schluppeck, D. (2013). Event-related fMRI at 7T reveals overlapping cortical representations for adjacent fingertips in S1 of individual subjects. *Hum. Brain Mapp.* 35, 2027–2043. doi: 10.1002/hbm.22310

Bona, S., Cattaneo, Z., and Silvanto, J. (2015). The causal role of the occipital face area (OFA) and lateral occipital (LO) cortex in symmetry perception. *J. Neurosci.* 35, 731–738. doi: 10.1523/JNEUROSCI.3733-14.2015

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Carbon, C.-C., and Leder, H. (2005). Face adaptation: changing stable representations of familiar faces within minutes? *Adv. Exp. Psychol.* 1, 1–7. doi: 10.5709/acp-0038-8

Carlin, J. D., and Kriegeskorte, N. (2017). Adjudicating between face-coding models with individual-face fMRI responses. *PLoS Comput. Biol.* 13:e1005604. doi: 10.1371/journal.pcbi.1005604

Chang, L., and Tsao, D. Y. (2017). The code for facial identity in the primate brain. *Cells* 169, 1013–1028.e14. doi: 10.1016/j.cell.2017.05.011

Davidenko, N., Remus, D. A., and Grill-Spector, K. (2012). Face-likeness and image variability drive responses in human face-selective ventral regions. *Hum. Brain Mapp.* 33, 2334–2349. doi: 10.1002/hbm.21367

Deng, W., Hu, J., Lu, J., and Guo, J. (2014). Transform-invariant PCA: a unified approach to fully automatic FaceAlignment, representation, and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 1275–1284. doi: 10.1109/TPAMI.2013.194

Dolci, C., Sansone, V. A., Gibelli, D., Cappella, A., and Sforza, C. (2021). Distinctive facial features in Andersen–Tawil syndrome: A three-dimensional stereophotogrammetric analysis. *Am. J. Med. Genet. A* 185, 781–789. doi: 10.1002/ajmg.a.62040

Elson, R., Schluppeck, D., Valstar, M., and Johnston, A. (2022). Taking face space to the extreme: assessing the effect of hyper-caricaturing faces on the fMRI response in the FFA. *Perception* 51, 354–364. doi: 10.1177/03010066221091992

Epstein, R. A., Higgins, J. S., Parker, W., Aguirre, G. K., and Cooperman, S. (2006). Cortical correlates of face and scene inversion: a comparison. *Neuropsychologia* 44, 1145–1158. doi: 10.1016/j.neuropsychologia.2005.10.009

Fischl, B. (2012). FreeSurfer. *Neuroimage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021

Freiwald, W. A., and Hosoya, H. (2021). Neuroscience: a Face's journey through space and time. *Curr. Biol.* 31, R13–R15. doi: 10.1016/j.cub.2020.10.065

Gad, A., Laurino, M., Maravilla, K. R., Matsushita, M., and Raskind, W. H. (2008). Sensorineural deafness, distinctive facial features, and abnormal cranial bones: a new variant of Waardenburg syndrome? *Am. J. Med. Genet. A* 146A, 1880–1885. doi: 10.1002/ajmg.a.32402

Gardner, J., Merriam, E., Schluppeck, D., Besle, J., and Heeger, D. (2018). mrTools: analysis and visualization package for functional magnetic resonance imaging data. *Zenodo.* doi: 10.5281/zenodo.1299483

Gardner, J. L., Sun, P., Waggoner, R. A., Ueno, K., Tanaka, K., and Cheng, K. (2005). Contrast adaptation and representation in human early visual cortex. *Neuron* 47, 607–620. doi: 10.1016/j.neuron.2005.07.016

Gilaie-Dotan, S., Gelbard-Sagiv, H., and Malach, R. (2010). Perceptual shape sensitivity to upright and inverted faces is reflected in neuronal adaptation. *NeuroImage* 50, 383–395. doi: 10.1016/j.neuroimage.2009.12.077

Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B. H., Marinkovic, K., and Rosen, B. R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Hum. Brain Mapp.* 7, 29–37. doi: 10.1002/(SICI)1097-0193(1999)7:1<29::AID-HBM3>3.0.CO;2-R

Hancock, P. J. B., Burton, A. M., and Bruce, V. (1996). Face processing: human perception and principal components analysis. *Mem. Cogn.* 24, 26–40. doi: 10.3758/BF03197270

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., and Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron* 22, 189–199. doi: 10.1016/S0896-6273(00)80690-X

Hill, M. Q., Parde, C. J., Castillo, C. D., Colón, Y. I., Ranjan, R., Chen, J.-C., et al. (2019). Deep convolutional neural networks in the face of caricature. *Nat. Mach. Intell.* 1, 522–529. doi: 10.1038/s42256-019-0111-7

James, T. W., Arcurio, L. R., and Gold, J. M. (2013). Inversion effects in face-selective cortex with combinations of face parts. *J. Cogn. Neurosci.* 25, 455–464. doi: 10.1162/jocn_a_00312

Jenkins, R., Dowsett, A. J., and Burton, A. M. (2018). How many faces do people know? *Proc. R. Soc. B Biol. Sci.* 285:20181319. doi: 10.1098/rspb.2018.1319

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Jiang, F., Blanz, V., and O'Toole, A. J. (2006). Probing the visual representation of faces with adaptation: a view from the other side of the mean. *Psychol. Sci.* 17, 493–500. doi: 10.1111/j.1467-9280.2006.01734.x

Jiang, F., Blanz, V., and O'Toole, A. J. (2007). The role of familiarity in three-dimensional view-transferability of face identity adaptation. *Vis. Res.* 47, 525–531. doi: 10.1016/j.visres.2006.10.012

Johnston, A., McOwan, P. W., and Benton, C. P. (1999). Robust velocity computation from a biologically motivated model of motion perception. *Proc. R. Soc. Lond. B Biol. Sci.* 266, 509–518. doi: 10.1098/rspb.1999.0666

Johnston, A., McOwan, P. W., and Buxton, H. (1992). A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proc. R. Soc. Lond. B Biol. Sci.* 250, 297–306. doi: 10.1098/rspb.1992.0162

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: 10.1523/JNEUROSCI.17-11-04302.1997

Kaufmann, J. M., and Schweinberger, S. R. (2012). The faces you remember: caricaturing shape facilitates brain processes reflecting the acquisition of new face representations. *Biol. Psychol.* 89, 21–33. doi: 10.1016/j.biopsycho.2011.08.011

Kleiner, M. (2007). What's new in psychtoolbox-3? *Perception* 36, 1–16. https://pure.mpg.de/rest/items/item_1790332/component/file_3136265/content (Accessed June 10, 2021).

Koyano, K. W., Jones, A. P., McMahon, D. B., Waidmann, E. N., Russ, B. E., and Leopold, D. A. (2021). Dynamic suppression of average facial structure shapes neural tuning in three macaque face patches. *Curr. Biol.* 31, 1–12.e5. doi: 10.1016/j.cub.2020.09.070

Lafer-Sousa, R., Conway, B. R., and Kanwisher, N. G. (2016). Color-biased regions of the ventral visual pathway lie between face-and place-selective regions in humans, as in macaques. *J. Neurosci.* 36, 1682–1697. doi: 10.1523/JNEUROSCI.3164-15.2016

Lee, K., Byatt, G., and Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: testing the face-space framework. *Psychol. Sci.* 11, 379–385. doi: 10.1111/1467-9280.00274

Leopold, D. A., Bondar, I. V., and Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* 442, 572–575. doi: 10.1038/nature04951

Leopold, D. A., O'Toole, A. J., Vetter, T., and Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat. Neurosci.* 4, 89–94. doi: 10.1038/82947

Leppänen, J. M., and Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychol. Res.* 69, 22–29. doi: 10.1007/s00426-003-0157-2

Little, A. (2012). Adaptation to Antifaces and the perception of correct famous identity in an average face. *Front. Psychol.* 3:19. doi: 10.3389/fpsyg.2012.00019

Loffler, G., Yourganov, G., Wilkinson, F., and Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nat. Neurosci.* 8, 1386–1391. doi: 10.1038/nn1538

Mauro, R., and Kubovy, M. (1992). Caricature and face recognition. *Mem. Cognit* 20, 433–440. doi: 10.3758/BF03210927

McKone, E., Jeffery, L., Boeing, A., Clifford, C. W. G., and Rhodes, G. (2014). Face identity aftereffects increase monotonically with adaptor extremity over, but not beyond, the range of natural faces. *Vis. Res.* 98, 1–13. doi: 10.1016/j.visres.2014.01.007

Morris, J. S., Friston, K. J., Büchel, C., Frith, C. D., Young, A. W., Calder, A. J., et al. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain* 121, 47–57. doi: 10.1093/brain/121.1.47

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 383, 812–815. doi: 10.1038/383812a0

Nagle, F., Griffin, H., Johnston, A., and McOwan, P. (2013). "Techniques for mimicry and identity blending using morph space PCA" in *Computer Vision - ACCV 2012 Workshops Lecture Notes in Computer Science*. eds. J.-I. Park and J. Kim (Berlin, Heidelberg: Springer), 296–307.

Nasr, S., and Tootell, R. B. H. (2012). Role of fusiform and anterior temporal cortical areas in facial recognition. *NeuroImage* 63, 1743–1753. doi: 10.1016/j.neuroimage.2012.08.031

Olivola, C. Y., Eubanks, D. L., and Lovelace, J. B. (2014). The many (distinctive) faces of leadership: inferring leadership domain from facial appearance. *Leadersh. Q.* 25, 817–834. doi: 10.1016/j.leaqua.2014.06.002

Over, H., and Cook, R. (2018). Where do spontaneous first impressions of faces come from? *Cognition* 170, 190–200. doi: 10.1016/j.cognition.2017.10.002

Parvizi, J., Jacques, C., Foster, B. L., Withoft, N., Rangarajan, V., Weiner, K. S., et al. (2012). Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci.* 32, 14915–14920. doi: 10.1523/JNEUROSCI.2609-12.2012

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10, 437–442.

Puce, A., Allison, T., Asgari, M., Gore, J. C., and McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J. Neurosci.* 16, 5205–5215. doi: 10.1523/JNEUROSCI.16-16-05205.1996

Ramon, M., Caharel, S., and Rossion, B. (2011). The speed of recognition of personally familiar faces. *Perception* 40, 437–449. doi: 10.1068/p6794

Rangarajan, V., Hermes, D., Foster, B. L., Weiner, K. S., Jacques, C., Grill-Spector, K., et al. (2014). Electrical stimulation of the left and right human fusiform gyrus causes different effects in conscious face perception. *J. Neurosci.* 34, 12828–12836. doi: 10.1523/JNEUROSCI.0527-14.2014

Rhodes, G., Brennan, S., and Carey, S. (1987). Identification and ratings of caricatures: implications for mental representations of faces. *Cognit. Psychol.* 19, 473–497. doi: 10.1016/0010-0285(87)90016-8

Rhodes, G., and Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vis. Res.* 46, 2977–2987. doi: 10.1016/j.visres.2006.03.002

Rhodes, G., Zebrowitz, L. A., Clark, A., Kalick, S. M., Hightower, A., and McKay, R. (2001). Do facial averageness and symmetry signal health? *Evol. Hum. Behav.* 22, 31–46. doi: 10.1016/S1090-5138(00)00060-X

Rodríguez, J., Bortfeld, H., Rudomín, I., Hernández, B., and Gutiérrez-Osuna, R. (2009). The reverse-caricature effect revisited: familiarization with frontal facial caricatures improves veridical face recognition. *Appl. Cogn. Psychol.* 23, 733–742. doi: 10.1002/acp.1539

Rossion, B., and Taubert, J. (2019). What can we learn about human individual face recognition from experimental studies in monkeys? *Vis. Res.* 157, 142–158. doi: 10.1016/j.visres.2018.03.012

Schulz, C., Kaufmann, J. M., Walther, L., and Schweinberger, S. R. (2012). Effects of anticaricaturing vs. caricaturing and their neural correlates elucidate a role of shape for face learning. *Neuropsychologia* 50, 2426–2434. doi: 10.1016/j.neuropsychologia.2012.06.013

Sirovich, L., and Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A* 4, 519–524. doi: 10.1364/JOSAA.4.000519

Susilo, T., McKone, E., and Edwards, M. (2010). What shape are the neural response functions underlying opponent coding in face space? A psychophysical investigation. *Vision Res.* 50, 300–314. doi: 10.1016/j.visres.2009.11.016

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., and Young, A. W. (2013). Social inferences from faces: ambient images generate a three-dimensional model. *Cognition* 127, 105–118. doi: 10.1016/j.cognition.2012.12.001

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., and Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nat. Neurosci.* 6, 989–995. doi: 10.1038/nn1111

Tsao, D. Y., Moeller, S., and Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proc. Natl. Acad. Sci.* 105, 19514–19519. doi: 10.1073/pnas.0809662105

Turk, M., and Pentland, A. (1991). Eigenfaces for recognition. *J. Cogn. Neurosci.* 3, 71–86. doi: 10.1162/jocn.1991.3.1.71

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol. A* 43, 161–204. doi: 10.1080/14640749108400966

Valentine, T., Lewis, M. B., and Hills, P. J. (2016). Face-space: a unifying concept in face recognition research. *Q. J. Exp. Psychol.* 69, 1996–2019. doi: 10.1080/17470218.2014.990392

Webster, M. A., and MacLeod, D. I. A. (2011). Visual adaptation and face perception. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 1702–1725. doi: 10.1098/rstb.2010.0360

Weibert, K., Flack, T. R., Young, A. W., and Andrews, T. J. (2018). Patterns of neural response in face regions are predicted by low-level image properties. *Cortex* 103, 199–210. doi: 10.1016/j.cortex.2018.03.009

Willis, J., and Todorov, A. (2006). First impressions: making up your mind after a 100-Ms exposure to a face. *Psychol. Sci.* 17, 592–598. doi: 10.1111/j.1467-9280.2006.01750.x

Yetter, M., Robert, S., Mammarella, G., Richmond, B., Eldridge, M. A., Ungerleider, L. G., et al. (2021). Curvilinear features are important for animate/inanimate categorization in macaques. *J. Vis.* 21:3. doi: 10.1167/jov.21.4.3

Yovel, G., and Freiwald, W. A. (2013). Face recognition systems in monkey and human: are they the same thing? *F1000prime Rep.* 5:10. doi: 10.12703/P5-10

Yovel, G., and Kanwisher, N. (2004). Face perception: domain specific, not process specific. *Neuron* 44, 889–898. doi: 10.1016/j.neuron.2004.11.018

Yovel, G., and Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Curr. Biol.* 15, 2256–2262. doi: 10.1016/j.cub.2005.10.072

Yue, X., Robert, S., and Ungerleider, L. G. (2020). Curvature processing in human visual cortical areas. *NeuroImage* 222:117295. doi: 10.1016/j.neuroimage.2020.117295

Zachariou, V., Del Giacco, A. C., Ungerleider, L. G., and Yue, X. (2018). Bottom-up processing of curvilinear visual features is sufficient for animate/inanimate object categorization. *J. Vis.* 18:3. doi: 10.1167/18.12.3

Zhu, Z., Luo, P., Wang, X., and Tang, X. (2013). Deep learning identity-preserving face space. *Presented at the Proceedings of the IEEE International Conference on Computer Vision*, pp. 113–120. https://www.cv-foundation.org/openaccess/content_iccv_2013/html/Zhu_Deep_Learning_Identity-Preserving_2013_ICCV_paper.html (Accessed April 26, 2022).

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 12 January 2023
DOI 10.3389/fpsyg.2022.953389

Check for updates

OPEN ACCESS

EDITED BY
Chihiro Saegusa,
Kao Corporation,
Japan

REVIEWED BY
Yuiko Sakuta,
Jissen Women's University, Japan
Ayahito Ito,
Kochi University of Technology, Japan

*CORRESPONDENCE
Miki Kamatani
✉ mmkamatani@eis.hokudai.ac.jp
Jun I. Kawahara
✉ jkawa@let.hokudai.ac.jp

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

RECEIVED 26 May 2022
ACCEPTED 12 December 2022
PUBLISHED 12 January 2023

CITATION
Kamatani M, Miyazaki Y and Kawahara JI (2023) Occlusion of faces by sanitary masks improves facial attractiveness of other races. *Front. Psychol.* 13:953389.
doi: 10.3389/fpsyg.2022.953389

Occlusion of faces by sanitary masks improves facial attractiveness of other races

Miki Kamatani[1]*, Yuki Miyazaki[2] and Jun I. Kawahara[1]*

[1]Faculty of Humanities and Human Sciences, Hokkaido University, Sapporo, Japan, [2]Department of Psychology, Fukuyama University, Hiroshima, Japan

Recent studies provide mixed results regarding whether the perception of facial attractiveness is increased or decreased by partial occlusion with a sanitary mask. One set of studies demonstrated that occluding the bottom half of a face increased facial attractiveness. This effect is thought to occur because the occluded area is interpolated by an average facial representation that is perceived as attractive. However, several groups of studies showed that partial occlusion can increase or decrease perceived attractiveness depending on the attractiveness of the original (unoccluded) face, due to regression to the mean. To reconcile this inconsistency, we propose that the occluded area is interpolated not by an average facial representation, but by a template of moderate attractiveness, shaped by the distribution of each viewer's experience. This hypothesis predicts an interaction between occlusion and the attractiveness of the original face so that occluded attractive faces are rated as less attractive, while occluded unattractive faces are rated as more attractive. To examine this hypothesis, the present study used attractiveness-rating tasks with mask-free versus masked faces in own-race and other-races categories. Viewers were familiar with own-race faces and unfamiliar with other-races faces. If moderate-attractiveness interpolation were the explanatory factor, the interaction between the occlusion and the attractiveness of the original face should be found only in the rating of own-race faces. Consistent with this hypothesis, the interaction between the occlusion and the attractiveness of the original faces was significant only for the own-race faces. Specifically, wearing a sanitary mask decreased the facial attractiveness of attractive faces in the own-race, while it increased the attractiveness regardless of the level of facial attractiveness in other-races. These findings suggest that the occluded area of own-race faces is interpolated by a facial template of moderate attractiveness. The other-races template could be developed using familiar exemplars such as celebrities. Thus, interpolation by such a template should result in elevated attractiveness relative to that by an own-race template. Accordingly, the apparent inconsistency in the literature regarding the effect of partial occlusion on physical attractiveness can be explained in terms of differences in the template involving interpolation of the occluded area.

KEYWORDS

sanitary mask, COVID-19, own-race, other-race, attractiveness, experience

1. Introduction

Faces convey a large amount of information during human interactions. The COVID-19 pandemic resulted in an increase in the use of sanitary masks worldwide. Face masks offer protection against disease by blocking the inhalation and exhalation of respiratory virus particles (e.g., Asadi et al., 2020; Chu et al., 2020; Li et al., 2021). However, they are disadvantageous in terms of interpersonal cognition and communication. For example, sanitary masks impede identification of individuals (Carragher and Hancock, 2020; Freud et al., 2020; Noyes et al., 2021) and recognition of emotional expressions (Carbon, 2020; Grundmann et al., 2021; Noyes et al., 2021; Parada-Fernández et al., 2022), and also affect the perceived facial attractiveness of the masked individual (Miyazaki and Kawahara, 2016; Patel et al., 2020; Kamatani et al., 2021; Hies and Lewis, 2022). These disadvantages are attributable to occlusion of the lower half of the face by the mask.

A set of recent studies revealed a systematic effect of occlusion of faces by masks on the perception of facial attractiveness. Specifically, a recent study (Patel et al., 2020) revealed that masked faces were perceived to be more attractive than the original faces. Participants viewed both a mask-free face and a masked face and rated the attractiveness on a scale of 1 (least attractive) to 10 (most attractive). The results revealed that masked faces were perceived as more attractive compared to mask-free faces when the original faces were unattractive or average in terms of attractiveness. Hies and Lewis (2022) also reported improvements in scores for masked faces in comparison to those that were mask-free. A similar effect was found to demonstrate that occluded faces were generally rated as more attractive than unoccluded faces (Sadr and Krowicki, 2019; Orghian and Hidalgo, 2020). Orghian and Hidalgo (2020) suggest that attractiveness is enhanced by an interpolation process. Some studies suggest that the average face is not the most attractive (e.g., Alley and Cunningham, 1991; Etcoff et al., 2011). However, given that facial averageness reflects developmental stability, and heterozygosity provides advantages for mating, it is intuitive that average faces created by morphing multiple images are perceived as attractive (e.g., Langlois and Roggman, 1990; Grammer and Thornhill, 1994; Rhodes and Tremewan, 1996; Apicella et al., 2007). Thus, by interpolating the occluded area with a representation of an average face, the resulting composite should be perceived more attractive than the unoccluded original. However, recent findings showed that this was not always the case.

A different group of researchers found an occlusion effect, but it was affected by the attractiveness of the original faces (Miyazaki and Kawahara, 2016; Kamatani et al., 2021). Specifically, Kamatani et al. (2021) found that the wearing of a sanitary mask increased the facial attractiveness of masked unattractive faces in comparison to mask-free faces. However, the mask also decreased the attractiveness of attractive faces. This effect of wearing masks occurs because the occlusion reduces the cues contributing to facial attractiveness. For example, a mask can occlude features related to low attractiveness such as asymmetric facial contours, misaligned or distorted facial features (e.g., Rhodes et al., 1998, 1999; Scheib et al., 1999; Little and Jones, 2003), and pimples and scars (e.g., Jaeger et al., 2018). Thus, negative ratings may trend toward the mean value. On the contrary, a mask on an attractive face can occlude features related to high attractiveness such as symmetric facial contours and smooth skin; thus, positive ratings may also trend toward the mean value.

Although the effect of occlusion by sanitary masks on perceived facial attractiveness, as reviewed above, has been reported (Miyazaki and Kawahara, 2016; Patel et al., 2020; Kamatani et al., 2021; Hies and Lewis, 2022), the occlusion effects of sanitary masks were inconsistent across studies. Importantly, Patel et al. (2020) and Hies and Lewis (2022) demonstrated a general improvement in attractiveness ratings by occlusion regardless of the attractiveness of the unmasked faces, whereas Kamatani et al. (2021) and Miyazaki and Kawahara (2016) found an interaction in that occlusion decreased the attractiveness of attractive faces and increased that of unattractive faces. This interaction cannot be explained by the same interpolation of an average face explaining why a general improvement in facial attractiveness is observed with wearing a mask. Likewise, the general improvement in attractiveness cannot be explained by a reduction of perception of cues of facial attractiveness.

We argue that the mixed results are attributable to the occluded area of a masked face being interpolated not by the representation of an average face, perceived as attractive, but rather by the representation of a face with moderate attractiveness. The findings of Miyazaki and Kawahara (2016) and Kamatani et al. (2021) support this idea. In their study, the participants and individuals presented in the stimuli were Japanese. The participants would develop a fine-tuned facial template representing a Japanese face of moderate attractiveness (Zhou et al., 2016). Instead of developing a single template of an averaged image, accumulated statistical evidence indicated that the participants saw moderately attractive faces most frequently. They were shown an ample number of moderately attractive features, including those with symmetric contours, typical global arrangements of facial parts, and smooth surface textures. In other words, viewing multiple faces rated as moderately attractive yielded statistical scores reflecting moderate attractiveness (scores of 40–50 on a scale of 1–100), rather than high attractiveness (scores of 90–100). This does not contradict previous research indicating that averaged faces are perceived as attractive. Instead, viewing multiple faces rated 40–50 on the attractiveness scale, i.e., viewing moderately attractive faces more frequently, results in the viewer imagining moderately attractive (ratings of 40–50) faces (or facial features) when asked to imagine or interpolate a face. This is distinct from the idea that viewing multiple faces leads to a single merged image. The participants could create such a template through observing faces of people in own-race group (Zhou et al., 2016). Therefore, when they encounter a masked face, they might have interpolated the occluded half using their moderate-attractiveness template. The use of a moderate-attractiveness template is reasonable in this case because it would represent the most statistically probable samples and has been shown to reduce modification error costs between predictions and real faces when the faces were revealed. However, this would not be likely to be the case when participants in previous studies were assumed to be recruited from diverse populations (Patel et al., 2020) and the stimuli contained racially heterogeneous faces (Patel et al., 2020; Hies and Lewis, 2022). In those cases, the participants would be expected to create a moderate-attractiveness template based on their experience of individuals with the same and different racial backgrounds. They would be likely to accumulate fine-tuned facial representations of individuals with the same racial background, but less likely to create a moderate-attractiveness template of the same quality for those of a different race, because they would be less sensitive to the fine details of faces of people of different races (Cross-Race effect; Apicella et al., 2007; Hourihan et al., 2012). Instead, they might have aggregated available knowledge from memory or been biased by other-race representatives such as celebrities (e.g., movie stars, athletes, and politicians). Consequently, the resulting viewer's moderate-attractiveness template could be biased toward attractive faces. Therefore, when the participants encountered a masked face, they may have interpolated the occluded half using a template biased toward attractive faces, resulting in an increase in the perception of attractiveness.

This idea is depicted in Figure 1. In this figure, the ordinate is frequency and the abscissa is facial attractiveness. The left two panels represent distributions of facial attractiveness of own-race (Figure 1A) and other-race (Figure 1B) individuals in the real world. Specifically (Figure 1A) represents the "real-world facial attractiveness" as a hypothetical dataset containing accumulated measurements of attractiveness scores rated by people in a given race group and (Figure 1B) represents "real-world facial attractiveness" as rated by people in a different race group. In Figure 1C shows a viewer's internal representation of the facial-attractiveness distribution of facial attractiveness generated *via* their own experience so that they see moderate attractive faces most frequently. The filled circles represent the encountered individuals. In Figure 1C, a viewer sees own-race faces of moderate attractiveness most frequently and encounters extremely attractive faces very rarely (extremely unattractive faces, as well). According to the Law of Large Numbers, the range of moderately attractive faces that the viewer conceives approximates to the range of moderate attractiveness in the real-world population. However, this does not apply to the faces of people of other races (Figure 1D). Viewers in environments with limited opportunities to see other-race faces would encounter a biased sample of attractive faces, such as movie stars, athletes, and politicians who are rated as highly attractive. This is because the number of samples of other-race faces that the viewer encounters is very small and biased toward attractive faces. The resulting distribution is skewed toward the highly attractive direction (right end of abscissa in Figure 1D), increasing the level of moderate attractiveness that functions as the viewer's standard.[1]

1 This can be analogously described in terms of height. We have a relatively accurate conception of the height distribution among people of our own race, but our perception of the height distribution among people of other races is likely to be influenced by images of actors or athletes, leading us to imagine higher average heights.

To examine this interpolation as explained by the moderate interpolation hypothesis, that is, the attractiveness rating of a masked face depends on its grouping (own-race or other-races), we established an experimental design in the present study that compares the attractiveness ratings of mask-free vs. masked faces between an own-race (Experiment 1: Japanese participants rated Japanese female faces) and other-races (Experiments 2–4: Japanese participants rated Black, White, and non-Japanese Asian female faces, in separate experiments). We focused on how the face template used for interpolation differs based on race. Sanitary masks affected attractiveness ratings for male faces similarly to female faces (Miyazaki and Kawahara, 2016; Hies and Lewis, 2022; Pazhoohi and Kingstone, 2022, Experiment 1). Therefore, we considered that only female faces used as the face stimuli were sufficient to examine the hypothesis of the present study. If the participants interpolated occluded faces based on a moderate-attractiveness template, masked other-races faces should be perceived as more attractive regardless of the baseline attractiveness because the moderate-attractiveness template of other-races is biased toward high attractiveness. Therefore, we predicted that the masked faces of other-races should be perceived as more attractive than the mask-free faces of those races. On the contrary, if the moderate-attractiveness template is involved in the interpolation process for masked faces in the own-race, the rating scores should reflect an interaction with the baseline attractiveness (i.e., improvement for unattractive faces and impairment for attractive faces).

There are two unresolved issues. First, the present model, informed by a moderate interpolation hypothesis, is based on the frequency with which viewers encounter faces of other races. This issue was addressed using a survey designed to validate such encounters. Second, the interpretation may not be consistent. The results of attractiveness ratings of masked other-races faces being greater than mask-free faces cannot be interpreted as an improvement in attractiveness due to the wearing of masks on those faces. Instead, another possibility is that those mask-free faces might be rated less attractive (i.e., devalued) due to an increase in exclusive attitudes toward persons from abroad as the COVID-19 pandemic expands (Yamagata et al., 2020), probably because individuals without masks are seen as a threat to anti-infection policies. For example, Yamagata et al. (2020) found an association between infection-avoidance tendencies and strong exclusionary attitudes towards persons from abroad. Therefore, we included the perceived vulnerability to disease (PVD) scale to examine whether the attitude of infection avoidance was related to negative ratings for mask-free faces. The PVD scale includes 15 items grouped into two internally consistent subscales, perceived infectability and germ aversion. The perceived infectability subscale assesses beliefs related to one's own susceptibility to infectious disease and the germ aversion subscale assesses beliefs related to one's emotional discomfort toward a high potential for pathogen transmission (Duncan et al., 2009; Fukukawa et al., 2014). We predicted that concern

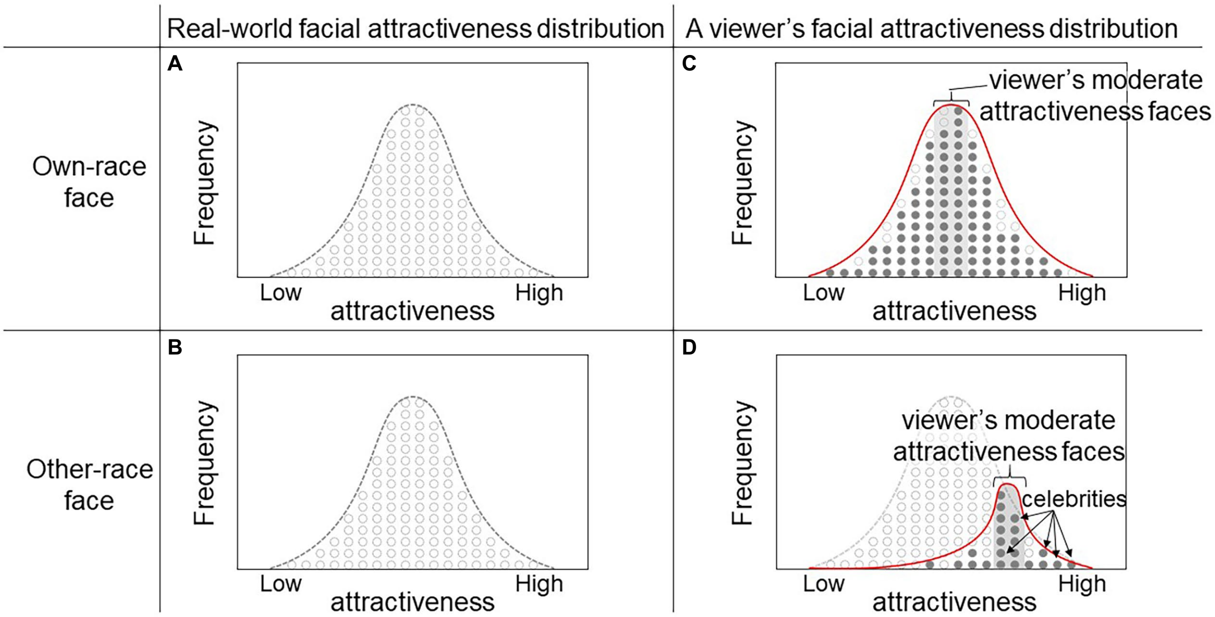


FIGURE 1
Distribution of real-world own-race **(A)** and other-race **(B)** facial attractiveness ratings. **(C)** Distribution (red solid line) of facial attractiveness ratings based on the viewer's own experience. The ordinate represents the frequency and the abscissa represents facial attractiveness ratings. Filled circles represent encountered individuals. Own-race moderately attractive faces are viewed most frequently. **(D)** The distribution (red solid line) of other-race facial attractiveness ratings is skewed toward high attractiveness because the viewer's experience is limited to attractive faces, such as celebrities. A viewer sees other-race highly attractive faces more often than moderately attractive faces.

about infectious diseases would be negatively related to the perceived attractiveness of mask-free other-races faces. Here, people who report higher PVD scores should perceive mask-free other-races faces as less attractive relative to others who report lower PVD scores.

2. Materials and methods

2.1. Participants

One hundred and fifty-six Japanese undergraduate and graduate students participated. They were assigned to the aforementioned four experiments (Experiment 1, $N=60$, $M=18.37$ years, $SD=0.63$, 31 females; Experiment 2, $N=32$, $M=20.50$ years, $SD=2.29$, 15 females; Experiment 3, $N=32$, $M=18.50$ years, $SD=0.82$, 19 females; Experiment 4, $N=32$, $M=18.78$ years, $SD=1.65$, 21 females). Participants in Experiment 2 were recruited from a participant pool at Hokkaido University. Experiments 1, 3 and 4 were implemented as coursework at the Health Sciences University of Hokkaido. Thus, the number of study participants depended on class attendance. The number of participants in Experiment 2 was equal to that in Experiments 3 and 4. All participants provided informed consent. No one participated in more than one experiment. The experimental protocol of the present study was approved by the ethical review board of Hokkaido University (31–2, 3–02). The experiments were conducted from June 26, 2020 until September 10, 2020, and on July 2, 2021.

2.2. Apparatus and stimuli

Stimuli were presented in a web browser of each participant's personal computer in Experiments 1, 3, and 4, and laboratory computer in Experiment 2 and were controlled by lab.js software (Henninger et al., 2022). Therefore, the size of the presented stimuli differed across experiments and participants, although no participants completed the experiment using their smartphones and tablets.

Different sets of faces images were used across experiments. For Experiment 1, Japanese female facial images [354 × 472 pixels; 8.1 cm (W) × 10.8 cm (H) on a 14-in laptop screen, color JPEG format] were selected from a homemade database of young female Japanese facial images (Kawahara and Kitazaki, 2013). For Experiments 2–4, Black, White, and non-Japanese Asian facial images [2,444 × 1,718 pixels; 11.0 cm (W) × 7.7 cm (H) on a 14-in laptop screen, color JPEG format] were selected from the Chicago Face Database (Ma et al., 2015)[2] based on attractiveness ratings obtained for 1,087 participants (516 Whites, 117 Asians, 74 Blacks, 72 bi- or multi-racial, 57

2 http://www.chicagofaces.org/

Latinos, 18 other, and 233 unreported races). Image size varied among the databases, so the image sizes in Experiment 1 were different from those in Experiments 2–4. A total of 48 images were selected from each image set so that 16 images were included in three attractiveness categories (low, middle, and high). All face images included the hair and had a neutral expression or slight smile. Both databases include slightly smiling faces, the frequency of which did not vary according to attractiveness. Therefore, it is unlikely that the attractiveness ratings were reduced as a result of shielding, which could have prevented positive facial expressions from being distinguished. Furthermore, a previous study showed the same results for both facial image databases with slightly smiling faces, and for a control database of faces with no expressions (Miyazaki and Kawahara, 2016, Studies 1a and 2). None of the faces wore eyeglasses. There were no lighting-related differences in facial shading between the two databases. The background was white for both databases; however, it was slightly darker for the Japanese dataset. The Chicago Face Database images were fully frontal, while some of the Japanese faces were at a slight angle. The procedure to create masked faces was identical to that in Miyazaki and Kawahara (2016). Namely, an image of a white sanitary mask was superimposed on the face. The superimposition and removal of unnatural edges between the face and mask was conducted using graphic editing software (Adobe Photoshop 2020). Figure 2 represents a sample image of masked and mask-free faces.

2.3. Procedure

In each experiment, a single race of facial images was presented: female Japanese faces for Experiment 1, female Black faces for Experiment 2, female White faces for Experiment 3, and female non-Japanese Asian faces for Experiment 4. In every trial, a single face was presented in the center of the screen, concurrently accompanied by a horizontal rating scale located below the image. Participants rated the attractiveness of the facial image from 1 (very unattractive) to 100 (very attractive) by moving the slider on the rating scale and clicked the submit button to report the score. The inter-trial interval was 500 ms.

Forty-eight facial images of different individuals were presented to each participant in random order. As shown in the Figure 2, half of the facial images were masked faces, and the others were mask-free faces. No participants viewed the same individual twice for both the mask-free and masked conditions. The presence or absence of masks was counterbalanced according to participant identity. The Japanese version of the PVD questionnaire (Fukukawa et al., 2014) was administered after the completion of the rating tasks of Experiments 2, 3, and 4. We did not administer the PVD questionnaire after Experiment 1 because we were not aiming to assess the relationship between the scale score and the rating score of attractiveness in the own-race condition. Moreover, our unpublished data found no correlations between PVD scores and attractiveness rating scores of masked and mask-free faces (Supplementary material A). Participants scored each item from 1 (strongly disagree) to 7 (strongly agree).

2.4. Statistical analysis

All analysis were conducted using R software (version 4.0.3.) and all Analysis of Variance (ANOVA) tests were performed using the anovakun function in R. In the analyses of correlations between PVD scores, included subscale scores, and the rating scores of mask-free and masked faces, five participants were excluded because they did not answer the PVD questionnaire. We computed the sum of scores for each participant.

3. Results

3.1. Attractiveness ratings

The mean rating scores were plotted as a function of baseline attractiveness for masked (solid line) and mask-free (dashed line) faces in each panel, as shown in Figure 3. A two-way (baseline attractiveness × mask presence) ANOVA with a within-participant design for the rating scores of Japanese faces (i.e., own-race faces) revealed a significant main effect of baseline attractiveness [$F(2,118) = 265.059$, $p < 0.001$, $\eta_p^2 = 0.817$]. Holm's test on this main effect revealed that the attractiveness rating was greater with an increase in baseline attractiveness [$ts(59) > 13.824$, $ps < 0.001$, $rs > 0.875$]. The main effect of mask presence was not significant [$F(1,59) = 0.842$, $p = 0.362$, $\eta_p^2 = 0.014$]. Importantly, the interaction between baseline attractiveness and mask presence was significant [$F(2,118) = 8.691$, $p < 0.001$, $\eta_p^2 = 0.128$]. Multiple comparisons of this interaction revealed that a sanitary mask decreases the facial attractiveness of faces with high attractiveness scores [$F(1,59) = 7.648$, $p = 0.007$, $\eta_p^2 = 0.114$]. For faces with low and middle attractiveness scores, the differences in perceived attractiveness between the masked and mask-free conditions were not significant [low: $F(1,59) = 3.524$, $p = 0.065$, $\eta_p^2 = 0.056$; middle: $F(1,59) = 0.323$, $p = 0.571$, $\eta_p^2 = 0.005$]. For Black, White, and non-Japanese Asian faces (i.e., other-races faces), the main effects of baseline attractiveness [$Fs(2,62) > 56.505$, $ps < 0.001$, $\eta_p^2s > 0.645$] and mask presence [$Fs(1,31) > 17.131$, $ps < 0.001$, $\eta_p^2 > 0.355$] were significant. Holm's tests revealed that the attractiveness rating was greater with an increase in baseline attractiveness [$ts(31) > 3.202$, $ps < 0.005$, $rs > 0.499$]. For other-races faces the interaction between baseline attractiveness and mask presence was not significant [$Fs(2,62) < 2.042$, $ps > 0.138$, $\eta_p^2s < 0.061$].

Overall, own-race facial attractiveness ratings decreased when a sanitary mask was worn. In contrast, facial attractiveness ratings for other races increased with sanitary masks, regardless of baseline

FIGURE 2
Examples of mask-free and masked faces. **(A)** Japanese females. **(B)** Black females. **(C)** White females. **(D)** Non-Japanese Asian females. The Japanese female faces (shown in panel **A**) were not used in the present study. Black, White, and non-Japanese Asian female faces shown in panels **B–D** were used in the present study.

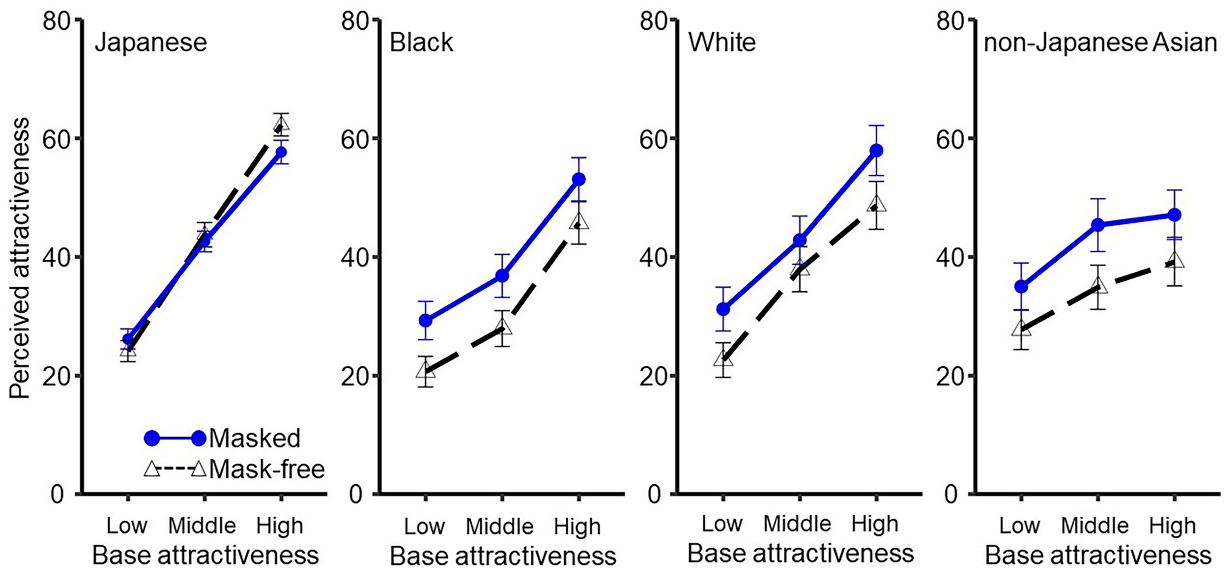


FIGURE 3
Attractiveness rating scores for masked and mask-free faces in each experiment. From left to right: Japanese, Black, White, non-Japanese Asian. Error bars represent standard errors.

facial attractiveness. The results for other races were identical to those obtained by rearranging the stimuli into three attractiveness categories for the 153 adults living in Japan (Supplementary material B). These results were consistent with our prediction based on the moderate interpolation hypothesis. Regarding the results of own-race facial attractiveness, the present study failed to demonstrate an exact replication of the findings of Kamatani et al. (2021), in that masked unattractive faces were perceived to be more attractive, whereas masked attractive faces were perceived to be less attractive. However, another study

(Morioka, 2022) found comparable interactions consistent with the results of Kamatani et al. (2021). Nonetheless, it is notable that the decrement of attractiveness in highly attractive faces was more pronounced relative to the increment of attractiveness in low-attractive faces. This pattern in the results is consistent with our previous data (Kamatani et al., 2021). Therefore, this asymmetry would be a common feature involved in this interaction.

3.2. Frequency of encounters with other-race faces

We conducted a survey to determine the frequency with which viewers encounter faces of other races. A group of 30 newly recruited Japanese undergraduate and graduate students participated ($M = 19.24$ years, $SD = 1.84$, 17 females). The participants never experienced Experiments 1, 2, 3, and 4. Stimuli were presented on a 24-inch LCD monitor (100-Hz refresh rate, 1920 × 1,080 pixels) and controlled by a Linux PC using MATLAB software (The MathWorks) with the Psychophysics Toolbox (Kleiner et al., 2007).

Forty facial images were used as stimuli in the present experiment; 10 were randomly selected from each race. Black, White, and non-Japanese Asian face images were resized to match the Japanese face images, i.e., 354 (W) × 472 (H) pixels. First, 10 Japanese face images were presented sequentially in the center of the screen at a rate of 3,000 ms/image, with a blank display of 500 ms between stimuli. A single other-race face was presented in the center of the screen, accompanied by a horizontal rating scale located below the image. Participants indicated the frequency of encounters with the different races on a scale ranging from 1 (never) to 100 (as frequently as Japanese faces). We conducted a one-sample t-test to analyze the mean race-specific scores. The results showed that the participants encountered individuals of other races infrequently (Black: $M = 26.253$, $t(29) = -30.206$, $p < 0.001$, $r = 0.985$; White: $M = 30.576$, $t(29) = -31.553$, $p < 0.001$, $r = 0.986$; non-Japanese Asian: $M = 51.073$, $t(29) = -15.391$, $p < 0.001$, $r = 0.944$). Since new participants were recruited, it was not possible to confirm whether the participants in Experiments 1–4 had less contact with people from other races. However, the results supported the moderate interpolation hypothesis. In particular, Japanese individuals had opportunities to form a detailed template for own-race faces, but lacked this opportunity for other races because of the low encounter frequency.

3.3. Correlations between the rating score and perceived vulnerability to disease scores

Cronbach's alpha for all 15 items was 0.74 (0.82: Duncan et al., 2009; 0.97: Fukukawa et al., 2014), Cronbach's alpha for the perceived infectability subscale was 0.86 (0.87: Duncan et al., 2009; 0.87 Fukukawa et al., 2014), and Cronbach's alpha for the germ aversion subscale was.75 (0.74: Duncan et al., 2009; 0.67 Fukukawa et al., 2014).

We calculated Pearson's correlation coefficients between the PVD score and rating scores of masked and mask-free faces. There was a significant positive correlation between the PVD score and the rating score for masked Black faces ($r = 0.365$, $p = 0.043$), whereas there was no significant correlation between the PVD score and the rating score for mask-free Black faces ($r = 0.317$, $p = 0.081$). For the White and non-Japanese Asian faces, there were no significant correlations between the PVD score and the rating scores of masked and mask-free faces (masked White: $r = 0.091$, $p = 0.623$; masked non-Japanese Asian: $r = 0.075$, $p = 0.696$; mask-free: White: $r = 0.043$, $p = 0.815$; mask-free non-Japanese Asian: $r = -0.064$, $p = 0.739$). For the perceived infectability and germ aversion subscale scores, we also computed Pearson's correlation coefficients. A positive correlation between the germ aversion score and the rating score for masked Black faces was significant ($r = 0.369$, $p = 0.040$). However, there was no significant correlation between the germ aversion score and the rating score for mask-free Black faces ($r = 0.298$, $p = 0.102$). Furthermore, there were no significant correlations between the germ aversion score and rating scores for masked and mask-free White and non-Japanese faces (masked White: $r = -0.046$, $p = 0.805$; mask-free White: $r = -0.043$, $p = 0.814$; masked non- Japanese Asian: $r = -0.184$, $p = 0.338$; mask-free non-Japanese Asian: $r = -0.157$, $p = 0.414$). No significant correlations were found between the score for perceived infectability and the attractiveness scores for other-races mask-free and masked faces (mask-free Black: $r = 0.163$, $p = 0.380$; masked Black: $r = 0.160$, $p = 0.388$; mask-free White: $r = 0.111$, $p = 0.549$; masked White: $r = 0.185$, $p = 0.316$; mask-free non-Japanese Asian: $r = 0.058$, $p = 0.763$; masked non-Japanese Asian: $r = 0.228$, $p = 0.233$). We show the graphs of the correlation results between the rating scores and PVD scores in Supplementary material C.

In general, there were no significant negative correlations between the PVD, germ aversion, and perceived infectability scores and the rating scores of mask-free other-races faces. These results were inconsistent with our prediction that people who report higher PVD scores should perceive mask-free other-races faces as less attractive relative to others who report lower PVD scores. Although the rating score for Black masked faces was positively correlated with the PVD and germ aversion scores, our replication study using a crowdsourcing service with the identical procedure as the original study ($N = 46$, 22 females, $M = 40.67$ years, $SD = 9.96$) did not identify significant positive correlations [PVD score: ρ $(46) = -0.035$, $p = 0.815$; germ aversion score: ρ $(46) = -0.038$, $p = 0.801$]. Therefore, care should be taken when interpreting these positive correlations.

4. Discussion

In the present study, we investigated the effects of two different factors on the attractiveness rating of masked faces. The first factor involved interpolation of the masked area. We predicted that the facial area occluded by a sanitary mask would be interpolated differently based on moderate-attractiveness templates between own-race and other-races faces. We argue that for own-race faces, respondents would have shaped a moderate-attractiveness template of good quality through experience and apply this template to interpolate the occluded area of the masked faces. Note that this moderate-attractiveness template differs from an averaged representation (Rhodes et al., 1987; Rhodes and Tremewan, 1996; Rhodes et al., 2007) of own-race faces that reflects developmental stability and heterozygosity, thus resulting in improvements in attractiveness when integrated into a whole-face representation. However, the moderate-attractiveness template represents a facial knowledge that is neither extremely attractive nor extremely unattractive but instead represents an average level of attractiveness. This results in an interaction with baseline attractiveness when the occluded area is interpolated using a moderate-attractiveness template. Specifically, interpolating masked attractive faces based on a moderate-attractiveness template should reduce attractiveness, whereas interpolating those of unattractive faces should increase attractiveness. This moderate-attractiveness interpolation process should not apply to other-races faces because we do not have counterpart templates for other-races faces due to lack of experience accumulated knowledge. Instead, the template used for interpolating other-races faces is biased toward highly attractive faces because of sparse exposure to other -races faces. Therefore, interpolation using this template should result in an improvement in attractiveness regardless of baseline attractiveness.

The second factor is an exclusive attitude toward the other-races. Because an exclusive attitude toward an other-races population is positively related to concern regarding the COVID-19 threat (e.g., Yamagata et al., 2020; Adam-Troian and Bagci, 2021; Mula et al., 2022), mask-free other-races faces should be rated less attractive if this factor is in effect. In particular, we predicted a negative correlation in that people who report higher PVD scores should perceive mask-free other-races faces as less attractive relative to others who report lower PVD scores, because of the association between infection-avoidance tendencies and strong exclusionary attitudes toward persons from abroad (Yamagata et al., 2020).

For the first interpolation-related factor, the present results supported the explanation of interpolation based on a moderate-attractiveness template and did not support that based on interpolation by an average representation. Specifically, mask-wearing did not simply improve perceived attractiveness. Rather, we found an interaction effect. Masked attractive own-race faces were rated less attractive than mask-free faces, whereas masked unattractive own-race faces tended to be rated more attractive than mask-free faces. By contrast, masked other-races faces were rated as more attractive than mask-free other-races faces regardless of the level of facial attractiveness. Importantly, the premise of the moderate interpolation hypothesis that the Japanese participants have opportunities to form a fine-detailed template for own-race faces, although they do not have such chance to train templates for other-races due to infrequent encounter to other-races faces was supported. Regarding the second factor of exclusive attitudes toward other-races, there were no negative correlations between the PVD, germ aversion, and perceived infectability scores and the rating scores for mask-free other-races faces. Overall, the race the faces belonged to was a primary factor that affected the attractiveness rating and the effect from the presence or absence of an exclusive attitude was negligible.

Although it is possible that differences between the two datasets influenced our results, we consider this unlikely because we conducted between-subjects experiments for each race and the participants could not compare images between the datasets. Moreover, the participants used different laptop computers for the web-based experiments and it was not feasible to control/match image quality for each race. Furthermore, there was no significant correlation between display size and attractiveness ratings in Experiments 3 and 4, which were conducted on the participants' personal computers (Supplementary material D). Our conclusions are based on the results obtained for images of varying quality that were available during the COVID-19 pandemic (when the study was conducted). The differences in image quality and presentation size among image sets were limitations of the present study, although it is unlikely that these differences alone could explain the findings. Ideally, image sets comprising photographs obtained under similar lighting conditions, and with similar resolutions, should have been used for rigorous testing of the effects of wearing masks for the different races.

The present findings supporting that the interpolation process is based on a moderate-attractiveness template can be bolstered by the following two theories. First, interpolation based on moderate attractiveness could explain the results of the present study better than the hypothesis relying on interpolation by average representation and regression to the mean. The moderate interpolation hypothesis posits that viewers shape their moderate-attractiveness template based on the distribution of their own experience and interpolate the occluded area based on this template. Specifically, when an occluded face belongs to the own-race, the distribution of the viewer's experience is similar to the real-world facial attractiveness distribution because the viewer experiences a variety of own-race faces. Therefore, unattractive occluded own-race faces are perceived as more attractive than the original faces, whereas attractive occluded own-race faces are perceived as less attractive than the original faces. When the occluded face is from an other-races, the distribution of the viewer's experience is biased toward highly attractive faces and is different from the real-world facial attractiveness distribution because their exposure is limited to attractive faces. Therefore, occluded faces are perceived as more attractive than the original faces regardless of baseline attractiveness. As mentioned above, moderate-attractiveness interpolation explains the results of the present study.

On the other hand, according to the average interpolation hypothesis stating that viewers interpolate the occluded area based on an average representation perceived as attractive, attractive occluded faces should be perceived as more attractive than the original faces, but this hypothesis is inconsistent with the present result for occluded attractive own-race faces. Moreover, with regression to the mean, occluded facial attractiveness is standardized by a decreasing cue for judging facial attractiveness, thus attractive occluded faces should be perceived as unattractive relative to the original faces, but this hypothesis is also inconsistent with the present results for occluded attractive other-races faces.

As a second theory, the moderate interpolation hypothesis may reduce the error between the perception of occluded faces and original faces. Living organisms, including humans, should behave in a way that minimizes the error between their own predictions and the external state as perceived through their sensory systems in their daily lives (e.g., Friston, 2010; Ohira, 2017). Therefore, the interpolation of a moderate-attractiveness template is reasonable because the template represents the statistically most probable samples and was advantageous in reducing modification errors between predictions and real faces when the faces were revealed.

Experience with facial categories (own- or other-race) could influence the development of a real-world moderate attractiveness template. It has been suggested that template formation depends on the viewer's experience. Apicella et al. (2007) reported that the Hadza people from Northern Tanzania did not prefer an average European face, but did prefer an average Hadza face. This preference for an average face of their own race was attributed to the fact that the Hadza were not sufficiently exposed to European faces to form an accurate impression regarding what constitutes an average European face. Furthermore, 5-year-old children were less influenced by average facial attractiveness than adults (Vingilis-Jaremko and Maurer, 2013). These findings suggest that there is no universal template for attractiveness; frequent encounters with different races are required to generate the average face template. In other words, increased encounters with a specific group of faces would generate knowledge. Encounters with particular types of other-race faces, such as those of movie stars and athletes, may lead the viewer to believe that those types of faces are common within the race in question. This results in a moderate-attractiveness template for other races biased toward highly attractive faces.

One limitation of the present study was the potential variation in facial attractiveness among image datasets. In particular, the highest attractiveness score for mask-free faces was 60 for own-race faces, with lower scores (around 30–50) seen for the other races. Previous studies have reported a preference for own-race faces over those of other races (Fisman et al., 2008; Liu et al., 2015; Rodway et al., 2019). Furthermore, own-race individuals are encountered more frequently than those of other races, which leads to the "Cross-Race effect," i.e., own-race faces can be discriminated more precisely than other-race faces. Therefore, attractiveness ratings for own-race faces are likely to vary more widely than those for other-race faces.

In conclusion, the present study examined the moderate interpolation hypothesis, and the results were consistent with the hypothesis. The hypothesis reconciles the mixed results of ratings of occluded facial attractiveness in previous studies. It should be noted that the present study only focused on Japanese participants. Therefore, future research should examine the moderate interpolation hypothesis for other races. For example, in a previous study, Cardiff University students perceived occluded faces, varied as to race, as being more attractive than unoccluded faces (Hies and Lewis, 2022). In another study, Dudarev et al. (2022) examined attractiveness ratings for masked White and Asian faces; they found that participants' attitudes toward mask-wearing were strongly associated with political orientation, thus affecting the judgment of attractiveness regarding masked faces. Given the variation in the progression of COVID-19 spread, the reactions to mask-wearing people differ considerably among countries. To avoid this contextual influence, it is important to occlude the face with a neutral object that does not evoke association with any attitudes (e.g., political attitude). Pursuing these research questions will contribute to a better understanding of the mechanism of why we observe occluded faces as more or less attractive.

Data availability statement

The datasets presented in this study can be found at: https://osf.io/5yzmh/.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethical Review Board of Hokkaido University (31–2; 3–02). The participants provided their written informed consent to participate in this study.

Author contributions

MK and YM processed the photographs and created stimuli. MK programmed the study and collected and analyzed the data. All authors designed the study and contributed to the manuscript text.

Funding

JK was funded by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (No. 20H04568). This work was partially supported by Graduate Grant Program of Graduate School of Humanities and Human

Sciences, Hokkaido University, Japan and JST SPRING, Grant Number JPMJSP2119 to MK.

Acknowledgments

We would like to thank Ryo Yoshioka for their help with the data collection.

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.953389/full#supplementary-material

References

Adam-Troian, J., and Bagci, S. C. (2021). The pathogen paradox: evidence that perceived COVID-19 threat is associated with both pro- and anti-immigrant attitudes. *Int. Rev. Soc. Psychol.* 34, 1–15. doi: 10.5334/irsp.469

Alley, T. R., and Cunningham, M. R. (1991). Article commentary: averaged faces are attractive, but very attractive faces are not average. *Psychol. Sci.* 2, 123–125. doi: 10.1111/j.1467-9280.1991.tb00113.x

Apicella, C. L., Little, A. C., and Marlowe, F. W. (2007). Facial averageness and attractiveness in an isolated population of hunter-gatherers. *Perception* 36, 1813–1820. doi: 10.1068/p5601

Asadi, S., Cappa, C. D., Barreda, S., Wexler, A. S., Bouvier, N. M., and Ristenpart, W. D. (2020). Efficacy of masks and face coverings in controlling outward aerosol particle emission from expiratory activities. *Sci. Rep.* 10:15665. doi: 10.1038/s41598-020-72798-7

Carbon, C. C. (2020). Wearing face masks strongly confuses counterparts in reading emotions. *Front. Psychol.* 11:566886. doi: 10.3389/fpsyg.2020.566886

Carragher, D. J., and Hancock, P. J. (2020). Surgical face masks impair human face matching performance for familiar and unfamiliar faces. *Cogn. Res. Principl. Implic.* 5, 59–15. doi: 10.1186/s41235-020-00258-x

Chu, D. K., Akl, E. A., Duda, S., Solo, K., Yaacoub, S., and Schünemann, H. J. (2020). Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *Lancet* 395, 1973–1987. doi: 10.1016/S0140-6736(20)31142-9

Dudarev, V., Kamatani, M., Miyazaki, Y., Enns, J. T., and Kawahara, J. I. (2022). The attractiveness of masked faces is influenced by race and mask attitudes. *Front. Psychol.* 13:864936. doi: 10.3389/fpsyg.2022.864936

Duncan, L. A., Schaller, M., and Park, J. H. (2009). Perceived vulnerability to disease: development and validation of a 15-item self-report instrument. *Personal. Individ. Differ.* 47, 541–546. doi: 10.1016/j.paid.2009.05.001

Etcoff, N. L., Stock, S., Haley, L. E., Vickery, S. A., and House, D. M. (2011). Cosmetics as a feature of the extended human phenotype: modulation of the perception of biologically important facial signals. *PLoS One* 6:e25656. doi: 10.1371/journal.pone.0025656

Fisman, R., Iyengar, S. S., Kamenica, E., and Simonson, I. (2008). Racial preferences in dating. *Rev. Econ. Stud.* 75, 117–132. doi: 10.1111/j.1467-937X.2007.00465.x

Freud, E., Stajduhar, A., Rosenbaum, R. S., Avidan, G., and Ganel, T. (2020). The COVID-19 pandemic masks the way people perceive faces. *Sci. Rep.* 10:22344. doi: 10.1038/s41598-020-78986-9

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Fukukawa, Y., Oda, R., Usami, H., and Kawahito, J. (2014). Development of a Japanese version of the perceived vulnerability to disease scale (in Japanese with English abstract). *Jpn. J. Psychol.* 85, 188–195. doi: 10.4992/jjpsy.85.13206

Grammer, K., and Thornhill, R. (1994). Human (Homo sapiens) facial attractiveness and sexual selection: the role of symmetry and averageness. *J. Comp. Psychol.* 108, 233–242. doi: 10.1037/0735-7036.108.3.233

Grundmann, F., Epstude, K., and Scheibe, S. (2021). Face masks reduce emotion-recognition accuracy and perceived closeness. *PLoS One* 16:e0249792. doi: 10.1371/journal.pone.0249792

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., and Hilbig, B. E. (2022). Lab.Js: a free, open, online study builder. *Behav. Res. Meth.* 54, 556–573. doi: 10.3758/s13428-019-01283-5

Hies, O., and Lewis, M. B. (2022). Beyond the beauty of occlusion: medical masks increase facial attractiveness more than other face coverings. *Cogn. Res. Prin. Implic.* 7:19. doi: 10.1186/s41235-021-00351-9

Hourihan, K. L., Benjamin, A. S., and Liu, X. (2012). A cross-race effect in metamemory: predictions of face recognition are more accurate for members of our own race. *J. Appl. Res. Mem. Cogn.* 1, 158–162. doi: 10.1016/j.jarmac.2012.06.004

Jaeger, B., Wagemans, F. M., Evans, A. M., and van Beest, I. (2018). Effects of facial skin smoothness and blemishes on trait impressions. *Perception* 47, 608–625. doi: 10.1177/0301006618767258

Kamatani, M., Ito, M., Miyazaki, Y., and Kawahara, J. I. (2021). Effects of masks worn to protect against COVID-19 on the perception of facial attractiveness. *Perception* 12:920. doi: 10.1177/20416695211027920

Kawahara, J., and Kitazaki, M. (2013). "The effect of variance in members' attractiveness on perceived group attractiveness." in *Proceedings of the ACM symposium on applied perception*. Trinity College, Dublin, Ireland: ACM, 141.

Kleiner, M., Brainard, D., and Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception* 36, 1–16.

Langlois, J. H., and Roggman, L. A. (1990). Attractive faces are only average. *Psychol. Sci.* 1, 115–121. doi: 10.1111/j.1467-9280.1990.tb00079.x

Li, Y., Liang, M., Gao, L., Ahmed, M. A., Uy, J. P., Cheng, C., et al. (2021). Face masks to prevent transmission of COVID-19: a systematic review and meta-analysis. *Am. J. Infect. Control* 49, 900–906. doi: 10.1016/j.ajic.2020.12.007

Little, A. C., and Jones, B. C. (2003). "Evidence against perceptual bias views for symmetry preferences in human faces," in *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270, 1759–1763. doi: 10.1098/rspb.2003.2445

Liu, S., Xiao, N. G., Quinn, P. C., Zhu, D., Ge, L., Pascalis, O., et al. (2015). Asian infants show preference for own-race but not other-race female faces: the role of infant caregiving arrangements. *Front. Psychol.* 6:593. doi: 10.3389/fpsyg.2015.00593

Ma, D. S., Correll, J., and Wittenbrink, B. (2015). The Chicago face database: a free stimulus set of faces and norming data. *Behav. Res. Methods* 47, 1122–1135. doi: 10.3758/s13428-014-0532-5

Miyazaki, Y., and Kawahara, J. I. (2016). The sanitary-mask effect on perceived facial attractiveness. *Jpn. Psychol. Res.* 58, 261–272. doi: 10.1111/jpr.12116

Morioka, Y. (2022). "*Influence of wearing sanitary-mask on perception of facial attractiveness after the COVID-19 pandemics* [poster presentation]," in *The 86th annual convention of the Japanese psychological association*. Tokyo, Japan: The Japanese psychological association, 3EV-054-PI.

Mula, S., Di Santo, D., Resta, E., Bakhtiari, F., Baldner, C., Molinario, E., et al. (2022). Concern with COVID-19 pandemic threat and attitudes towards immigrants: the mediating effect of the desire for tightness. *Curr. Res. Ecol. Soc. Psychol.* 3:100028. doi: 10.1016/j.cresp.2021.100028

Noyes, E., Davis, J. P., Petrov, N., Gray, K. L., and Ritchie, K. L. (2021). The effect of face masks and sunglasses on identity and expression recognition with super-recognizers and typical observers. *R. Soc. Open Sci.* 8:201169. doi: 10.1098/rsos.201169

Ohira, H. (2017). Predictive coding, interoception, and affect. (in Japanese with English abstract). *Emot. Stud.* 3, 2–12. doi: 10.20797/ems.3.1_2

Orghian, D., and Hidalgo, C. A. (2020). Humans judge faces in incomplete photographs as physically more attractive. *Sci. Rep.* 10, 110–112. doi: 10.1038/s41598-019-56437-4

Parada-Fernández, P., Herrero-Fernández, D., Jorge, R., and Comesaña, P. (2022). Wearing mask hinders emotion recognition, but enhances perception of attractiveness. *Personal. Individ. Differ.* 184:111195. doi: 10.1016/j.paid.2021.111195

Patel, V., Mazzaferro, D. M., Sarwer, D. B., and Bartlett, S. P. (2020). Beauty and the mask. *Plast. Reconstr. Surg. Glob. Open* 8:e3048. doi: 10.1097/GOX.0000000000003048

Pazhoohi, F., and Kingstone, A. (2022). Unattractive faces are more attractive when the bottom-half is masked, an effect that reverses when the top-half is concealed. *Cogn. Res.* 7, 6–10. doi: 10.1186/s41235-022-00359-9

Rhodes, G., Brennan, S., and Carey, S. (1987). Identification and ratings of caricatures: implications for mental representations of faces. *Cogn. Psychol.* 19, 473–497. doi: 10.1016/0010-0285(87)90016-8

Rhodes, G., Maloney, L. T., Turner, J., and Ewing, L. (2007). Adaptive face coding and discrimination around the average face. *Vis. Res.* 47, 974–989. doi: 10.1016/j.visres.2006.12.010

Rhodes, G., Proffitt, F., Grady, J. M., and Sumich, A. (1998). Facial symmetry and the perception of beauty. *Psychon. Bull. Rev.* 5, 659–669. doi: 10.3758/BF03208842

Rhodes, G., Sumich, A., and Byatt, G. (1999). Are average facial configurations attractive only because of their symmetry? *Psychol. Sci.* 10, 52–58. doi: 10.1111/1467-9280.00106

Rhodes, G., and Tremewan, T. (1996). Averageness, exaggeration, and facial attractiveness. *Psychol. Sci.* 7, 105–110. doi: 10.1111/j.1467-9280.1996.tb00338.x

Rodway, V., Tatham, B., and Guo, K. (2019). Effect of model race and viewing perspective on body attractiveness and body size assessment in young Caucasian women: an eye-tracking study. *Psychol. Res.* 83, 347–356. doi: 10.1007/s00426-018-1138-9

Sadr, J., and Krowicki, L. (2019). Face perception loves a challenge: less information sparks more attraction. *Vis. Res.* 157, 61–83. doi: 10.1016/j.visres.2019.01.009

Scheib, J. E., Gangestad, S. W., and Thornhill, R. (1999). "Facial attractiveness, symmetry and cues of good genes." in *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266, 1913–1917. doi: 10.1098/rspb.1999.0866

Vingilis-Jaremko, L., and Maurer, D. (2013). The influence of averageness on children's judgments of facial attractiveness. *J. Exp. Child Psychol.* 115, 624–639. doi: 10.1016/j.jecp.2013.03.014

Yamagata, M., Teraguchi, T., and Miura, A. (2020). The relationship between infection-avoidance tendencies and exclusionary attitudes toward foreigners: a panel study of the COVID-19 outbreak in Japan. *PsyArXiv.* doi: 10.31234/osf.io/x5emj

Zhou, X., Short, L. A., Chan, H. S., and Mondloch, C. J. (2016). Judging normality and attractiveness in faces: direct evidence of a more refined representation for own-race, young adult faces. *Perception* 45, 973–990. doi: 10.1177/0301006616652044