Check for updates

# Statistical context learning in tactile search: Crossmodally redundant, visuo-tactile contexts fail to enhance contextual cueing

Siyi Chen*, Zhuanghua Shi, Gizem Vural, Hermann J. Müller and Thomas Geyer

General and Experimental Psychology, Department of Psychology, Ludwig-Maximilians-Universität Munich, Munich, Germany

In search tasks, reaction times become faster when the target is repeatedly encountered at a fixed position within a consistent spatial arrangement of distractor items, compared to random arrangements. Such "contextual cueing" is also obtained when the predictive distractor context is provided by a non-target modality. Thus, in tactile search, finding a target defined by a deviant vibro-tactile pattern (delivered to one fingertip) from the patterns at other, distractor (fingertip) locations is facilitated not only when the configuration of tactile distractors is predictive of the target location, but also when a configuration of (collocated) visual distractors is predictive—where intramodal-tactile cueing is mediated by a somatotopic and crossmodal-visuotactile cueing by a spatiotopic reference frame. This raises the question of whether redundant multisensory, tactile-plus-visual contexts would enhance contextual cueing of tactile search over and above the level attained by unisensory contexts alone. To address this, we implemented a tactile search task in which, in 50% of the trials in a "multisensory" phase, the tactile target location was predicted by both the tactile and the visual distractor context; in the other 50%, as well as a "unisensory" phase, the target location was solely predicted by the tactile context. We observed no redundancy gains by multisensory-visuotactile contexts, compared to unisensory-tactile contexts. This argues that the reference frame for contextual learning is determined by the task-critical modality (somatotopic coordinates for tactile search). And whether redundant predictive contexts from another modality (vision) can enhance contextual cueing depends on the availability of the corresponding spatial (spatiotopic-visual to somatotopic-tactile) remapping routines.

## 1. Introduction

### 1.1. Contextual cueing in the individual modalities of vision and touch

Attention is guided by a number of separable mechanisms that can be categorized as bottom-up driven—such as guidance by salient physical properties of the current stimuli—or top-down controlled—such as guidance by observers' "online" knowledge about (search-) critical object properties (Wolfe and Horowitz, 2017). These processes are augmented by the

automatic extraction of statistical co-occurrences of objects in the environment, rendering attention-guiding spatial long-term (LT) memories. For instance, repeatedly encountering a searched-for target item at a particular location within a visual scene of consistently arranged distractor items leads to the formation of LT relational distractor-target memories, that, upon being activated by the currently viewed search display, (relatively) efficiently direct attentional scanning toward the target location (Goujon et al., 2015; Sisk et al., 2019). This effect was first described by Chun and Jiang (1998), who, in their seminal study, had participants search for a target letter "T" (left- or right-rotated) among a set of (orthogonally rotated) distractor letters "L". In half of the trials, the spatial arrangements of the distractor and target stimuli were repeated, permitting participants to learn the invariant distractor-target relations to guide their search (repeated/predictive displays). In the other half, the distractors were distributed randomly on each trial, rendering their arrangement non-predictive of the target's position in the search array (non-repeated/non-predictive displays). Chun and Jiang's (1998) critical finding was that the reaction times (RTs) taken to find and respond to the target were faster for repeated vs. non-repeated display arrangements or "contexts". This effect referred to as "contextual cueing", subsequently was confirmed and elaborated in a plethora of studies using behavioral, computational, and neuroscientific measures (Chun and Jiang, 1999; Chun, 2000; Shi et al., 2013; Zinchenko et al., 2020; Chen et al., 2021a). In the first instance, of course, effective contextual cueing requires successful retrieval of the respective (search-guiding) LT-memory representation. Thus, for example, when the time for which the spatial distractor-target layout can be viewed is limited (Zang et al., 2015) or when encoding of the display layout is hampered by competing visual task demands (Manginelli et al., 2013), the retrieval of acquired context memories may be prevented, abolishing contextual facilitation.

Interestingly, contextual cueing is not limited to the visual modality: tactile predictive contexts can facilitate search, too. For instance, Assumpção et al. (2015) showed that people can become better at finding an odd-one-out vibrotactile target within arrays of repeated vs. non-repeated (homogeneous) vibrotactile distractor items delivered to participants' fingertips (where the vibrotactile distractor-target arrangements consisted of two stimulated fingers, excepting the thumb, on each hand). As revealed by postural manipulations of the hands (Assumpção et al., 2018), tactile contextual cueing is rooted in a somatotopic reference frame: spatial target-distractor associations acquired during training transfer to a test phase (with crossed or flipped hands) only if the target and distractors are located at the same fingers, rather than the same external spatial locations. This finding implies that search in repeated vs. non-repeated tactile distractor-target arrangements evokes, in default mode, a somatosensory reference frame, which is different from (default) spatiotopic encoding of distractor-target relations in visual search (Chua and Chun, 2003). However, while the learning of statistical co-occurrences of target and distractor items is bound to the currently task-relevant sensory modality, the brain has the ability to adapt and reorganize connectivity between different sensory modalities in response to consistent changes in input or

experience—referred to as "crossmodal plasticity" (Bavelier and Neville, 2002; Nava and Röder, 2011). Thus, an interesting question arises, namely, whether the encoding of statistical regularities in one modality would facilitate search in another modality through the engagement of crossmodal-plasticity mechanisms. For instance, given that optimal task performance may depend on the use of all available sources of information, spatial learning in the tactile modality might be enhanced by congruent, redundant-signal information in the visual modality (Ho et al., 2009), and this may involve changes in the strength (and number) of connections between neurons in the visual and somatosensory regions of the brain. The possibility of such crossmodal spatial regularity/contextual learning is the question at issue in the current study.

## 1.2. Crossmodal contextual cueing across visual and tactile modalities

Initial evidence indicates that the mechanisms underlying contextual cueing may support the functional reorganization of one sensory modality following statistical learning in another modality. For example, Kawahara (2007) presented participants with meaningless speech sounds followed by a visual search display during a training phase. The location of the search was predictable from the preceding auditory stimulus. In the subsequent test phase, this auditory–visual association was either removed for one (inconsistent-transfer) group or maintained for another (consistent-transfer) group. The results revealed the search RTs to be increased for the inconsistent-transfer group but decreased for the consistent group—suggesting that visual attention can be guided implicitly by crossmodal association. In another study, Nabeta et al. (2003) had their participants first search for a T-type target among L-type distractors *visually* in a learning phase, which was followed by a test phase in which they had to search *haptically* for T- vs. L-shaped letters. The haptic search arrays (which were carved on wooden boards and covered by an opaque curtain) were arranged in the same or different configurations compared to the visual displays during initial visual learning. Nabeta et al. (2003) found that target-distractor contexts learned during visual search also facilitated haptic search in the absence of visual guidance. It should be noted, though, that Nabeta et al.'s haptic search involved active exploration, involving serial finger movements to sense the local items. Haptic search may thus have required participants to set up and continually update a visual working-memory representation of the scene layout, and the initially learned contexts may have come to interact with this representation, guiding the haptic exploration toward the target location. However, this would not work with tactile search scenarios involving spatially *parallel*, passive sensing, such as those explored by Assumpção et al. (2015, 2018). Passive tactile sensing and active manual exploration have been shown to involve distinct processes (Lederman and Klatzky, 2009). Accordingly, being based on active exploration, the findings of Nabeta et al. (2003) provide no clear answer as to whether and how target-distractor contexts
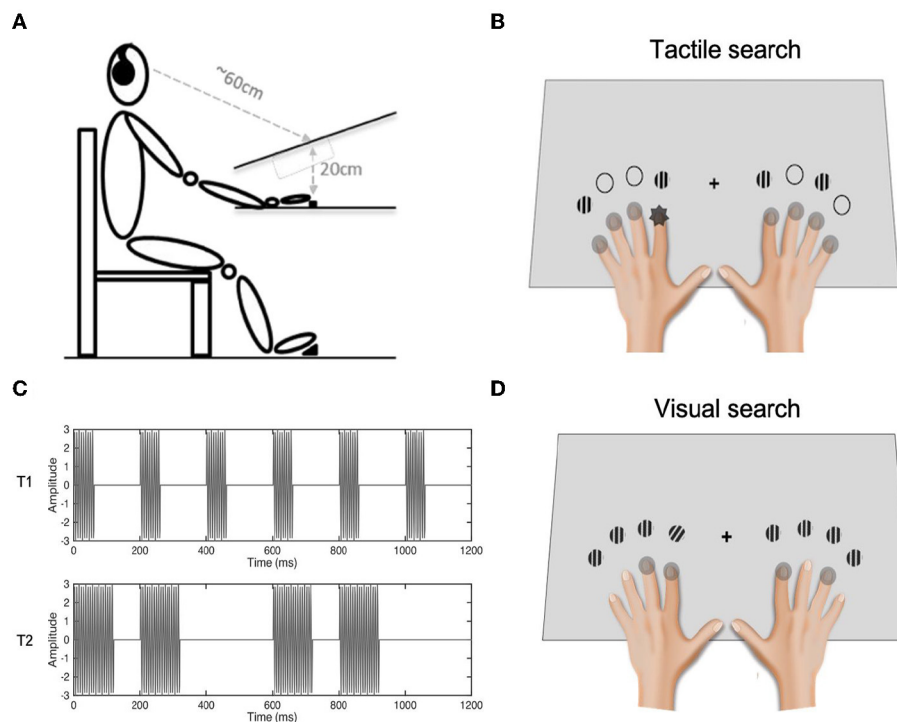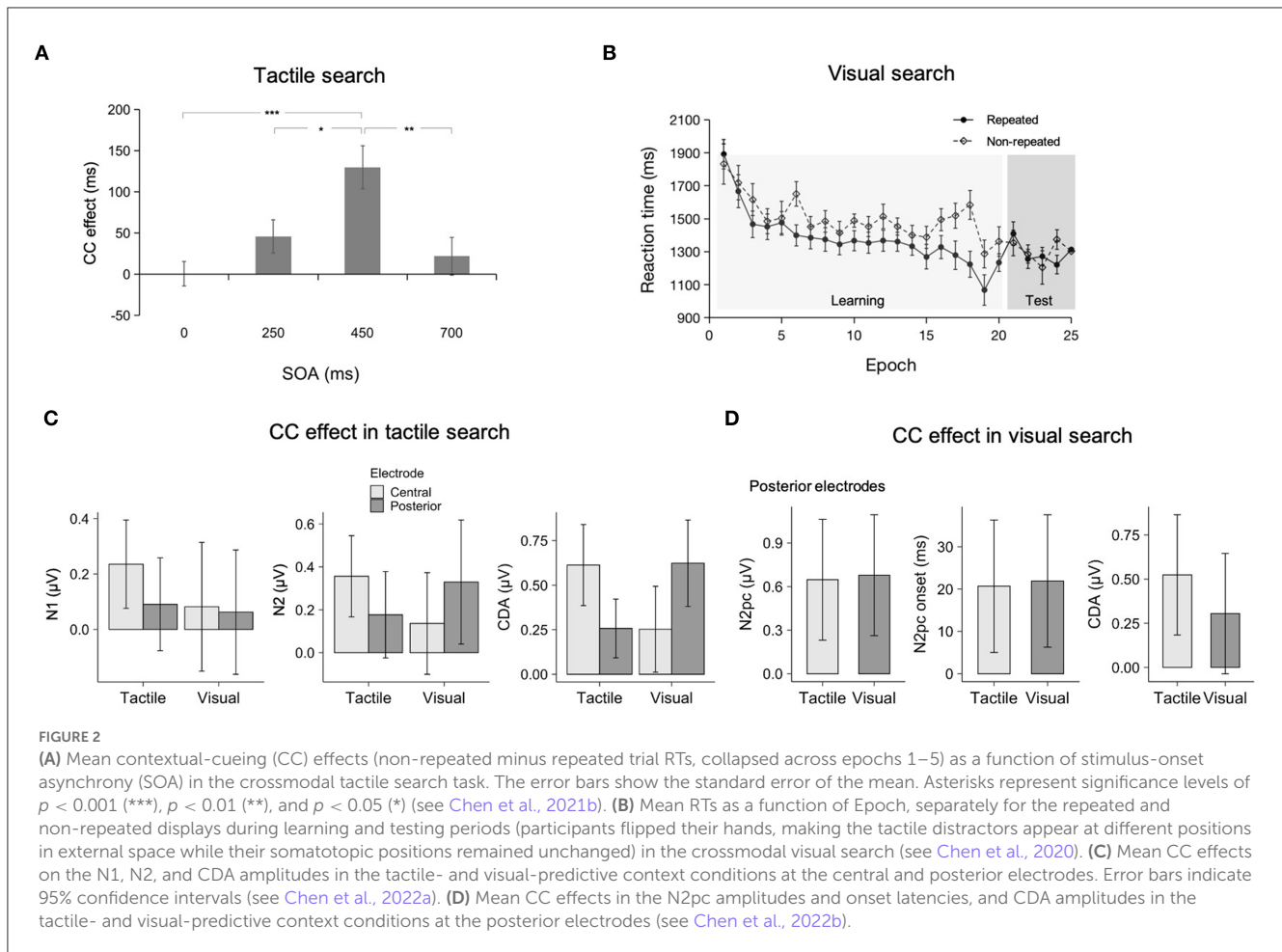
**FIGURE 1**
Illustration of the experimental set-up. As illustrated in Panel **(A)**, the height difference between the visual and tactile presentation planes was some 20 cm. Visual stimuli were presented on a white canvas surface tilted about 20° toward the observer. The viewing distance was 60 cm. Participants placed their fingers (except the thumbs) on the eight solenoid actuators and responded to the identity of the tactile singleton target *via* a designated foot pedal. Panel **(B)** depicts the visual-tactile stimuli in a tactile search task. The search display consisted of one tactile target (the dark "spark") with seven homogenous distractor vibrations (light gray circles), accompanied by a configuration of four distractor Gabor patches (and four empty circles). The locations of the tactile target for the tactile search and Gabor patches for the visual search varied depending on whether the displays were repeated or not. In the real setting, the hands were placed on the plane below the visual plane, as illustrated in Panel **(A)**. Panel **(C)** depicts the waveforms of the two possible tactile targets in a tactile search task. The upper panel depicts the waveform of target 1 (T1): a 5-Hz square wave with a 30% duty cycle delivered *via* 150-Hz vibrations. The lower panel illustrates the waveform of target 2 (T2): a 5-Hz square wave with an average 60% duty cycle, also composed of 150-Hz vibrations. The distractors were constant vibrations of 150 Hz. Panel **(D)** depicts the visual-tactile stimuli in a visual search task. One visual target was embedded among seven homogenous distractors, with a configuration of four vibrotactile stimulations delivered to two (selected) fingers (gray circles) of each hand.

acquired during visual search would transfer to parallel, passive tactile search.

Recently, Chen et al. (2021b) aimed to directly address this question by adopting a similar tactile-search paradigm to Assumpção et al. (2015, 2018), delivering vibrotactile stimulation to participants' fingertips instead of requiring active manual exploration. In addition, the visual search displays, projected on a white canvas on the top of the tactile array, were collocated with the tactile stimuli (Figures 1A, B). The visuotactile search arrays were constructed in such a way that only the visual configuration was predictive of the tactile target location (Figure 1C). Chen et al. (2021b) found that repeated visual contexts came to facilitate tactile search as the experiment progressed, but only if the tactile items were presented some 250–450 ms prior to the visual elements (Figure 2A). Chen et al. (2021b) attributed this tactile preview time to the need to recode the (somatotopically-sensed) tactile array in a visual reference frame, in order for a search to benefit from the predictive context provided by the visual distractor elements (sensed in spatiotopic format).

Using a similar visual-tactile setup (Figure 1A), Chen et al. (2020) investigated whether a predictive tactile context could

facilitate visual search. Participants had to search for a visual odd-one-out target, a Gabor patch differing in orientation (clockwise or counter-clockwise) from seven homogeneous vertical Gabor distractors (see Figure 1D). Critically, unbeknown to participants, visual targets were paired with repeated tactile contexts in half of the trials, and with newly generated tactile contexts in the other half. Again, the visual-tactile display onset asynchrony was varied. Similar to Chen et al. (2021b), the repeated tactile context had to be presented before the visual target in order for crossmodal contextual cueing to manifest—again suggesting that a preview time was required for remapping the somatotopically encoded tactile context into the visual spatiotopic reference frame in which the target is encoded. Of note, in a control experiment, Chen et al. (2020) found that under conditions in which participants flipped their hands, but the visual target and tactile distractors were kept unchanged with respect to somatotopic coordinates, the crossmodal contextual-cueing effect was diminished (Figure 2B). This supports the idea that, with multisensory presentations, the predictive tactile context was remapped into a spatiotopically organized visual—i.e., target-appropriate—format (Kennett et al., 2002; Kitazawa,

**FIGURE 2**
**(A)** Mean contextual-cueing (CC) effects (non-repeated minus repeated trial RTs, collapsed across epochs 1–5) as a function of stimulus-onset asynchrony (SOA) in the crossmodal tactile search task. The error bars show the standard error of the mean. Asterisks represent significance levels of $p < 0.001$ (***), $p < 0.01$ (**), and $p < 0.05$ (*) (see Chen et al., 2021b). **(B)** Mean RTs as a function of Epoch, separately for the repeated and non-repeated displays during learning and testing periods (participants flipped their hands, making the tactile distractors appear at different positions in external space while their somatotopic positions remained unchanged) in the crossmodal visual search (see Chen et al., 2020). **(C)** Mean CC effects on the N1, N2, and CDA amplitudes in the tactile- and visual-predictive context conditions at the central and posterior electrodes. Error bars indicate 95% confidence intervals (see Chen et al., 2022a). **(D)** Mean CC effects in the N2pc amplitudes and onset latencies, and CDA amplitudes in the tactile- and visual-predictive context conditions at the posterior electrodes (see Chen et al., 2022b).

2002; see also Azañón and Soto-Faraco, 2008; Heed et al., 2015).

But is the remapping process still helpful when predictive contexts are concurrently available in two sensory modalities? Recently, Chen et al. (2021a) investigated this issue by presenting redundant visual-tactile contexts intermixed with single visual contexts in a visual search task. Following Chen et al. (2020), the tactile context was presented 450 ms prior to the visual context to promote tactile-to-visual remapping. Interestingly, Chen et al. (2021a) found that contextual facilitation of search was increased with multisensory, i.e., visuotactile, contexts relative to predictive visual contexts alone—suggesting that multisensory experiences facilitate unisensory learning.

Taken together, previous studies (Chen et al., 2020, 2021a,b) investigating visual and tactile search in multisensory arrays consisting of visual and tactile items established that contextual cues available in one—distractor—modality can be utilized in the other—target—modality. Further, redundant contexts consisting of identically positioned visual and tactile elements can enhance visual learning of the relational position of the visual target item over and above that deriving from predictive visual contexts alone.

## 1.3. ERP evidence on crossmodal contextual cueing

Evidence for crossmodal cueing comes also from recent electrophysiological studies (Chen et al., 2022a,b). For example, when using the crossmodal search paradigm sketched in Figure 1, Chen et al. (2022a) found that in a tactile search task, facilitation of search RTs by repeated visual contexts was also seen in well-established electrophysiological markers of the allocation of visuospatial attention, in particular, the N2pc (Luck et al., 2000) and CDA (Töllner et al., 2013) measured at parietal-posterior ("visual") electrodes; however, the lateralized event-related potentials (ERPs) in the respective time windows were less marked at central ("somatosensory") electrodes (Figure 2C). In contrast, statistical learning of the unimodal (tactile) context led to enhanced attention allocation (indexed by the N1/N2cc/CDA) at central ("somatosensory") electrodes, whereas these effects were less prominent at posterior ("visual") electrodes. These findings indicate that both somatosensory and visual cortical regions contribute to contextual cueing of tactile search, but their involvement is differentially weighted depending on the sensory modality that contains the predictive context information. There is a stronger reliance on or weighting of, either a visual or somatotopic

coordinate frame depending on the currently available sensory regularities that support contextual cueing in tactile-visual search environments. Worth mentioning is also the work of Chen et al. (2021b), who observed that crossmodal (tactile) context learning in a visual search resulted in enhanced amplitudes (and reduced latencies) of the lateralized N2pc/CDA waveforms at posterior ("visual") electrodes (see Figure 2D); both components correlated positively with the RT facilitation. These effects were comparable to the unimodal (visual context) cueing conditions. In contrast, motor-related processes indexed by the response-locked LRP at central ("somatosensory") electrodes contributed little to the RT effects. This pattern indicates that the crossmodal-tactile context is encoded in a visual format for guiding visual search.

## 1.4. Goals of the current study

The studies reviewed thus far show that search is not "a-historic". Rather, LT-memory representations about the searched-for target's relational position within a repeatedly encountered distractor context are accumulated across trials, and then expedite behavioral RTs and enhance lateralized ERP markers—both reflecting the more effective allocation of attention in repeated search displays. Importantly, statistical LT memories can be established in a crossmodal fashion, enabling re-occurring distractor configurations in one sensory modality (e.g., touch) to facilitate search in another (e.g., visual) modality. Theoretically, there are at least two principal accounts for this. One possibility is that crossmodal adaptation processes are set by the sensory modality that is dominant in a given performance function. Accordingly, given that spatial judgments are the province of the visual modality, items from non-visual modalities will be remapped into the coordinate system of this modality in spatial learning tasks (hypothesis 1). An alternative possibility is that functional reorganization of modalities is contingent on the modality that is relevant to the task at hand, i.e., the modality in which the target is defined (hypothesis 2). Critically, these two possibilities would make the opposite predictions regarding measurable indices of crossmodal learning in a tactile search task with redundant—i.e., both tactile and visual—distractor items presented in consistent (and thus learnable) configurations throughout performance of the task (see below for details). Hypothesis 1 would predict the remapping of the tactile items into a visual format and, thus, crossmodal facilitation of unisensory learning. In contrast, hypothesis 2 would predict no or at best a minimal benefit deriving from the presence of additional visual-predictive distractors alongside the tactile predictive distractors in a tactile search task. To decide between these alternatives, the present study implemented a *tactile* search task in which the visual as well as the tactile context were predictive of the target location (on multisensory trials), in order to investigate what context would be learned and in which modality-specific coordinate system the context would be encoded and retrieved to facilitate performance.

In more detail, we conducted two experiments (differing only in the stimulus-onset asynchrony, SOA, between the visual and tactile contexts) to examine the impact of multisensory, visuotactile (relative to unisensory, tactile-only) contexts on
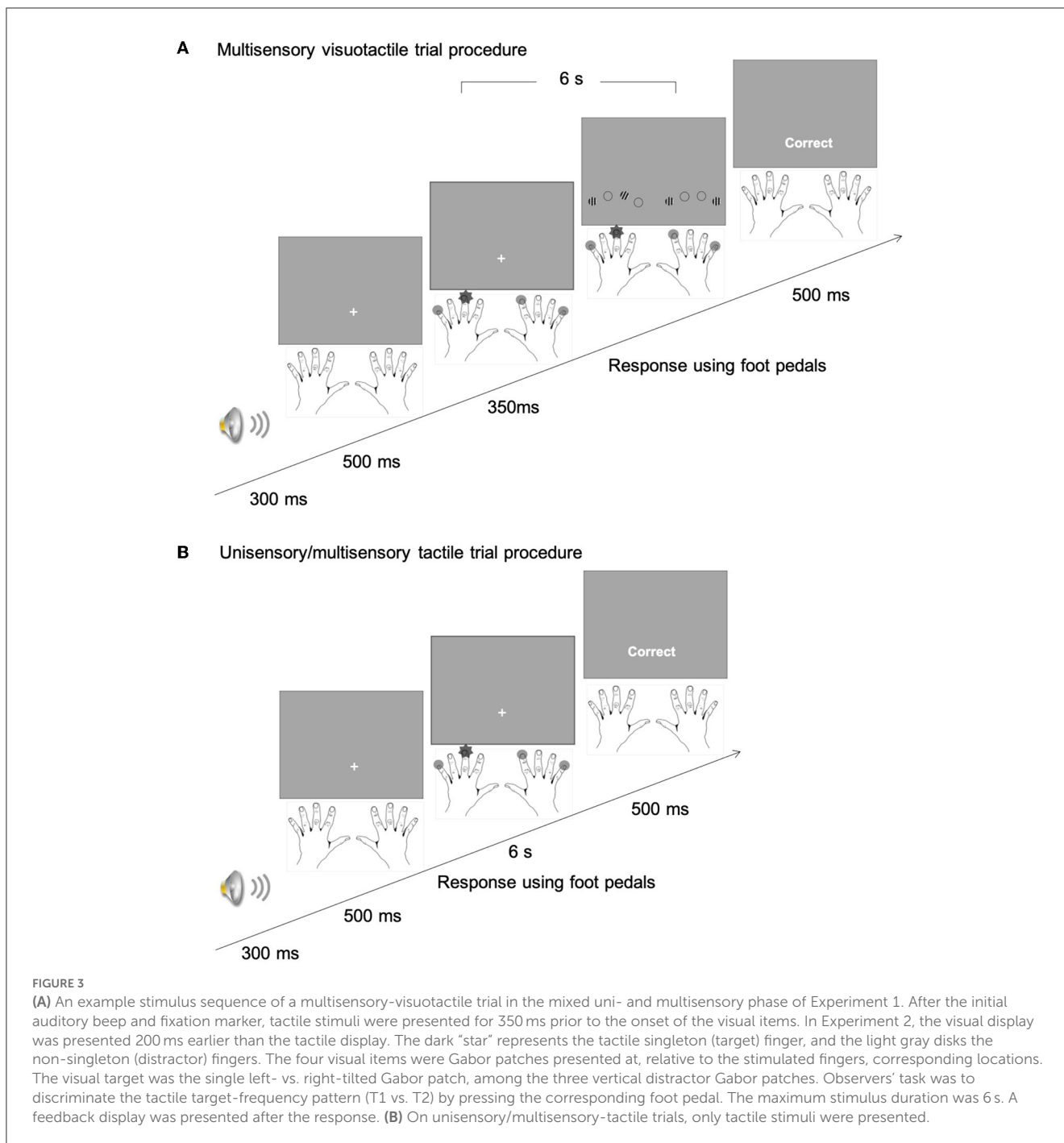
contextual facilitation learning in a tactile search task. Adopting a well-established, and demonstrably successful, multisensory learning protocol (Seitz et al., 2006; Kim et al., 2008; Shams et al., 2011; Chen et al., 2021a), observers had to search for and respond to a tactile odd-one-out target item appearing together, in a configuration, with three homogeneous tactile distractor items (see Figure 3). In 50% of the trials, the target-distractor configuration was fixed, i.e., the target appeared at a fixed location relative to the consistent distractor context (there were four such predictive, i.e., learnable contexts); in the other 50%, while the target position was also fixed, the locations of the distractors were randomly generated anew on each trial (there was the same number of such non-predictive contexts). Introducing this basic set-up, we tested contextual cueing in two separate, pure unisensory and mixed, uni- plus multisensory, phases. In the unisensory phase, the search was performed under the pure (unisensory) *tactile* task conditions just described; in contrast, in the mixed, uni- plus multisensory phase, trials with unisensory tactile stimulus arrays were presented randomly intermixed with trials with multisensory visuotactile contexts (the random mixing of trials ensured that participants adopted a consistent set to search for a tactile target). On the latter, visuotactile trials, the visual stimuli consisted of a configuration of three uniform distractor Gabor patches and one odd-one-out target Gabor patch, which were collocated with the positions of the tactile distractor and target stimuli. It is important to note that, in visuotactile studies of contextual cueing, the visual and tactile stimuli need to be collocated—which necessarily limits the number of (collocated) stimuli in the display. Nevertheless, previous work from our group has consistently shown reliable cueing effects using this multi-modal set-up (Chen et al., 2020, 2021b, 2022a,b), as well as with easy, "pop-out" visual search tasks (Geyer et al., 2010; Harris and Remington, 2017). Thus, by comparing contextual facilitation of RTs in tactile search with redundant, i.e., visual and tactile, distractor contexts vs. single, i.e., tactile-only, distractor contexts, we aimed to decide between the two alternative accounts (outlined above) of crossmodal contextual cueing in search tasks.

## 2. Materials and methods

### 2.1. Participants

Twenty-eight university students were recruited, and randomly assigned to Experiment 1 (14 participants; six males; $M = 27.4$ years, SD = 5.1 years) and Experiment 2 (14 participants; eight males; $M = 25.8$ years, SD = 3.95 years); they were all right-handed, had normal or corrected-to-normal vision, and reported normal tactile sensation. The sample sizes were determined by a-priori power analysis based on (effect size) $d_z = 0.81$ for a facilitatory effect of multisensory statistical learning in a similar study of multisensory context cueing (Chen et al., 2021a). According to the power estimates computed with G*Power (Erdfelder et al., 1996), a minimum sample size of 13 participants was required (with α = 0.05, and power = 0.85). All participants provided written informed consent before the experiment and were paid 9.00 Euro per hour for their services. The study was approved by the Ethics Committee of the LMU Munich Faculty of Psychology and Pedagogics.

**FIGURE 3**
**(A)** An example stimulus sequence of a multisensory-visuotactile trial in the mixed uni- and multisensory phase of Experiment 1. After the initial auditory beep and fixation marker, tactile stimuli were presented for 350 ms prior to the onset of the visual items. In Experiment 2, the visual display was presented 200 ms earlier than the tactile display. The dark "star" represents the tactile singleton (target) finger, and the light gray disks the non-singleton (distractor) fingers. The four visual items were Gabor patches presented at, relative to the stimulated fingers, corresponding locations. The visual target was the single left- vs. right-tilted Gabor patch, among the three vertical distractor Gabor patches. Observers' task was to discriminate the tactile target-frequency pattern (T1 vs. T2) by pressing the corresponding foot pedal. The maximum stimulus duration was 6 s. A feedback display was presented after the response. **(B)** On unisensory/multisensory-tactile trials, only tactile stimuli were presented.

## 2.2. Apparatus and stimuli

Both experiments were conducted in a sound-attenuated testing chamber, dimly lit by indirect incandescent lighting, with a Windows computer using Matlab routines and Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). The tactile and visual items were presented at spatially corresponding locations at vertically offset (i.e., a lower, tactile and an upper, visual) presentation planes (Figure 1A). Visual stimuli (and task instructions/ feedback) were projected onto a white canvas in front of the participant, using an Optoma projector (HD131Xe;

screen resolution: 1,024 × 720 pixels; refresh rate: 60 Hz), mounted on the ceiling of the experimental booth. The canvas was fixed on a wooden frame and tilted about 20° toward the observer. The viewing distance was fixed at about 60 cm with a chin rest. Responses were recorded using foot pedals (Heijo Research Electronics, UK).

Participants placed their eight fingers (except the thumbs) on eight solenoid actuators (each of a diameter of 1.8 cm, with a distance of 2 cm between adjacent actuators; see also Assumpção et al., 2015; Chen et al., 2020). The actuators activated lodged metal tips, vibrating a pin by 2–3 mm upon the magnetization of

the solenoid coils, controlled by a 10-Channel Amplifier (Dancer Design) connected to the computer with a MOTU analog-output card. Four vibro-tactile stimulations were presented to four fingers, two of each hand, with the tactile target delivered to one finger and tactile distractors to the three other fingers. Each distractor was a constant 150-Hz vibration, while the target was one of the following vibration patterns (see Figure 1C): target 1 (T1) was a 5-Hz square wave with 30% duty cycle, composed of 150-Hz vibrations, and target 2 (T2) a 5-Hz square wave with an average 60% duty cycle, also made up of 150-Hz vibrations. To make T2 distinguishable from T1, a blank gap of 200 ms was inserted between every two cycles in T2 (the mean frequency of T2 was thus 3.3 Hz).

Visual stimuli in multisensory, visuo-tactile, trials consisted of four (striped black and white; Michelson contrast of 0.96, spatial frequency of 2 cpd) and four empty circles, each subtending 1.8° of visual angle, presented on a gray background (36.4 cd/m$^2$). Of the four Gabor patches, one patch was an odd-one-out orientation item, deviating by +30° or −30° from the vertical: the visual "target"; and the other three were orientation-homogeneous, vertical visual "distractor" patches (see also Chen et al., 2021a). The visual Gabor and empty-circle items were presented at the eight "virtual" (i.e., collocated) finger positions on the upper display plane, with a distance of about 1.9° of visual angle between adjacent items. The "target" and "distractor" Gabor positions exactly matched the vibro-tactile target and distractor stimuli, i.e.,: the response-relevant tactile target position was signaled redundantly by a collocated visual Gabor singleton. Importantly, cross-modally redundant target-location signaling was realized with both predictive and non-predictive distractor contexts. This also applied to the pairing of a particular vibro-tactile target (T1 or T2) with a particular visual Gabor orientation (+30° or −30°); this pairing was fixed for a particular participant (and counterbalanced across participants). Keeping these conditions the same with both predictive and non-predictive distractor contexts was designed to rule out any potential influences of space- and identity/response-based crossmodal correspondences (e.g., Spence and Deroy, 2013) on the dependent measure: contextual facilitation.[1] During task performance, participants wore headphones (Philips SHL4000, 30-mm speaker drive) delivering white noise (65 dBA) to mask the (otherwise audible) sound produced by the tactile vibrations. The white noise started and stopped together with the vibrations.

## 2.3. Procedure

Experiments 1 and 2 only differed in the stimulus-onset asynchrony (SOA) between the visual and tactile displays. In Experiment 1, the *tactile* display was presented 350 ms prior to the visual display, similar to the setting in our previous work (Chen et al., 2020, 2021a). Conversely, in Experiment 2, the *visual* display

---

1  To avoid potential response incompatibility, the response to the tactile target T1 or T2 (either left or right) was mapped to the orientation of the tactile-matched visual singleton (left or right), across phases. That is, if T1 was paired to the left-tilted Gabor patch in the mixed, uni- plus multisensory phase for a given participant, the target T1 was assigned to the left key, and T2 to the right key for that participant.

was presented 200 ms before the tactile display. A pilot experiment (run in preparation for the current study) with the unisensory *visual* displays showed that the 200-ms presentation of the displays was sufficient to produce a contextual cueing effect, with four repeated target-distractor configurations. This is consistent with a previous study of ours (Xie et al., 2020), which demonstrated a contextual cueing effect with a 300-ms presentation, even though the search displays were more complex (consisting of 1 T-shaped target among 11 L-shaped distractors) and there were 12 repeated target-distractor configurations. Moreover, evidence from neuro-/electrophysiological studies indicate that the allocation of spatial attention diverges as early as 100–200 ms post-display onset between repeated and novel target-distractor configurations (e.g., Johnson et al., 2007; Chaumon et al., 2008; Schankin and Schubö, 2009).

### 2.3.1. Practice tasks

Participants first practiced the response mapping of the foot (i.e., response) pedals to the tactile targets (T1 or T2). The target-pedal assignment was fixed for each participant but counterbalanced across participants. The practice phase consisted of four tasks: (1) tactile target identification (32 trials); (2) tactile search (32 trials); (3) visual search (32 trials); and (4) multisensory search (64 trials, half of which presented only tactile targets and the other half redundantly defined, visuotactile targets). Participants had to reach a response accuracy of 85% in a given task before proceeding to the next task (all participants achieved this criterion with one round of training).

In the tactile target-identification task, one vibrotactile target (either T1 or T2 lasting 6 s) was randomly delivered to one of the eight fingers. Participants had to respond, as quickly and accurately, as possible by pressing the corresponding foot pedal to discriminate the tactile target. During this task, the tactile array was always accompanied by the correct target label, "T1" or "T2", on the screen, to aid identification of the tactile target (T1 vs. T2) and mapping it onto the required (left vs. right foot-pedal) response. In the tactile-search task, four vibrotactile stimuli, one target and three distractors, were delivered to two fingers of each hand. Participants had to identify T1 or T2 as quickly and accurately as possible by pressing the associated foot pedal. Given the experimental task proper consisted of redundant visuo-tactile displays, the visual-search practice was designed to familiarize participants with the visual target (and distractor) stimuli and, so, ensure that they would not simply be ignored on multisensory trials in the experiment proper (in which the task could be performed based on the tactile stimuli alone). In the visual search task, eight visual items (four Gabors and four empty circles) were presented on the screen. Participants were asked to identify the target Gabor orientation (tilted to the left or the right) as rapidly as possible by pressing the corresponding foot pedal.

In the practice of the search task under mixed, uni- and multisensory conditions, participants were presented with 50% unisensory tactile trials (identical to the tactile-search practice) and 50% multisensory visuotactile trials (presenting both one target and three distractors in each, the visual and the tactile, modality), randomly interleaved (see Figure 3). In Experiment 1, the visual

items were presented 350 ms after the tactile stimuli; in Experiment 2, they were presented 200 ms prior to the tactile stimuli. Importantly, although the multisensory displays had two collocated targets singled out in the two sensory modalities, participants were expressly instructed to set themselves for tactile search, even though the visual stimuli (provided only on multisensory trials) could provide cues to solving for the tactile task. This instruction was meant to ensure that the "tactile" task set was identical across uni- and multisensory trials[2]—allowing us to examine for any beneficial effects of multisensory vs. tactile stimulation on statistical context learning. To reflect *tactile* search, the RTs were recorded from the onset of the tactile stimuli in both experiments.

### 2.3.2. Experimental tasks

Immediately following the practice, each participant performed two experimental phases: a pure unisensory phase and a mixed, uni- and multisensory phase. The unisensory phase presented only tactile trials, and the mixed phase included both tactile and visuotactile trials, randomly intermixed. And the repeated target-distractor configurations were identical for the tactile-only and visuotactile trials in the mixed phase. The trial procedure was the same as in the respective practice tasks (see Figure 3). Each trial began with a 600-Hz beep (65 dBA) for 300 ms, followed by a short fixation interval of 500 ms. A search display (tactile or visuotactile) was then presented until a foot-pedal response was made or for a maximum of 6 s. Participants were instructed to respond as quickly and accurately as possible to the tactile target. Following observers' responses, accuracy feedback with the word "correct" or "wrong" was presented in the center of the screen for 500 ms (Figure 3). After an inter-trial interval of 1,000 ms, the next trial began. Eight consecutive trials constituted one trial block, consisting of the presentation of each of the four predictive display configurations plus four non-predictive configurations, in randomized order. After every two blocks, double beeps (2 × 200 ms, 1,000 Hz, 72 dBA, separated by an 800 ms silent interval) cued the accuracy feedback, with the mean accuracy attained in the previous two blocks shown in the center of the screen for 1,000 ms.

Half of the participants started with the pure unisensory phase and the other half with the mixed uni- and multisensory phase; each phase consisted of the same number of trials [Experiment 1: 256 trials per phase, with 128 repetitions per (repeated/non-repeated display) condition for the pure unisensory phase, and 64 repetitions per condition for the mixed, uni- plus multisensory phase; Experiment 2: 320 trials per phase with 160 repetitions per condition for the pure unisensory phase and 80 repetitions per condition for the mixed phase], to equivalent numbers of trials with tactile information between the pure unisensory and the mixed, uni- and multisensory phase. We increased the number of trials in Experiment 2 in order to extend the opportunity for contextual learning, i.e.,: would the enhanced contextual facilitation by multisensory information become observable with more trials— i.e., repetitions of each predictive display arrangement—per

learning "epoch"? Recall that each of the four predictive display arrangements is presented once per block, intermixed with four non-predictive displays in Experiment 1. So, in Experiment 2, an experimental epoch combined data across five blocks of trials (i.e., five repetitions of each predictive display), compared with four blocks (i.e., four repetitions of each predictive display) in Experiment 1.
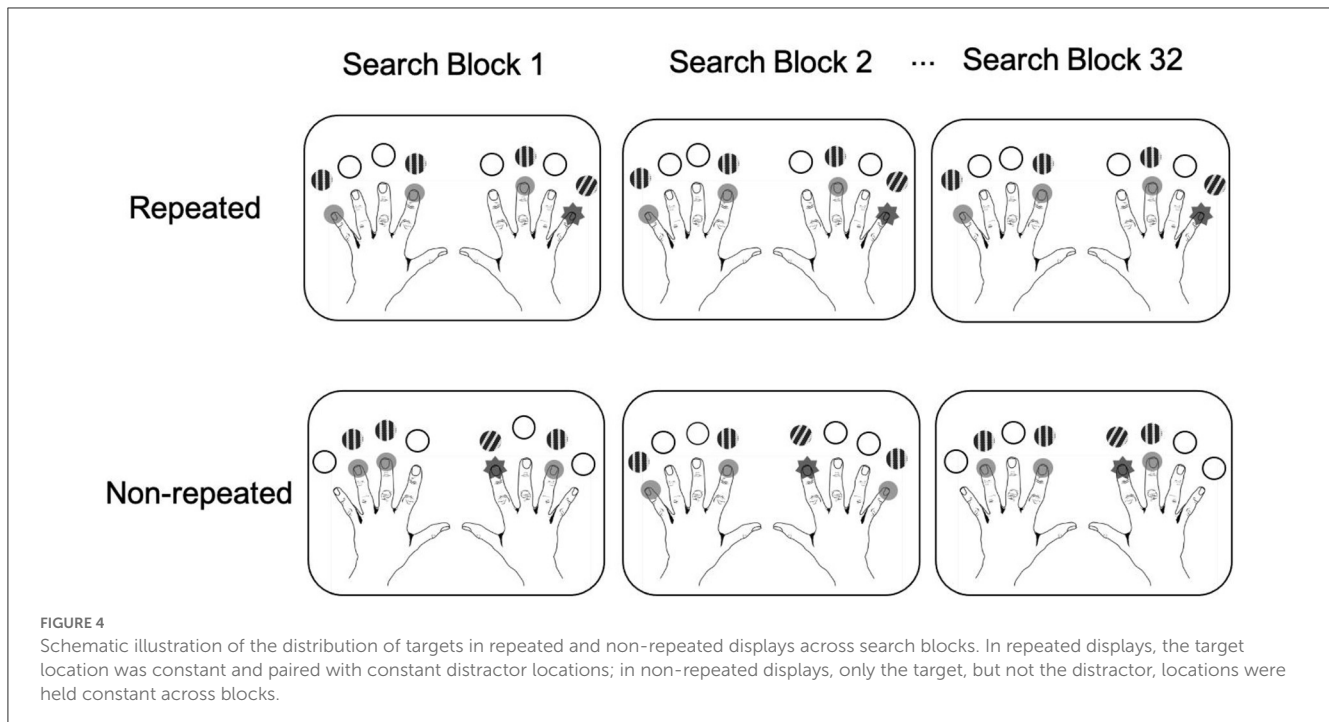
## 2.4. Design

To balance stimulus presentations between the left and right sides, the search arrays always consisted of two distractors on one side, and one target and one distractor on the other side. There were 144 possible displays in total to be sampled from. For the repeated contexts, we randomly generated two different sets of four displays for each participant, one set for the pure unisensory phase (hereafter Set 1) and one for the mixed uni- and multisensory phase (Set 2). Separate sets of repeated displays were generated to minimize potentially confounding transfer effects across phases. For "repeated" displays (of both sets), the target and distractor positions were fixed and repeated in each phase. For "non-repeated" displays, by contrast, the pairing of the target location with the three distractor positions was determined randomly in each block; these displays changed across blocks, making it impossible for participants to form spatial distractor-target associations. Note, though, that target locations were repeated equally in non-repeated and repeated displays (see Figure 4). That is, in each block of four repeated and four non-repeated trials, four positions, two from each side, were used for targets in the repeated condition, and the remaining four positions (again two on each side) for non-repeated displays (we also controlled the eccentricity of the target locations to be the same, on average, for repeated and non-repeated trials; see Supplementary material for an analysis of the eccentricity effects). This was designed to ensure that any performance gains in the "repeated" conditions could only be attributed to the effects of repeated spatial distractor-target arrangements, rather than repeated target locations, in this condition (see, e.g., Chun and Jiang, 1998, for a similar approach).

## 2.5. Data analysis

Trials with errors or RTs below 200 ms or above three standard deviations from the mean were excluded from RT analysis. Mean accuracies and RTs were submitted to repeated-measures analyses of variance (ANOVAs) with the factors Modality (unisensory-tactile, multisensory-tactile, multisensory-visuotactile), Display (repeated vs. non-repeated), and Epoch (1–8; one experimental epoch combining data across four consecutive trial blocks in Experiment 1 and 5 blocks in Experiment 2). Greenhouse-Geisser-corrected values were reported when the sphericity assumption was violated (Mauchley's test, $p < 0.05$). When interactions were significant, least-significant-difference *post-hoc* tests were conducted for further comparisons. The contextual-cueing effect was defined as the RT difference between repeated and non-repeated displays. We conducted one-tailed *t*-tests to examine the

---

2   Of course, this was also the only set permitting the task to be performed consistently, without set switching, on both types of—randomly interleaved—trials.

**FIGURE 4**
Schematic illustration of the distribution of targets in repeated and non-repeated displays across search blocks. In repeated displays, the target location was constant and paired with constant distractor locations; in non-repeated displays, only the target, but not the distractor, locations were held constant across blocks.

significance of the contextual-cueing effect (i.e., testing it against zero), given contextual cueing is, by definition, a directed effect: search RTs are expected to be faster for repeated vs. non-repeated search-display layouts (Chun and Jiang, 1998). We additionally report Bayes factors (Bayes inclusion for ANOVA) for non-significant results to further evaluate the null hypothesis (Harold Jeffreys, 1961; Kass and Raftery, 1995).

## 3. Results and discussion

### 3.1. Accuracy

The mean accuracies in Experiment 1 (in which the visual items were presented after the tactile items) were 90, 91, and 94%, for the unisensory-tactile, multisensory-tactile, and multisensory-visuotactile conditions, respectively. A repeated-measures ANOVA revealed no significant effects, all $p$s > 0.31, $\eta_p^2$s < 0.09, $BF_{incl}$s < 0.16.

In Experiment 2 (where the visual items were presented before the tactile items), the mean accuracies were 93, 91, and 97% for the unisensory-tactile, multisensory-tactile, and multisensory-visuotactile conditions, respectively. A repeated-measures ANOVA revealed the main effect of Modality to be significant, $F_{(1.3,16.94)}$ = 7.63, $p$ = 0.009, $\eta_p^2$ = 0.37: accuracy was higher for multisensory-visuotactile trials compared to both unisensory-tactile and multisensory-tactile trials (two-tailed, $p$s < 0.008, $d_z$s > 0.83); there was no significant difference between the latter two conditions ($p$ = 0.37, $d_z$ = 0.25, $BF_{10}$ = 0.39). Thus, accurately responding to the tactile target was generally enhanced by the preceding visual display (whether or not this was predictive). Further, accuracy was overall slightly higher for repeated (94.4%) vs. non-repeated (93.5%) displays, $F_{(1,13)}$ = 4.66, $p$ = 0.05, $\eta_p^2$ = 0.26, $BF_{incl}$ = 0.10, though the Bayes factor argues in favor of a null

effect. No other effects were significant, all $p$s > 0.1, $\eta_p^2$s < 0.16, and $BF_{incl}$s < 0.31.

### 3.2. RTs

Trials with extreme RTs were relatively rare: only 0.4% had to be discarded in Experiment 1 and 0.5% in Experiment 2. Figure 5A depicts the correct mean RTs for repeated and non-repeated displays as a function of Epoch, separately for the unisensory-tactile, multisensory-tactile, and multisensory-visuotactile trials, for Experiments 1 and 2, respectively. By visual inspection, both experiments show a procedural-learning effect: a general (i.e., condition-non-specific) improvement of performance with increasing practice of the task. Importantly, in contrast to Experiment 1, there was a clear contextual-cueing effect (over and above the general performance gain) in the multisensory-visuotactile as well as unisensory-tactile and multisensory-tactile search conditions (witness the differences between the corresponding solid and dashed lines) in Experiment 2; in Experiment 1, by contrast, there appeared to be no cueing effect in the multisensory-visuotactile condition. Recall, the only difference between Experiments 2 and 1 was the order in which the visual and tactile (context) stimuli were presented on multisensory-visuotactile trials: the visual context preceded the tactile context in Experiment 2, whereas it followed the tactile context in Experiment 1.

### 3.2.1. Experiment 1

A repeated-measures ANOVA of the mean RTs in Experiment 1 revealed significant main effects of Display (repeated vs. non-repeated), $F_{(1,13)}$ = 11.24, $p$ = 0.005, $\eta_p^2$ = 0.46, and Epoch, $F_{(7,91)}$
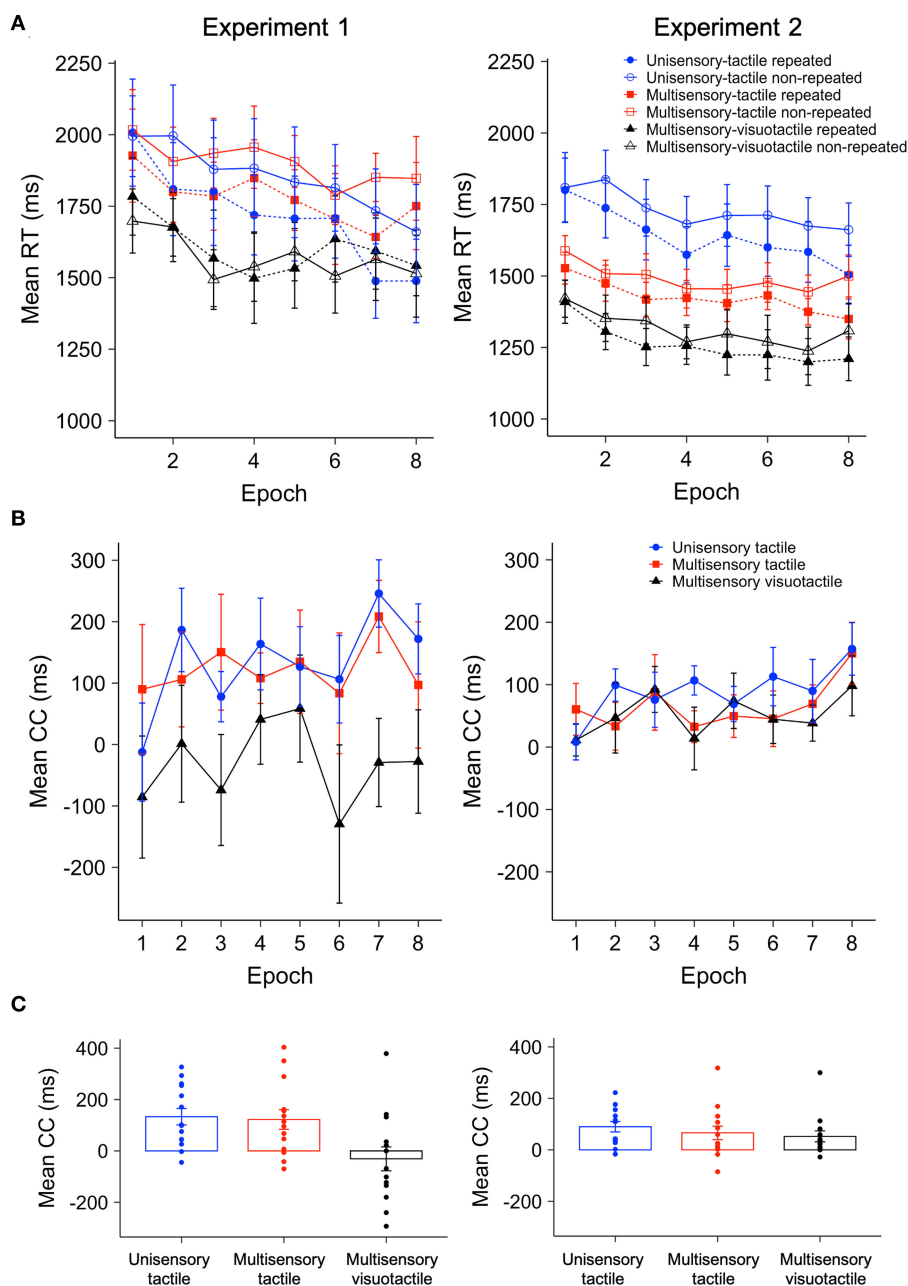
**FIGURE 5**
**(A)** Mean RTs for repeated and non-repeated contexts across epochs (one epoch collapses the RT data across four consecutive blocks in Experiment 1 and five blocks in Experiment 2) for unisensory-tactile, multisensory-tactile, and multisensory-visuotactile trials in Experiments 1 and 2. **(B)** Mean contextual-cueing (CC) effect as a function of epoch, for unisensory-tactile, multisensory-tactile, and multisensory-visuotactile trials in Experiments 1 and 2. **(C)** Mean contextual cueing (CC) and individual participants' data as a function of modality in Experiments 1 and 2. The error bars denote the within-subject standard error of the mean in panel A and the between-subject standard error of the mean in panels **(B, C)**.

$= 3.30$, $p = 0.004$, $\eta_p^2 = 0.2$. RTs were generally shorter to repeated vs. non-repeated displays, indicative of contextual cueing (mean contextual-cueing effect $= 75$ ms); and they decreased (linearly) across the task epochs, indicative of procedural task learning. The main effect of Modality was non-significant, $F_{(1.4,18.24)} = 0.97$, $p = 0.37$, $\eta_p^2 = 0.07$, $BF_{incl} = 0.63$. Importantly, the Modality × Display interaction turned out significant, $F_{(2, 26)} = 5.35$, $p = 0.01$, $\eta_p^2 = 0.29$, due to the RT difference between repeated and non-repeated

displays (i.e., contextual facilitation) being more pronounced for both unisensory-tactile and multisensory-tactile trials compared to multisensory-visuotactile trials (two-tailed, $ps < 0.047$, $d_z$s > 0.59), without a difference between the former two (purely tactile) conditions ($p = 0.82$, $d_z = 0.06$, $BF_{10} = 0.28$). No other interactions were significant (all $ps > 0.45$, $\eta_p^2$s $< 0.07$, $BF_{incl}$s $< 0.1$). Of note, contextual facilitation was reliably greater than zero in the unisensory-tactile ($133 \pm 27$ (SE) ms, one-tailed $t_{(13)} = 4.18$, $p <$

0.001, $d = 1.12$) and multisensory-tactile ($122 \pm 45$ ms, $t_{(13)} = 3.20$, $p = 0.007$, $d = 0.86$) conditions, but not the multisensory-visuotactile condition [$-31 \pm 45$ ms, $t_{(13)} = -0.66$, $p = 0.52$, $d = -0.18$, $BF_{10} = 0.18$; Figure 5C].

Despite the non-significant Modality main effect, responding appeared to be generally faster in the multisensory-visuotactile condition, and this may have curtailed any contextual-facilitation effect. Therefore, to more "fairly" compare contextual cueing among the three modality conditions, we normalized the cueing effect by dividing the mean RT facilitation (RT$_{non-repeted}$ – RT$_{repeated}$) by the mean RT for the respective modality condition, for each observer. Not surprisingly, the comparisons again revealed contextual facilitation to be much smaller (and, in fact, absent) for the multisensory-visuotactile condition ($-2.6 \pm 2.7\%$) compared to both the unisensory-tactile ($6.5 \pm 1.3\%$) and multisensory-tactile ($7 \pm 2\%$) conditions (see Supplementary Figure S2), $F_{(2, 26)} = 5.91$, $p = 0.008$, $\eta_p^2 = 0.31$.

Additional comparisons confined to the very first epoch of learning revealed no significant context-based facilitation for any of the three modality conditions, $ps > 0.20$, $d_z s < 0.24$, $BF_{10}s < 0.58$. In other words, contextual facilitation in the unisensory-tactile and multisensory-tactile conditions required more than four repetitions of each predictive tactile context (i.e., the number of repetitions in Epoch 1) to evolve.

Experiment 1 thus showed that predictive tactile contexts alone could facilitate tactile search in both the pure unisensory and mixed, uni- and multisensory phases of the experiment, whereas redundant predictive visuotactile contexts (with the visual display following the tactile array) failed to facilitate tactile search. Note that, in the mixed multisensory phase, the purely tactile and the visuotactile contexts involved exactly the same predictive tactile item configurations. Accordingly, the absence of contextual facilitation on visuotactile trials, which contrasts with the manifestation of facilitation on purely tactile trials (where the two types of trial were presented randomly interleaved), indicates that it is not the lack of contextual learning that is responsible for lack of cueing on the former trials; instead, this is likely due to retrieval of successfully learnt contexts being blocked when the visual context is presented after the tactile search array—consistent with previous findings (Zang et al., 2015).

### 3.2.2. Experiment 2

A repeated-measures ANOVA of the mean RTs in Experiment 2 again revealed significant main effects of Display, $F_{(1,13)} = 15.82$, $p = 0.002$, $\eta_p^2 = 0.55$, and Epoch, $F_{(1.93,25.13)} = 5.14$, $p = 0.014$, $\eta_p^2 = 0.28$. RTs were faster to repeated vs. non-repeated displays (mean contextual-cueing effect = 70 ms), and task performance improved generally with time-on-task. However, different from Experiment 1, the main effect of Modality was also significant, $F_{(1.15,14.91)} = 5.29$, $p = 0.03$, $\eta_p^2 = 0.29$: responding was substantially faster on multisensory-visuotactile trials ($1,286 \pm 23$ ms) compared to both multisensory-tactile ($1,459 \pm 19$ ms) and unisensory-tactile trials ($1,683 \pm 31$ ms; $ps < 0.024$, $|d_z|s > 0.68$), without any significant differences between the latter two conditions ($p = 0.12$, $d_z = 0.45$, $BF_{10} = 0.83$). There were no significant interactions, all $p$'s > 0.11, $\eta_p^2 s < 0.12$, $BF_{incl}s < 0.31$. Further one-sample $t$-tests

revealed the contextual-facilitation effect to be reliable (i.e., greater than zero; see Figure 5C) and statistically comparable [$F_{(2, 26)} = 1.15$, $p = 0.33$, $\eta_p^2 = 0.08$, $BF_{incl} = 0.38$] in all three modality conditions [unisensory-tactile trials: $90 \pm 23$ ms, $t_{(13)} = 4.50$, $p < 0.001$, $d = 1.20$; multisensory-tactile trials: $66 \pm 19$ ms, $t_{(13)} = 2.53$, $p = 0.013$, $d = 0.68$; multisensory-visuotactile trials: $53 \pm 9$ ms, $t_{(13)} = 2.46$, $p = 0.014$, $d = 0.66$]. Again, given the differences in the general, baseline-RT levels among the three conditions and the facilitation effects scaling with the baseline RTs, we further examined the normalized facilitation effects (see Supplementary Figure S2). These were $5.8 \pm 1.4\%$, $4.5 \pm 1.6\%$, and $4.5 \pm 1.4\%$ for the unisensory-tactile, multisensory-tactile, and multisensory-visuotactile conditions, respectively, and did not differ among the three conditions, $F_{(1.37,17.82)} = 0.57$, $p = 0.51$, $\eta_p^2 = 0.04$, $BF_{incl} = 0.26$. This pattern indicates when the visual context is presented prior to the tactile context, it neither enhances nor suppresses contextual cueing. Again, comparisons within Epoch 1 revealed no significant contextual facilitation in any of the three conditions, $ps > 0.14$, $d_z s < 0.38$, $BF_{10}s < 0.92$—so, more than five repetitions of each predictive tactile configuration were required to engender a cueing effect.

The significant Modality effect is interesting: It was due to the preceding visual display generally enhancing both response speed and accuracy (see the accuracy results above). However, this effect (in both RTs and accuracy) is independent of whether the visual context is predictive or non-predictive of the target location in the tactile array, i.e., it does not impact the contextual-cueing effect (the Modality × Display interaction was non-significant). Thus, the visual display likely just acts like an additional "warning signal" (Posner, 1978) over and above the auditory beep and fixation cross at the start of the trial, boosting observers' general preparedness for processing the impending tactile array.

### 3.2.3. Experiment 1 vs. Experiment 2

A further ANOVA comparing the *normalized* contextual-facilitation effects between Experiments 1 and 2, with the within-subject factors Modality and Epoch and the between-subject factor Experiment (see Supplementary Figure S2), revealed a significant main effect of Modality, $F_{(2, 52)} = 5.6$, $p = 0.006$, $\eta_p^2 = 0.18$, owing to a much reduced contextual-facilitation effect for multisensory-visuotactile trials vs. both multisensory-tactile (mean difference = 4.8%) and unisensory-tactile trials (mean difference = 5.2%; $ps < 0.01$, $|d_z|s > 0.71$); there was no difference between the latter two conditions (mean difference = 0.4%, $p = 0.83$, $d_z = 0.06$, $BF_{10} < 0.1$). This pattern was mainly attributable to Experiment 1 (rather than Experiment 2), as attested by the Modality × Experiment interaction, $F_{(2, 52)} = 4.38$, $p = 0.018$, $\eta_p^2 = 0.14$. And there was no significant difference in the normalized facilitation effects between experiments in the unisensory-tactile and multisensory-tactile conditions ($ps > 0.33$, $|d|s < 0.37$, $BF_{10}s < 0.51$), but a significantly reduced effect in the multisensory-visuotactile condition in Experiment 1 vs. Experiment 2 ($p = 0.026$, $d = -0.89$). No other effects were significant, all $p$'s > 0.38, $\eta_p^2 s < 0.04$, $BF_{incl}s < 0.1$. The results pattern remained the same when examining the original (non-normalized) contextual-cueing scores (see Figure 5B). Thus, the multisensory-visuotactile

condition engendered less (if any) contextual facilitation in Experiment 1 compared to Experiment 2. Given the analysis unit of an "Epoch" is somewhat arbitrary and, arguably, to examine for procedural learning effects, the cueing effect between the very first epoch of learning (in which participants had encountered the repeated arrangements only a few times) and the very last epoch (by which they had the maximum opportunity to acquire the contextual regularities) was compared by an ANOVA on the *normalized* contextual-facilitation effects with the within-subject factors Modality and Epoch (Epoch 1, Epoch 8) and the between-subject factor Experiment. The results revealed a significant main effect of Modality, $F_{(2, 52)} = 4.87$, $p = 0.012$, $\eta_p^2 = 0.16$, and a significant main effect of Epoch (Epoch 1, Epoch 8 ), $F_{(1,26)} = 4.46$, $p = 0.04$, $\eta_p^2 = 0.15$, with a larger cueing effect in Epoch 8 than in Epoch 1 (mean difference = 4.9%). No other main effects or interaction effects were significant, all $p$'s $> 0.15$, $\eta_p^2$s $< 0.08$, $BF_{incl}$s $< 0.22$. This result pattern is indicative of an increased effect of the contextual learning across the experiment for all three modality conditions, in both Experiments 1 and 2.

## 4. General discussion

The question at issue in the present study was to examine what context would be learned, and in which modality the context would be encoded and retrieved if both visual and tactile contexts are available in principle to guide tactile search. To address this, in two experiments, we compared the impact of multisensory, relative to unisensory, predictive contexts on the performance of a tactile search task. The two experiments differed in the order in which the visual and tactile contexts were presented on multisensory-visuotactile trials: the visual context followed the task-critical context in Experiment 1 but preceded it in Experiment 2. Critically, in the mixed uni- and multisensory phase of the task, we randomly intermixed tactile-only and visuotactile trials using identical predictive configurations in both trial types. Both experiments revealed reliable contextual cueing when the tactile context was shown alone, whether in a separate (unisensory) phase or randomly intermixed with visuotactile trials in the mixed (uni- and multisensory) phase, replicating previous findings (Assumpção et al., 2015, 2018). However, presenting both identically positioned visual items and the tactile target-distractor configuration together on multisensory trials did not enhance the contextual-cueing effect over and above the presentation of the tactile array alone, i.e.,: there was no redundancy gain from multisensory-visuotactile contexts. Indeed, the expression of the cueing effect was impeded when the visual display was presented after the tactile array in Experiment 1. We take the lack of a redundancy gain even under optimal conditions (with the visual display preceding the tactile array in Experiment 2) to indicate that, despite the availability of redundant, visual and tactile predictive item configurations, statistical learning of distractor-target contingencies is driven (solely) by the task-relevant, tactile modality.

Presenting the visual display after the tactile array on multisensory trials in Experiment 1 abolished the contextual-cueing effect. Given that the same predictive tactile configurations significantly facilitated tactile search on tactile-only trials in

the multisensory (i.e., the mixed, uni- plus multisensory) phase of the experiment, the lack of a contextual cueing effect on multisensory trials may be owing to the (delayed) presentation of the visual display interfering with tactile-context retrieval, likely by diverting attention away from the tactile modality (see also Manginelli et al., 2013; Zang et al., 2015). Whatever the precise explanation, the differential effects between Experiments 1 and 2 agree with the hypothesis that which modality is selected for the encoding of contextual regularities is determined by the task at hand.

Recall that in the existing studies of crossmodal contextual cueing (Chen et al., 2020, 2021a,b, 2022a,b), search was either visual (Chen et al., 2020, 2021a, 2022b) or only visual predictive contexts were presented to inform tactile search (Chen et al., 2021b, 2022a). Those studies consistently showed that learning predictive distractor contexts in one modality can facilitate search in the other, target modality while highlighting the aptness of the spatiotopic visual reference frame for crossmodal spatial learning. A question left open by these studies was how statistical context learning develops in the presence of redundant context stimuli encoded in different reference frames in search of a tactile target—in particular, predictive visual and predictive tactile contexts sensed in spatiotopic and somatotopic frames, respectively. We take the pattern of findings revealed in the present study to provide an answer: The spatiotopic reference frame of the visual modality is not the default system for multisensory contextual learning. Rather, when the task requires a search for a tactile target, contextual memories are formed within the somatotopic frame of the tactile modality—even when the target location is redundantly predicted by both the tactile and the visual item configuration.

In the previous study, Chen et al. (2021a) had observed enhanced contextual cueing when the task-critical visual item configuration was preceded by predictive tactile contexts (vs. predictive visual contexts alone) in a *visual* search task. Extrapolating from this result, a multisensory contextual redundancy gain would also have been expected in the present study, at least when the visual display preceded the tactile array. Chen et al. (2021a) argued that presenting the tactile context prior to the visual context in the visual search task permitted the predictive tactile array to be remapped into spatiotopic-external coordinates, i.e., the reference system of the visual modality. Accordingly, the remapped tactile-predictive array could be combined with the visual-predictive display, enhancing visual contextual cueing over and above the level rendered by the unisensory visual context alone (Kennett et al., 2002; Heed et al., 2015; Chen et al., 2021b). By analogy, in the present study, encoding of the preceding visual configuration could conceivably have engendered visual-to-tactile remapping, thus adding to the cue provided by the task-relevant tactile arrangement to enhance contextual facilitation (based on the common somatotopic reference system). However, our results are at odds with this possibility: although the prior onset of the visual array boosted performance (accuracy and speed) in general, it did not enhance contextual cueing. A likely reason for this is an asymmetry in coordinate-frame remapping: while somatotopic ("tactile") coordinates can be efficiently remapped into spatiotopic-external ("visual") coordinates, there may be no ready routines for

remapping spatiotopic-external coordinates into somatotopic coordinates (Pouget et al., 2002; Eimer, 2004; Ernst and Bülthoff, 2004). Given this, the present findings demonstrate a limit of multisensory signal processing in contextual cueing: multisensory redundancy gains require that both the visual-predictive and the tactile-predictive contexts can be coherently represented in a reference frame that is supported by the task-critical, target modality. Our results show that predictive visual contexts fail to meet this (necessary) condition when the task requires a search for a tactile target.

We acknowledge a possible limitation of the current study, namely: the fact that participants underwent only a relatively short multisensory phase of task performance. Recall that, even in this phase, the critical, multisensory-visuotactile displays occurred only on half the trials (the other half being designed to enforce a tactile task set, as well as providing a unisensory-tactile baseline condition against which to assess any multisensory-visuotactile redundancy gains). Thus, it cannot be ruled out that multisensory contextual facilitation of tactile search might be demonstrable with more extended training regimens (for indirect evidence of the facilitatory effect of consistent audio-visual training on the subsequent performance of a pure visual search task, see Zilber et al., 2014). Accordingly, with respect to the present visuotactile scenario, future work might examine whether tactile cueing of target-distractor regularities would be enhanced by concurrent visual-predictive items when an extended training schedule, perhaps coupled with a pre-/post-test design (cf. Zilber et al., 2014), is implemented.

In sum, when both visual-predictive and tactile-predictive contexts are provided in a tactile search task, the tactile context dominates contextual learning. Even giving the visual contexts a head-start does not facilitate tactile learning, likely because there are no ready routines for remapping the visual item configuration into the somatotopic coordinates underlying the tactile task. We conclude that the task-critical—i.e., target—modality determines the reference frame for contextual learning; and whether or not redundant predictive contexts provided by another modality can be successfully exploited (to enhance contextual cueing) depends on the availability of the requisite spatial remapping routines.

## 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: https://osf.io/73ejx/.

## Ethics statement

The studies involving human participants were reviewed and approved by LMU Munich Faculty of Psychology and Pedagogics. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

SC: conception, experimental design, data collection, data analysis, results interpretation, and drafting. ZS, TG, and HM: conception, experimental design, results interpretation, and revision. GV: data collection, data curation, and methodology. All authors contributed to the article and approved the submitted version.

## Conflict of interest

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

## Publisher's note

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/fcogn.2023.1124286/full#supplementary-material

## References

Assumpção, L., Shi, Z., Zang, X., Müller, H. J., and Geyer, T. (2015). Contextual cueing: implicit memory of tactile context facilitates tactile search. *Attent. Percept. Psychophys.* 77, 1212–1222. doi: 10.3758/s13414-015-0848-y

Assumpção, L., Shi, Z., Zang, X., Müller, H. J., and Geyer, T. (2018). Contextual cueing of tactile search is coded in an anatomical reference frame. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 566–577. doi: 10.1037/xhp0000478

Azañón, E., and Soto-Faraco, S. (2008). Changing reference frames during the encoding of tactile events. *Curr. Biol.* 18, 1044–1049. doi: 10.1016/j.cub.2008.06.045

Bavelier, D., and Neville, H. J. (2002). Cross-modal plasticity: where and how? *Nat. Rev. Neurosci.* 3, 443–452. doi: 10.1038/nrn848

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Chaumon, M., Drouet, V., and Tallon-Baudry, C. (2008). Unconscious associative memory affects visual processing before 100 ms. *J. Vis.* 8, 1–10. doi: 10.1167/8.3.10

Chen, S., Geyer, T., Zinchenko, A., Müller, H. J., and Shi, Z. (2022a). Multisensory rather than unisensory representations contribute to statistical context learning in tactile search. *J. Cogn. Neurosci.* 34, 1702–1717. doi: 10.1162/jocn_a_01880

Chen, S., Shi, Z., Müller, H. J., and Geyer, T. (2021a). Multisensory visuo-tactile context learning enhances the guidance of unisensory visual search. *Sci. Rep.* 11, 9439. doi: 10.1038/s41598-021-88946-6

Chen, S., Shi, Z., Müller, H. J., and Geyer, T. (2021b). When visual distractors predict tactile search: the temporal profile of cross-modal spatial learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 47, 1453–1470. doi: 10.1037/xlm0000993

Chen, S., Shi, Z., Zang, X., Zhu, X., Assumpção, L., Müller, H. J., et al. (2020). Crossmodal learning of target-context associations: when would tactile context predict visual search? *Attent. Percept. Psychophys.* 82, 1682–1694. doi: 10.3758/s13414-019-01907-0

Chen, S., Shi, Z., Zinchenko, A., Müller, H. J., and Geyer, T. (2022b). Cross-modal contextual memory guides selective attention in visual-search tasks. *Psychophysiology* 59, e14025. doi: 10.1111/psyp.14025

Chua, K.-P., and Chun, M. M. (2003). Implicit scene learning is viewpoint dependent. *Percept. Psychophys.* 65, 72–80. doi: 10.3758/BF03194784

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends Cogn. Sci.* 4, 170–178. doi: 10.1016/S1364-6613(00)01476-5

Chun, M. M., and Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cogn. Psychol.* 36, 28–71. doi: 10.1006/cogp.1998.0681

Chun, M. M., and Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychol. Sci.* 10, 360–365. doi: 10.1111/1467-9280.00168

Eimer, M. (2004). Multisensory integration: how visual experience shapes spatial perception. *Curr. Biol.* 14, R115–R117. doi: 10.1016/j.cub.2004.01.018

Erdfelder, E., Faul, F., and Buchner, A. (1996). GPOWER: a general power analysis program. *Behav. Res. Methods Inst. Comp.* 28, 1–11. doi: 10.3758/BF03203630

Ernst, M. O., and Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn. Sci.* 8, 162–169. doi: 10.1016/j.tics.2004.02.002

Geyer, T., Zehetleitner, M., and Müller, H. J. (2010). Contextual cueing of pop-out visual search: when context guides the deployment of attention. *J. Vis.* 10, 20. doi: 10.1167/10.5.20

Goujon, A., Didierjean, A., and Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. *Trends Cogn. Sci.* 19, 524–533. doi: 10.1016/j.tics.2015.07.009

Harold Jeffreys, S. (1961). *Theory of Probability*. Oxford: Clarendon Press.

Harris, A. M., and Remington, R. W. (2017). Contextual cueing improves attentional guidance, even when guidance is supposedly optimal. *J. Exp. Psychol. Hum. Percept. Perform.* 43, 926–940. doi: 10.1037/xhp0000394

Heed, T., Buchholz, V. N., Engel, A. K., and Röder, B. (2015). Tactile remapping: from coordinate transformation to integration in sensorimotor processing. *Trends Cogn. Sci.* 19, 251–258. doi: 10.1016/j.tics.2015.03.001

Ho, C., Santangelo, V., and Spence, C. (2009). Multisensory warning signals: when spatial correspondence matters. *Exp. Brain Res.* 195, 261–272. doi: 10.1007/s00221-009-1778-5

Johnson, J. S., Woodman, G. F., Braun, E., and Luck, S. J. (2007). Implicit memory influences the allocation of attention in visual cortex. *Psychon. Bull. Rev.* 14, 834–839. doi: 10.3758/BF03194108

Kass, R. E., and Raftery, A. E. (1995). Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795. doi: 10.1080/01621459.1995.10476572

Kawahara, J.-I. (2007). Auditory-visual contextual cuing effect. *Percept. Psychophys.* 69, 1399–1408. doi: 10.3758/BF03192955

Kennett, S., Spence, C., and Driver, J. (2002). Visuo-tactile links in covert exogenous spatial attention remap across changes in unseen hand posture. *Percept. Psychophys.* 64, 1083–1094. doi: 10.3758/BF03194758

Kim, R. S., Seitz, A. R., and Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS ONE* 3, e1532. doi: 10.1371/journal.pone.0001532

Kitazawa, S. (2002). Where conscious sensation takes place [Review of Where conscious sensation takes place]. *Conscious. Cogn.* 11, 475–477. doi: 10.1016/S1053-8100(02)00031-4

Lederman, S. J., and Klatzky, R. L. (2009). Haptic perception: a tutorial. *Attent. Percept. Psychophys.* 71, 1439–1459. doi: 10.3758/APP.71.7.1439

Luck, S. J., Woodman, G. F., and Vogel, E. K. (2000). Event-related potential studies of attention. *Trends Cogn. Sci.* 4, 432–440. doi: 10.1016/S1364-6613(00)01545-X

Manginelli, A. A., Langer, N., Klose, D., and Pollmann, S. (2013). Contextual cueing under working memory load: selective interference of visuospatial load with expression of learning. *Attent. Percept. Psychophys.* 75, 1103–1117. doi: 10.3758/s13414-013-0466-5

Nabeta, T., Ono, F., and Kawahara, J.-I. (2003). Transfer of spatial context from visual to haptic search. *Perception* 32, 1351–1358. doi: 10.1068/p5135

Nava, E., and Röder, B. (2011). Adaptation and maladaptation insights from brain plasticity. *Prog. Brain Res.* 191, 177–194. doi: 10.1016/B978-0-444-53752-2.00005-9

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366

Posner, M. I. (1978). *Chronometric Explorations of Mind*. Hillsdale, NJ: Lawrence Erlbaum.

Pouget, A., Ducom, J. C., Torri, J., and Bavelier, D. (2002). Multisensory spatial representations in eye-centered coordinates for reaching. *Cognition* 83, B1–B11. doi: 10.1016/S0010-0277(01)00163-9

Schankin, A., and Schubö, A. (2009). Cognitive processes facilitated by contextual cueing: evidence from event-related brain potentials. *Psychophysiology* 46, 668–679. doi: 10.1111/j.1469-8986.2009.00807.x

Seitz, A. R., Kim, R., and Shams, L. (2006). Sound facilitates visual learning. *Curr. Biol.* 16, 1422–1427. doi: 10.1016/j.cub.2006.05.048

Shams, L., Wozny, D. R., Kim, R., and Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Front. Psychol.* 2, 264. doi: 10.3389/fpsyg.2011.00264

Shi, Z., Zang, X., Jia, L., Geyer, T., and Muller, H. J. (2013). Transfer of contextual cueing in full-icon display remapping. *J. Vis.* 13, 2. doi: 10.1167/13.3.2

Sisk, C. A., Remington, R. W., and Jiang, Y. V. (2019). Mechanisms of contextual cueing: a tutorial review. *Attent. Percept. Psychophys.* 81, 2571–2589. doi: 10.3758/s13414-019-01832-2

Spence, C., and Deroy, O. (2013). How automatic are crossmodal correspondences? *Conscious. Cogn.* 22, 245–260. doi: 10.1016/j.concog.2012.12.006

Töllner, T., Conci, M., Rusch, T., and Müller, H. J. (2013). Selective manipulation of target identification demands in visual search: the role of stimulus contrast in CDA activations. *J. Vis.* 13, 23. doi: 10.1167/13.3.23

Wolfe, J. M., and Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nat. Hum. Behav.* 1, 0058. doi: 10.1038/s41562-017-0058

Xie, X., Chen, S., and Zang, X. (2020). Contextual cueing effect under rapid presentation. *Front. Psychol.* 11, 603520. doi: 10.3389/fpsyg.2020.603520

Zang, X., Jia, L., Müller, H. J., and Shi, Z. (2015). Invariant spatial context is learned but not retrieved in gaze-contingent tunnel-view search. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 807–819. doi: 10.1037/xlm0000060

Zilber, N., Ciuciu, P., Gramfort, A., Azizi, L., and van Wassenhove, V. (2014). Supramodal processing optimizes visual perceptual learning and plasticity. *NeuroImage* 93 (Pt 1), 32–46. doi: 10.1016/j.neuroimage.2014.02.017

Zinchenko, A., Conci, M., Töllner, T., Müller, H. J., and Geyer, T. (2020). Automatic guidance (and misguidance) of visuospatial attention by acquired scene memory: evidence from an N1pc polarity reversal. *Psychol. Sci.* 31, 1531–1543. doi: 10.1177/0956797620954815