

Cross-modal contextual memory guides selective attention in visual-search tasks

Siyi Chen¹  | Zhuanghua Shi^{1,2} | Artyom Zinchenko¹ | Hermann J. Müller^{1,2} | Thomas Geyer^{1,2}

¹General and Experimental Psychology, Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

²Munich Center for Neurosciences—Brain & Mind, Ludwig-Maximilians-Universität München, Munich, Germany

Correspondence

Siyi Chen, Allgemeine und Experimentelle Psychologie, Department Psychologie, LMU Munich, Leopoldstr. 13, D-80802 München, Germany.
Email: siyi.chen@psy.lmu.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: GE 1889/5-1 and SH166/7-1

Abstract

Visual search is speeded when a target item is positioned consistently within an invariant (repeatedly encountered) configuration of distractor items (“contextual cueing”). Contextual cueing is also observed in cross-modal search, when the location of the—visual—target is predicted by distractors from another—tactile—sensory modality. Previous studies examining lateralized waveforms of the event-related potential (ERP) with millisecond precision have shown that learned visual contexts improve a whole cascade of search-processing stages. Drawing on ERPs, the present study tested alternative accounts of contextual cueing in tasks in which distractor-target contextual associations are established *across*, as compared to, *within* sensory modalities. To this end, we devised a novel, cross-modal search task: search for a *visual* feature singleton, with repeated (and nonrepeated) distractor configurations presented either within the same (*visual*) or a different (*tactile*) modality. We found reaction times (RTs) to be faster for repeated versus nonrepeated configurations, with comparable facilitation effects between visual (unimodal) and tactile (crossmodal) context cues. Further, for repeated configurations, there were enhanced amplitudes (and reduced latencies) of ERPs indexing attentional allocation (PCN) and postselective analysis of the target (CDA), respectively; both components correlated positively with the RT facilitation. These effects were again comparable between uni- and crossmodal cueing conditions. In contrast, motor-related processes indexed by the response-locked LRP contributed little to the RT effects. These results indicate that both uni- and crossmodal context cues benefit the *same*, *visual* processing stages related to the selection and subsequent analysis of the search target.

KEYWORDS

CDA, contextual cueing, event-related potentials, multisensory processing, PCN, selective attention

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Psychophysiology* published by Wiley Periodicals LLC on behalf of Society for Psychophysiological Research.



1 | INTRODUCTION

Visual attention can be top-down controlled and guided by observers' "online" knowledge of search-critical object properties. For example, if a searched-for target is repeatedly encountered in an invariant arrangement of distractor elements, observers can learn these configurations and use them to expedite their search—an effect termed "contextual cueing" (Chun, 2000; Chun & Jiang, 1998). In the contextual-cueing paradigm (Chun & Jiang, 1998), participants search for a target item embedded in an arrangement of distractor items whose spatial locations are either repeated or newly generated throughout the course of the search experiment (repeated and nonrepeated conditions, respectively). The important finding is that reaction times (RTs) are faster to repeated (or "old") as compared to nonrepeated (or "new") displays, because learnt distractor-target spatial associations (stored in long-term memory) come to guide the search, cueing attention to—or predicting—the target location (Chun & Jiang, 1998). Furthermore, guidance of selective attention by learnt contexts is assumed to be implicit and automatic, as participants are typically unable to reliably discern repeated from nonrepeated in post-experimental (yes/no) recognition tasks (e.g., Goujon et al., 2015) and they persist to deploy attention to the learnt location even after consistent repositioning of the target to some other location (e.g., Zinchenko, Conci, Hauser, et al., 2020). Chun and Jiang's (1998) attention account of contextual cueing receives support from many subsequent studies using a variety of behavioral and electro-/physiological measures (e.g., Brockmole & Henderson, 2006; Chen et al., 2021a; Geyer et al., 2010; Giesbrecht et al., 2013; Johnson et al., 2007; Peterson & Kramer, 2001; Schlagbauer et al., 2017; Tseng & Li, 2004; Zinchenko, Conci, Töllner, et al., 2020). However, there are other findings suggesting that contextual cueing might also facilitate later, response-selection and/or motor-execution stages of processing, when participants make a decision about which motor (hand) effector is required for a correct manual response (i.e., response-selection account of contextual cueing; see, e.g., Chen et al., 2021a; Hout & Goldinger, 2012; Kunar et al., 2007; Schankin & Schubö, 2010).

Contextual learning of spatial distractor-target relations has been demonstrated to occur not only when the stimuli are visual but also within the tactile modality (Assumpção et al., 2015, 2018). More recently, Chen and colleagues (Chen et al., 2020, 2021b) have examined contextual cueing across sensory modalities, that is, under conditions when an invariant distractor context is defined in one modality and the search target in another modality. They found that repeatedly encountered (invariant) tactile distractor patterns facilitated search for a visual

target embedded in an array of nonrepeated (randomly arranged) visual distractors (Chen et al., 2020). A similar context-based facilitation of search has been shown when the target is defined in the tactile modality and the predictive distractors are visual (Chen et al., 2021b). Additional tests using hand-gesture manipulations (flipped/crossed hands) revealed that cross-modal contextual cueing is mediated by an environmental reference frame and that additional time is required for the tactile distractors to be remapped from the initial, somatotopically sensed format (Assumpção et al., 2018). These observations are consistent with context memory being supported by a supramodal, most likely visuospatial, representation that maintains modality-independent relational information. However, an alternative conception is also feasible, namely, that supramodal memory, albeit supported by a visual reference frame, is generated only when the target and distractors are defined in different, the visual and tactile, modalities (e.g., Shams & Seitz, 2008). That is, supramodal context memory would be only a specific instance of visual memory, which would not necessarily be identical with the memory that supports unimodal visual contextual cueing.

The main goal of the present study was to determine any similarities—as well as dissimilarities—of contextual cueing of visual search when distractor-target relational memories were acquired with unisensory versus multisensory training. Because reaction times in a visual search experiment can be affected by any processing stages related to the sensors (eyes) and response output (feet), we examined certain lateralized event-related potentials (ERPs) recorded over visual and motor areas that index the operation of attention- and response-related processes. This permitted us to track the processes that work more efficiently with learnt (vs. non-learnt) distractor-target arrays and compare these between uni- and multisensory training conditions. This question is of theoretical importance, because current theories of the memory-guided search and contextual cueing almost entirely draw on findings from visual search tasks (for reviews, see, e.g., Wolfe, 2019; Wolfe & Horowitz, 2017). Thus, tracking attentional and post-attentional processes over time in both unisensory and multisensory search environments would significantly impact on psychological theory building, i.e., advance our understanding about any commonalities versus differences in the processes that are more efficient in repeated search arrays when long-term contextual memories are established in an unisensory versus multisensory fashion.

One component of prime interest is the posterior contralateral negativity (PCN), which is taken to reflect enhanced focal-attentional selection of the search target within 200–300 ms of stimulus presentation (Luck & Hillyard, 1994; Töllner et al., 2008; Wascher & Wauschkuhn, 1996;

Woodman & Luck, 1999, 2003). In contextual-cueing paradigms, the PCN has been shown to index more efficient shifts of attention toward learnt target locations, with repeated contexts eliciting larger PCN amplitudes (Johnson et al., 2007; Zinchenko, Conci, Töllner, et al., 2020). Of note, however, these previous studies concentrated only on the amplitude of the PCN component, and not on its latency. The PCN latency (which is in the focus of the present study) might provide additional insight into the pre-attentive perceptual processing stages (Töllner et al., 2012) that may operate more efficiently with repeated arrays. Further, the individual contextual-cueing effects in the PCN amplitude have been found to correlate positively with both the behavioral contextual-cueing effects (Schankin & Schubö, 2009) and blood-oxygen-level-dependent (BOLD) signals in the medial temporal lobes (MTL; Kasper et al., 2015), the latter being considered as the neural generator site of the contextual-cueing effect (Geyer et al., 2012; Preston & Gabrieli, 2008). Thus, the PCN would be indicative of an effect of context memory on the (more efficient) attentional selection of the target item in repeated arrays.

An additional component of interest was the contralateral delay activity (CDA), a sustained negativity post stimulus presentation that is thought to reflect postselective (focal-attentional) processing of items held in working memory (WM; Mazza et al., 2007; Töllner et al., 2013; Vogel & Machizawa, 2004; Woodman & Vogel, 2008). Zinchenko, Conci, Töllner, et al. (2020) recently showed that repeated, relative to nonrepeated, configurations of items give rise to enhanced CDAs. They took this to reflect facilitated decision-making about the orientation of the target—that is, the response-critical property of the target item: typically, the character T rotated 90° to the left or the right—at a postselective processing stage. For instance, the matching of the target item against perceptual templates (held in WM) for deciding of left-versus right-orientation may be facilitated in repeated sensory arrays.

Another focus of the EEG data was the lateralized readiness potential (LRP; e.g., Coles, 1989; Eimer & Coles, 2003) recorded over the motor area contralateral to the side of a manual response, computed relative to response onset (i.e., response-locked LRP), which is known to be an online marker of motor-related processes in search tasks (Schankin & Schubö, 2009; Töllner et al., 2012). The LRP onset indexes the time required to execute the motor response (Töllner et al., 2011).

Thus, in summary, the present study was designed to examine the neural correlates of uni- and cross-modal contextual learning. Participants' electrophysiological brain activity was recorded during their performance of a visual search task (with eyes fixed at the center) while their eight fingers (except the thumbs) were at the same time placed at, relative to the visual items, corresponding

spatial locations on a lower plane and received vibro-tactile stimulation (see Figure 1). In the unimodal, visual–visual, condition, search arrays consisted of one visual target Gabor item, whose orientation had to be discriminated, presented amongst three non-target, distractor, Gabor items. For example, if the visual target was presented in the left display half, there was a second visual distractor item in this half plus two additional distractor items in the right display half (and vice versa for right-hemifield targets). Importantly, even in the unimodal condition, visual items were presented concurrently with a set of eight, though spatially homogeneous/indiscriminative, tactile distractors delivered at all four fingers (except the thumb) of each hand (see Figure 1). In the crossmodal, tactile–visual, condition, a given search array consisted of eight homogenous visual stimuli (one target and seven distractors, with four stimuli in each hemifield) arranged along the horizontal display axis; this visual array was presented together with four, spatially variably structured, tactile stimuli delivered at two fingers of each hand on the lower plane. Thus, both the uni- and cross-modal conditions used identical stimulations in terms of number of display items (4 visual +8 tactile items in the unimodal condition and 8 visual +4 tactile items in the cross-modal condition). Of note, the tactile stimulation always occurred 350 ms before the visual stimuli in order to compensate for processing differences across modalities and to allow for tactile-to-visual remapping to take place (see Chen et al., 2020, 2021a, 2021b; Colonius & Diederich, 2004). Further, 50% of the target items were presented on the left and 50% on the right half in order to avoid physical stimulus confounds in the measurement of lateralized ERPs (cf. Woodman, 2010). However, the uni- and crossmodal conditions differed with respect to the sensory modality that provided the repeated contexts: while the visual target was presented together with predictive visual distractors in the unimodal condition, it appeared amongst predictive tactile distractors in the crossmodal condition. Note that visual and tactile items were co-located with one-to-one correspondence, though they were presented on different planes (see Figure 1). This way, we were able to examine how search for a visual target element is aided by repeated distractor—context—information when distractor-target associations are formed within or across the sensory modalities of vision and touch.

2 | METHOD

2.1 | Participants

Twenty-three participants took part in this study (12 females; age: $M = 27.2$ years; all right-handed and all with

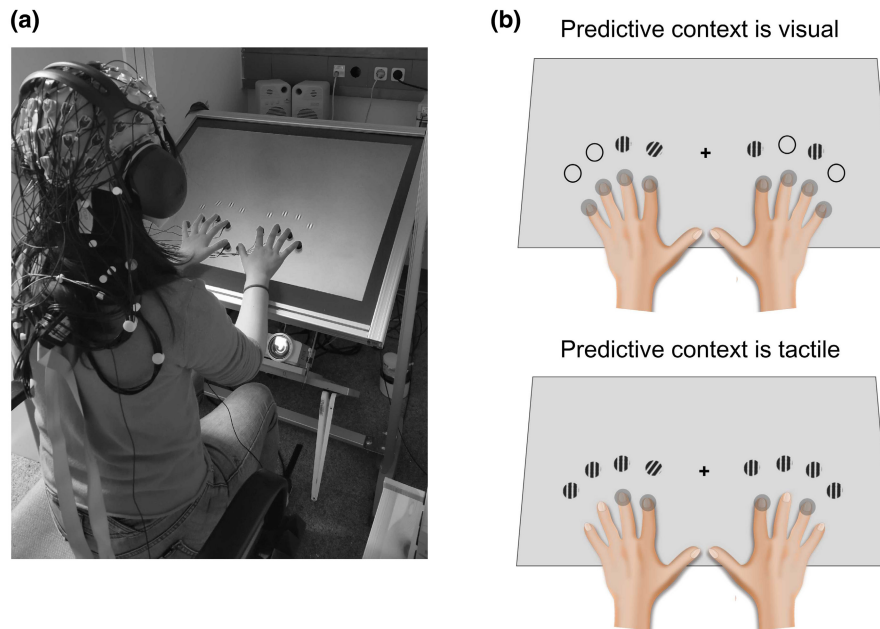


FIGURE 1 Illustration of the experimental setup and stimuli. (a) the real experimental setup from a photo of one participant. Participants placed their fingers on eight solenoid actuators (dancer design) delivering tactile stimulation (the two actuators under the thumbs were disabled). The vibration frequency was held constant at 150 Hz. Visual stimuli were presented on a semitransparent Plexiglas table tilted about 60° towards the observer. The participants wore headphones (Philips SHL4000, 30-mm speaker drive), through which white noise (65 dBA) was delivered to mask the tactile vibrations. (b) depicts the stimuli of the two sessions. Visual stimuli were Gabor patches with the target defined by an orientation difference relative to the distractor patches. These stimuli were presented at eight locations positioned along two virtual “curves” (one to the left and one to the right) over the horizontal axis, corresponding to the locations of the eight actuators below. Observers’ task was to respond to the (left-/right-tilt) orientation of the visual target via corresponding foot pedals. In the visual session, in which the predictive context was visual (upper panel), the search display consisted of one target with three distractor Gabor patches and four empty circles, accompanying tactile stimulation over all eight fingers. In the tactile session, in which the predictive context was tactile (lower panel), one visual target was embedded amongst seven homogenous distractors, with four vibrotactile stimulations delivered to two (selected) fingers of each hand. The gray circles represent stimulated fingers. The locations of the Gabor patches (for the visual session) and the vibrotactile stimulations (for the tactile session) varied depending on whether the configurations were repeated or not

normal or corrected-to-normal visual acuity). The sample size was determined based on previous, visual and tactile contextual-cueing studies (e.g., Assumpção et al., 2015; Chun & Jiang, 1998, 1999; Geyer et al., 2010) and previous studies that examined contextual cueing in combination with EEG (e.g., Schankin & Schubö, 2009, 2010; Zinchenko, Conci, Töllner, et al., 2020), aiming for 85% power to detect a relatively large effect size ($f[U] = 0.8$) in a repeated-measures analysis of variance (ANOVA; $\eta_p^2 = 0.4$) with an alpha level of .05. Power estimates were computed using G*Power (Erdfeider et al., 1996). The study was approved by the Ethics Committee of the Department of Psychology, Ludwig-Maximilians-Universität München. All participants provided written informed consent and received €9/h for taking part in the study.

2.2 | Apparatus and stimuli

The experiment, conducted in a sound-attenuated testing chamber dimly lit by indirect incandescent lighting,

was run on a Windows computer using Matlab routines and Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Visual stimuli and task instructions/feedback were projected onto a semitransparent Plexiglas table (size, 70 × 60 cm; height, 84 cm; Figure 1a) by a projector (Sharp XR-32X-L). The table (screen) was tilted about 60° toward the observer. The viewing distance was fixed at about 55 cm. The tactile and visual items were presented at eight spatially corresponding locations positioned along two virtual “curves” (one to the left and one to the right) on a lower (tactile) and upper (visual) horizontal axis of the respective presentation plane (Figure 1b). The visual stimuli consisted of Gabor patches (Michelson contrast 0.96, spatial frequency of 2 cpd), each subtending about 1.8° of visual angle, presented on a gray background (mean luminance of 36.4 cd/m²). The orientation of the distractor Gabor patches was homogeneously vertical (tilt degree: 0°). The singleton target was defined by orientation of ±9.2° (left- or right-tilted) from the vertical. Tactile stimuli were delivered via vibro-tactile stimulators (solenoid actuators with a diameter of 1.8 cm, Dancer

Design) to the fingers (thumbs excluded). The vibration frequency was held constant at 150 Hz. The actuators activated lodged metal tips vibrating a pin 2–3 mm following the magnetization of the solenoid coils, controlled by a 10-Channel Tactor Amplifier (Dancer Design) connected to a National Instrument computer with a MOTU analog output card. The spacing between adjacent items (on each side) was set at about 1.9° of visual angle, while separation between co-located visual stimuli and solenoid actuators was about 1.5° . During the experiment, participants were asked to wear headphones (Philips SHL4000, 30-mm speaker drive), through which white noise (65 dBA) was delivered to mask the tactile vibrations that would otherwise have been audible in the sound-insulated testing cabin.

2.3 | Design and procedure

Each participant performed two sessions in a counterbalanced order. In the “predictive-visual” session (Figure 1b,

upper panel), the search display consisted of one target with three distractor Gabor patches and four empty circles, accompanied by tactile stimulation over all four fingers on the left and right hand (except the thumbs). In the “predictive-tactile” session (Figure 1b, lower panel), one visual target was embedded in seven homogenous visual distractors, with four vibrotactile stimuli delivered to two (pre-defined) fingers of each hand. The locations of the Gabor patches (for the visual-predictive session) and the vibrotactile stimuli (for the tactile-predictive session) varied depending on whether the configurations were repeated or not (see Figure 2). For the repeated condition, the positions of both the target and distractors were fixed throughout the entire session. For new configurations, the positions of the search distractors were randomly generated on each trial anew; so, these positions had no predictive information regarding the target location, making it impossible for participants to form consistent spatial distractor-target associations. Note, though, that target positions were repeated equally often in repeated and nonrepeated configurations. That is, in each block of

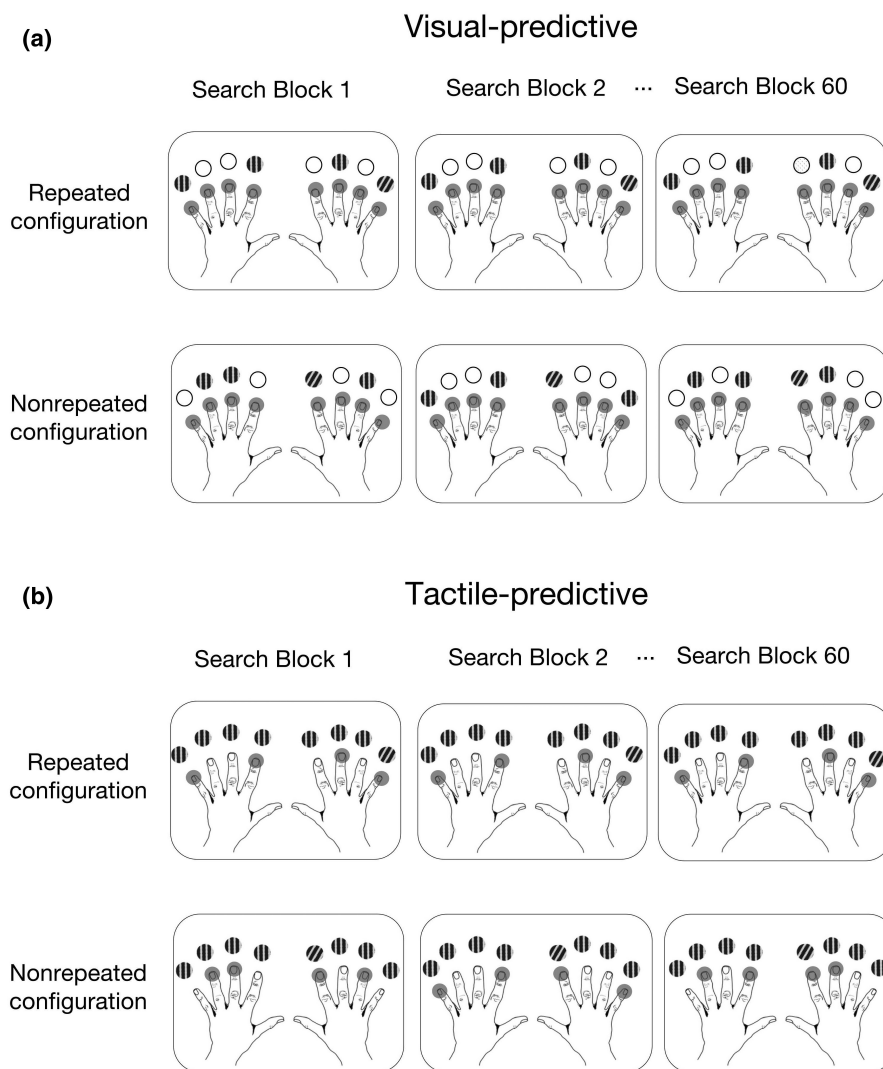


FIGURE 2 Schematic illustration of the distribution of targets in repeated and nonrepeated configurations across search blocks when the predictive distractors were visual (a, unimodal, visual–visual condition) and tactile (b, crossmodal, visual–tactile condition), respectively. In repeated configurations, the target location was constant and paired with constant (visual/tactile) distractor locations; in new configurations, by contrast, only the target, but not the (visual/tactile) distractor, locations were held constant across repetitions. The gray circles represent stimulated fingers

four repeated and four nonrepeated 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 nonrepeated configurations (Figure 2). This was intended to ensure that any performance gains in the repeated conditions could only be attributed to the effects of repeated spatial arrangements, rather than repeated target positions, in this condition (see, e.g., Chun & Jiang, 1998, for a similar approach), and also to balance stimulus presentations between the left and right hemifields (hands). The configurations used for the visual-predictive session were exactly the same as those used for the tactile-predictive session. Repeated and nonrepeated configurations were presented in each 50% of trials and randomly intermixed within each session for each participant.

A session consisted of 60 blocks of 8 trials each, with four repeated and four nonrepeated configurations. Each of the eight possible target locations was used and associated with the predictive or nonpredictive distractor configurations equally often in every block and throughout the experiment. Each trial began with a beep (600 Hz) for 300 ms to indicate the start of the trial. After a short interval of 300 ms with fixation, actuators began to vibrate 350 ms before the onset of the visual search array which was presented for 700 ms. The visual target was defined randomly as left- or right-tilted Gabor patch relative to the distractor orientation. Participants were asked to respond to the orientation of the target Gabor patch as quickly and accurately as possible. Responses were recorded using foot pedals (Heijo Research Electronics, UK). For example, when the tilt of the target was left (right), the participant had to press the left (right) foot pedal. Target-pedal assignment was counterbalanced across participants. A blank screen following the visual and tactile stimuli was presented until a response was executed or until a maximum duration of 800 ms had elapsed (participants could respond within 1500 ms from the onset of the visual + tactile search array). Next, a feedback screen (indicating “correct” or “wrong” response) was presented for 500 ms. After an inter-trial interval of 1000 to 1500 ms, the next trial began. Participants took a short break every 6 blocks.

Participants were not informed in any way of the aims of the experiment. Following written and verbal instructions, each observer was familiarized with the experimental setup. Before each session, they performed 24 practice trials. Participants then went on to perform the main experimental task of that session only if they achieved an accuracy level > 85%. Otherwise, participants were required to repeat the practice trials. After the experiment, participants were first asked to report anything they had noticed about the experimental task, whereupon they were administered an explicit (yes/no) recognition test consisting

of 32 trials in which they had to indicate whether they had already perceived a given display layout—consisting of the visual target, the tactile distractors, and the visual distractors—during the prior search experiment. In this recognition test, half of the trial displays included predictive visual/tactile configurations from the previous search task, and the other half newly generated configurations not presented before.

2.4 | EEG recording

The EEG was continuously sampled at 1 kHz using Ag/AgCl active electrodes (actiCAP system; Brain Products, Munich, Germany) from 64 scalp sites in accordance with the international 10–10 system. To monitor for blinks and eye movements, we additionally recorded the electrooculogram by means of electrodes placed at the outer canthi of the eyes and, respectively, the superior and inferior orbits. All electrophysiological signals were amplified using BrainAmp amplifiers (Brain Products) with a 0.1-Hz to 250-Hz band-pass filter. During data acquisition, all electrodes were referenced to FCz and re-referenced off-line to averaged mastoids. All electrode impedances were kept below 5 k Ω .

Prior to being segmented, the raw data were visually inspected in order to manually remove nonstereotypical noise; subsequently, the data were band-pass filtered using a 0.1-Hz to 40-Hz Butterworth infinite-impulse response filter (24 dB/octave). Next, an infomax independent-component analysis (ICA) was run to identify components representing blinks and horizontal eye movements and to remove these artifacts before back projection of the residual components (1% of all trials were removed because of eye-movement artifacts). Note that we also calculated the amount of horizontal eye-fixations based on the Gratton-and-Coles algorithm (Gratton et al., 1983) before the ICA procedure to actively test whether these eye movements were at play in the present investigation and may have impacted on performance even after ICA analysis and found that the proportion of horizontal eye movements was minimal and comparable across conditions: 0.71%, 0.71%, 0.72%, 0.74% for visual-repeated, visual-nonrepeated, tactile-repeated, tactile-nonrepeated, respectively. A repeated-measures analysis of variance (ANOVA) with the factors Predictive Context (visual, tactile) and Display (repeated, nonrepeated) revealed no significant effects, $F_s < 0.01$, $p_s > .9$, $\eta_p^2s < .001$, $BF_{10}s < 0.3$. ERPs were calculated time-locked to the onset of the visual stimuli, with segments extending from 200 ms before visual stimulus onset until 1000 ms afterwards and response-locked with segments extending from 600 ms prior to the response until 100 ms afterwards. In both stimulus- and response-locked

analyses, baseline correction was performed using the 200 ms interval preceding visual target onset. Only trials with correct responses and without artifacts (any signal exceeding $\pm 60 \mu\text{V}$), bursts of electromyographic activity (as defined by voltage steps or sampling points larger than $50 \mu\text{V}$), and activity lower than $0.5 \mu\text{V}$ within intervals of 500 ms (indicating dead channels) were accepted for further analysis on an individual-channel basis before averaging the ERP waves.

To extract the PCN and CDA components independently of the spatial location of the target in the left/right hemifield, we subtracted ERPs from parietooccipital electrodes (PO7 and PO8) ipsilateral to the target's location from contralateral ERPs. The latencies of the PCN component were defined individually as the maximum negatively directed deflection in the time ranges of 180 ms to 350 ms after visual stimulus presentation. We computed PCN amplitudes by averaging 5 sample points, respectively, before and after the maximum deflection. The CDA amplitudes were computed by averaging activity over the time range of 400 ms to 700 ms after visual stimulus presentation as the mean RT was around 700 ms for both unimodal and cross-modal conditions. This suggests that the 400–700 ms time period was highly accurate in revealing processes associated with the perceptual analysis of the selected item for extracting the response-critical feature (indexed by the CDA).

The LRP was calculated relative to the onset of the response (response-locked LRP). The LRP component was computed by subtracting ERPs measured at medial central electrodes (C3/C4) contralateral to the unipodal response side from ipsilateral ERPs, given that foot responses generate a more negative readiness potential over the motor cortex ipsilateral rather than contralateral to the responding foot (Böcker et al., 1994). Due to the somatotopic representation of the lower extremities within the longitudinal fissure, this paradoxical lateralization arises because the current dipole, located in the hemisphere contralateral to the responding foot, points to the ipsilateral hemisphere (cf. Brunia & Van den Bosch, 1984). Peak latency and peak amplitude were measured as the global maximal voltage at electrodes placed over the motor cortex (C3/C4). The onset latencies of the LRPs were determined by the jackknife-based scoring method (Miller et al., 1998), according to which the LRP onset is indicated when the LRP amplitude meets a specific criterion. As recommended by Miller et al. (1998), we used 90% of the maximum LRP activation as optimal criteria for defining the response-locked LRP onset latencies (see also Töllner et al., 2008, 2012). The amplitudes of the LRPs were calculated by averaging five sample points before and after the maximum deflection obtained for response-locked LRPs.

3 | RESULTS

3.1 | Behavioral data

For the RT analysis, trials with errors and with RTs below 200 ms and above 2.5 standard deviations from the mean were excluded from the analysis, leading to the removal of 3.3% of all trials (2.89% for the visual-predictive session; 3.7% for the tactile-predictive session). Mean error rates and RTs were submitted to a repeated-measures analysis of variance (ANOVA) with the factors Predictive Context (visual, tactile), Display (repeated, nonrepeated), and Epoch (1–5; one experimental epoch combining data across 12 consecutive trial blocks). Greenhouse–Geisser-corrected values are reported when Mauchly's test of sphericity was significant ($p < .05$). When interactions were significant, Bonferroni-corrected post-hoc tests were conducted for further comparisons. We additionally report Bayes factors (BF_{10}) for nonsignificant results to further evaluate, i.e., confirm, the null hypothesis (see Jeffreys, 1961; Kass & Raftery, 1995).

Figure 3 depicts the mean RTs for repeated and nonrepeated displays as a function of epoch, separately for the visual- and tactile-predictive contexts. The ANOVA revealed a significant main effect of Display, $F(1, 22) = 14.33$, $p < .001$, $\eta_p^2 = .39$, indicative of a RT contextual-cueing effect. The main effect of Epoch, $F(4, 88) = 14.45$, $p < .001$, $\eta_p^2 = .40$, was also significant, indicative of non-configural, that is, general procedural learning of (how to perform) the task at hand. Importantly, the main effect of (visual vs. tactile predictive) Context and all (two- or three-way) interactions involving this factor were nonsignificant (all $ps > .39$, $\eta_p^2s < .05$, $BF_{10}s < 0.11$). This suggests that both repeated visual and repeated tactile contexts were equally successful in facilitating visual search. Note that there was no context-based facilitation of RTs in Block 1 in either the uni- or the crossmodal condition, $F_s < 1$, $ps > .5$, $\eta_p^2s < .02$, $BF_{10}s < 0.36$, though the contextual-cueing effect was statistically significant in Epoch 1 (i.e., after >2 repetitions of each repeated display) for both visual and tactile contexts: visual, $F(1, 22) = 5.5$, $p = .028$, $\eta_p^2 = .2$; tactile, $F(1, 22) = 7.88$, $p = .01$, $\eta_p^2 = .26$.

The overall rate of response errors was 17% for visual contexts and 14.6% for tactile contexts; this is relatively high by the standards of search RT experiments but modest when the limited display exposure time and the prevention of eye movements are taken into account (see also Zinchenko, Conci, Hauser, et al., 2020; Zinchenko, Conci, Töllner, et al., 2020). A repeated measures ANOVA on the mean error rates revealed a significant main effect of Epoch, $F(4, 88) = 6.82$, $p < .001$, $\eta_p^2 = .24$, with error rates decreasing across epochs, and a main effect of Display, $F(1, 22) = 12.16$, $p = .002$, $\eta_p^2 = .36$, with lower error rates

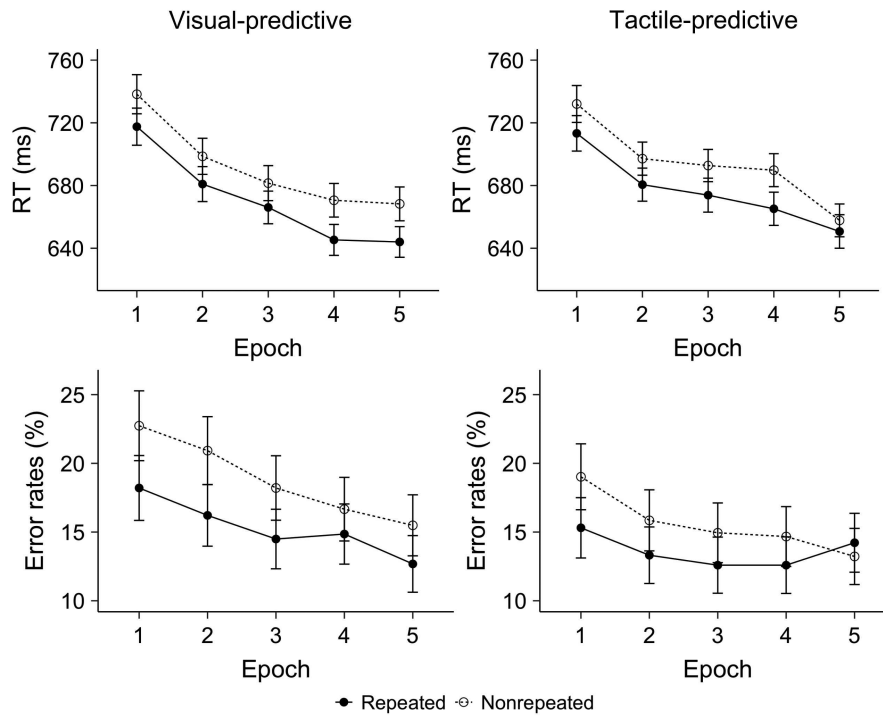


FIGURE 3 Mean reaction time (RT) and error rates for repeated and nonrepeated contexts as a function of epoch, separately for visual and tactile contexts. Error bars indicate 95% confidence intervals

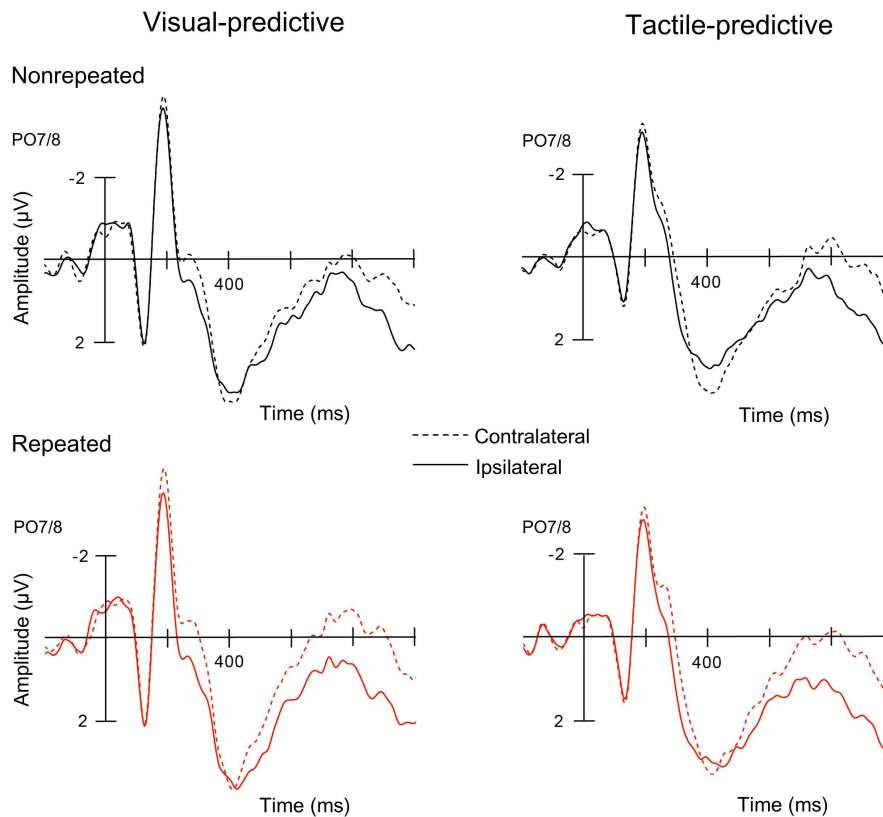


FIGURE 4 Grand-average ERPs at electrodes contra- and ipsilateral to the target (PO7 and PO8) are shown separately for nonrepeated (black) and repeated (red) displays, separately for the visual and tactile contexts

for repeated versus nonrepeated displays. Although the error rates were numerically higher for visual contexts, the effect of Context was not significant, $F(1, 22) = 2.69$, $p = .12$, $\eta_p^2 = .11$, $BF_{10} = 1.58$ (though the Bayes factor provided only anecdotal evidence). No interaction effects were significant (all p 's $> .12$, η_p^2 's $< .11$, BF_{10} 's < 0.3). The results of error analysis effectively rule out confounding of the RT effects by speed/accuracy trade-offs.

3.2 | Electrophysiological data

3.2.1 | Visual ERPs

Figure 4 presents the visual ERP waves contralateral and ipsilateral to the target for nonrepeated and repeated displays when the predictive contexts were visual and tactile, respectively. Of particular relevance are the latency and amplitude of the PCN, obtained at electrodes PO7/ PO8 from 180–350 ms post-display onset, both of which are indicative of more efficient allocation of attention to the search target in repeated versus nonrepeated displays (compare the contralateral minus ipsilateral difference waves in Figure 5). Interestingly, this attentional-guidance effect was effectively unaffected by whether the repeated context was visual or tactile. Our statistical analyses supported these observations. Entering the ERP waveforms in separate 2 (Display: repeated, nonrepeated) \times 2 (Predictive Context: visual, tactile) repeated-measures ANOVAs, we only found significant

main effects of Display for the PCN latency, $F(1, 22) = 12.38$, $p = .002$, $\eta_p^2 = .36$, and the PCN amplitude, $F(1, 22) = 23.00$, $p < .001$, $\eta_p^2 = .51$. No other (main or interaction effects) were significant (all p 's $> .60$, η_p^2 's $< .01$, BF_{10} 's < 0.29). An additional correlational analysis revealed a strong positive relationship between the contextual-cueing effect (nonrepeated minus repeated displays) in the PCN amplitude and the corresponding RT cueing effect. Of note, this effect was seen for both tactile ($r = 0.55$, $p = .005$) and visual predictive contexts ($r = .49$, $p = .009$; see Figure 6). The statistical significance of the correlation coefficient was determined by comparing the observed correlations with results derived from 20,000 permutations of the two variables excluding the influence from any outliers in the data (also for below).

Analysis of the CDA amplitude in the time window 400–700 ms post search-array onset also revealed a significant main effect of Display, $F(1, 22) = 7.74$, $p = .01$, $\eta_p^2 = .26$. Again, this effect was independent of whether the predictive context was visual or tactile (nonsignificant Display \times Context interaction, $p = .36$, $\eta_p^2 = .04$, $BF_{10} = 0.4$; see Figure 5). And the effect of Context was not significant, $p = .39$, $\eta_p^2 = .03$, $BF_{10} = 0.28$. Contextual facilitation of CDA amplitudes also correlated significantly with the RT facilitation, for both visual, $r = .41$, $p = .025$, and tactile predictive contexts, $r = .68$, $p < .001$, respectively (see Figure 6). In sum, the results of both the PCN and CDA analyses revealed a facilitatory effect for repeated displays which was uninfluenced by the type of—visual or tactile—predictive arrays.

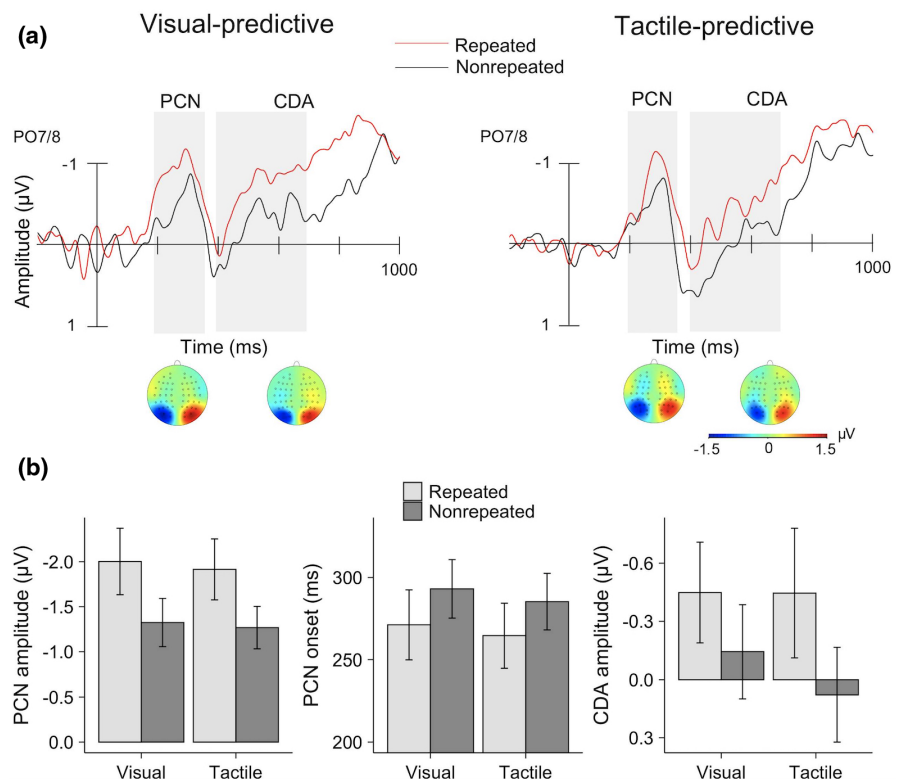


FIGURE 5 (a) Event-related potential difference waveforms (contralateral—Ipsilateral; electrodes PO7 and PO8) for repeated and nonrepeated displays, separately for the visual and tactile contexts. The shaded gray areas illustrate the timing of the posterior contralateral negativity (PCN) and contralateral delay activity (CDA). Each component is depicted with a corresponding scalp distribution. (b) Mean peak amplitudes and onset latencies in the PCN and mean amplitudes in the CDA for repeated and nonrepeated displays in the visual and tactile contexts. Error bars indicate 95% confidence intervals

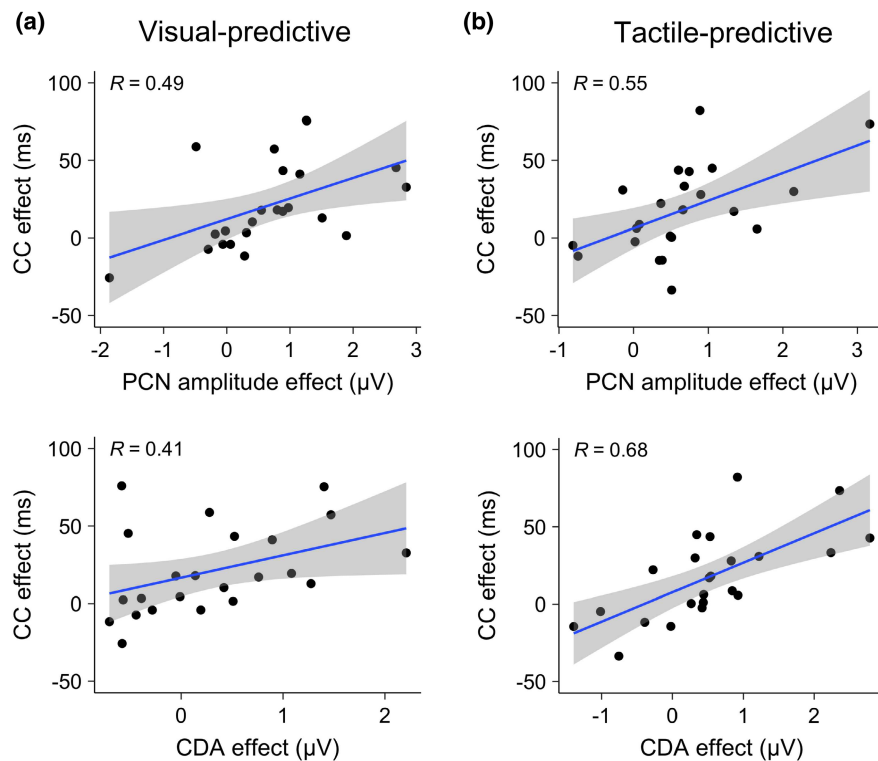


FIGURE 6 Correlations. The scatterplots in (a) and (b) show the relations between the individual PCN/CDA amplitude effects (difference between nonrepeated and repeated displays) and the individual RT contextual-cueing effects (CC effect) for the visual and tactile contexts, respectively. Solid lines indicate best-fitting regressions, shaded regions the 95% confidence intervals

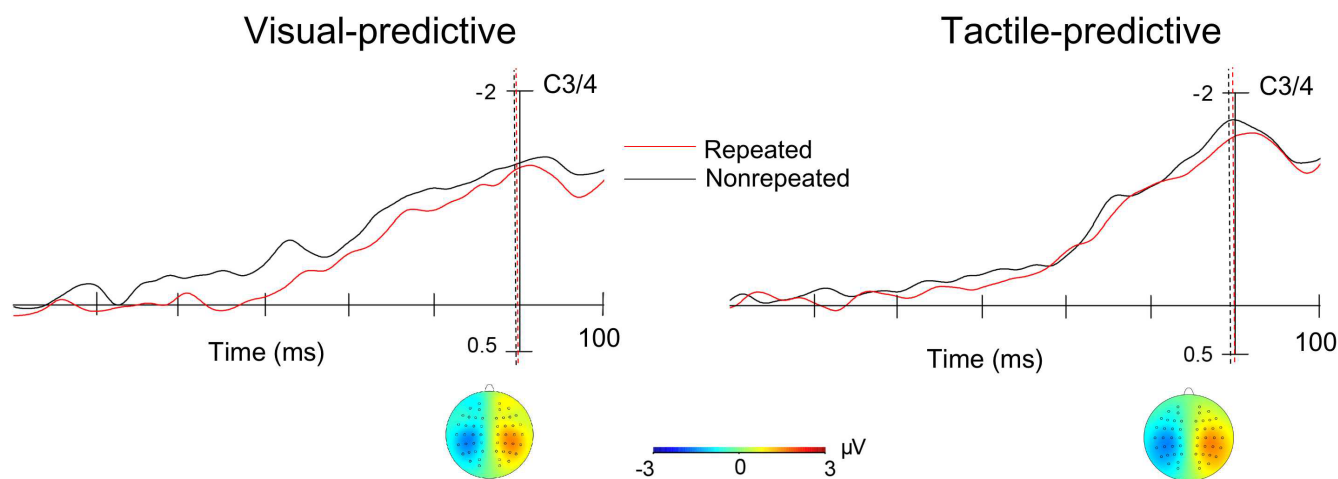


FIGURE 7 Grand average lateralized readiness potentials (LRP), synchronized to the onset of the response (response-locked LRP), measured at central electrodes (C3/C4), separately for visual and tactile contexts. The onsets of the LRPs are marked by vertical dashed lines

3.2.2 | Motor ERPs

Figure 7 shows the difference waves for the response-locked LRP components (electrodes C3/C4 over the motor area). As concerns the response-locked LRPs, there was a marginal Context \times Display interaction for the peak amplitudes, $F(1, 22) = 3.36$, $p = .08$, $\eta_p^2 = .13$, $BF_{10} = 0.58$: these were numerically, but not significantly, larger for nonrepeated versus repeated displays when the predictive context was visual (two-tailed $t[22] = -1.88$, $p = .073$, $d = -0.39$, $BF_{10} = 0.98$), but not when it was tactile ($t[22] = 0.15$, $p = .88$, $d = 0.03$, $BF_{10} = 0.22$). The main

effects of Context and Display were nonsignificant, both $ps > .30$, $\eta_p^2s < .05$, $BF_{10}s < 0.36$. Similarly, the ANOVA of the onset latencies yielded no significant effects: Context, $F(1, 22) = 3.14$, $p = .09$, $\eta_p^2 = .12$, $BF_{10} = 1.13$ (suggesting only inconclusive evidence); Display, $F(1, 22) = 0.17$, $p = .68$, $\eta_p^2 = .008$, $BF_{10} = 0.25$; and Context \times Display interaction, $F(1, 22) = 2.55$, $p = .13$, $\eta_p^2 = .1$, $BF_{10} = 0.88$. Importantly, though, neither the onset latency nor the peak amplitude effects (i.e., the difference in onset latency or peak amplitude between repeated and nonrepeated displays) of the response-locked LRP correlated significantly with the RT contextual-cueing effect, onset latency: visual

context $r = .28$, $p = .1$, $BF_{10} = 1$, tactile context $r = .35$, $p = .052$, $BF_{10} = 1.68$; peak amplitude: visual context $r = -.15$, $p = .76$, $BF_{10} = 0.16$, tactile context $r = -.27$, $p = .90$, $BF_{10} = 0.12$. Thus, any subtle effects attributable to response-related components contribute very little, if anything, to the behavioral contextual cueing effect.

3.3 | Recognition test

None of the participants spontaneously reported having noticed the display repetition during the search task. Participants' explicit recognition performance—that is, their ability to tell apart repeated tactile displays (“signals”) from nonrepeated displays (“noise”)—was assessed by the signal-detection sensitivity parameter d' (Green & Swets, 1966), treating correct recognition of repeated displays as “hit” responses and incorrect ‘recognition’ of nonrepeated displays as “false-alarm” responses. Across participants, d' was relatively small and statistically indistinguishable from zero for both visual and tactile contexts, visual: $d' = -0.14$, $t(22) = -0.93$, $p = .37$, $d = -0.19$, $BF_{10} = 0.32$; tactile: $d' = 0.07$, $t(22) = 0.59$, $p = .56$, $d = 0.12$, $BF_{10} = 0.26$. There was, thus, no evidence of explicit context memory in the current experiment, consistent with previous findings (e.g., Chun & Jiang, 1998).

4 | DISCUSSION

In the present study, we examined a series of lateralized visual and motor ERP components to elucidate the mechanisms involved in the acquisition of uni- and crossmodal spatial distractor-target associations. We found that implicit and automatic context-based guidance of search was expressed by enhanced PCN and CDA waves, indicating greater early allocation of attention to and postselective processing of the target in repeated displays. Importantly, these effects were seen—and statistically indistinguishable—in both predictive context conditions. These findings imply that spatial contextual associations can be formed successfully across sensory modalities, that is, when a visual target is predicted by tactile distractor configurations, and subsequently improve the efficiency of visual search. This was also supported by correlation analyses that showed a significant positive relationship between RT contextual cueing and each of the two visual ERP waveforms. Taken together, we found a striking relationship between the RT contextual-cueing and ERP effects, whether distractor-target contextual associations were acquired under uni- or cross-sensory learning conditions; but contextual facilitation of response execution (reflected

in the response-locked LRP) was not evident in neither uni- and crossmodal conditions.

Concerning the visual ERPs, we also found contextual cueing to lead to an earlier onset of the PCN component. Assuming that the timing of the PCN reflects the transition from pre-attentive sensory coding to focal-attentional selection (Töllner et al., 2011), the finding of reduced PCN latencies maybe taken to suggest that statistical learning of both unimodal and crossmodal distractor-target contexts increases the speed with which attention can be shifted towards the visual target item. This is consistent with the notion that contextual memory allows attention to be reliably and quickly shifted to the hemifield containing the target (e.g., Chun & Jiang, 1998, 1999; Chun & Nakayama, 2000). Interestingly, there have been no previous reports of contextual-cueing effects on PCN latencies (e.g., Johnson et al., 2007). One reason might be that, in comparison with previous studies, we used fewer repeated displays (4 instead of 12 as in Chun & Jiang, 1998, and Johnson et al., 2007). This effectively increases the number of repetitions of each individual display and thus the statistical power to reveal contextual learning, assuming that, normally, not all repeated displays are learnt in contextual-cueing experiments (see, e.g., Geyer et al., 2013; Smyth & Shanks, 2008). Overall, the results of PCN analysis support the attentional-guidance account of contextual cueing (Chun & Jiang, 1998, 1999; Chun & Nakayama, 2000). That is, repeating the search display leads to more efficient attentional deployment toward the visual target item, in terms of both expedited (as indicated by shorter PCN latencies) and enhanced engagement of attention at the target location (as indicated by the enhanced PCN amplitudes; see Zivony et al., 2018), independently of whether distractor-target contextual associations were formed within the visual or across the tactile and visual modalities. This interpretation lines in well with a recent study by Zinchenko, Conci, Töllner, et al. (2020) showing that the effects of display repetition in attention-related components start already with an early posterior negativity (N1pc, 80–180 ms) preceding the PCN (180–350 ms).

In addition to the PCN, uni- and crossmodal contexts also gave rise to a similar pattern in the CDA component: the CDA was enhanced over parietal-occipital areas when the target was presented in a repeated versus a nonrepeated array. This can be taken to reflect more efficient ‘focal-attentional’ processing of selected items (i.e., items represented in visual working memory, VWM) in repeated search displays (e.g., Zinchenko, Conci, Töllner, et al., 2020), perhaps as a result of the enhanced engagement of attention reflected in the PCN. Following selection of the target into VWM, postselective processes would include establishing that the selected item is actually a searched-for target (here: an off-vertical Gabor)

and, if so, extracting the relevant target feature (the left/right orientation of the target Gabor) to make a response decision (e.g., Mazza et al., 2007; Töllner et al., 2013; Wolfe, 2021; Woodman & Vogel, 2008). The enhanced CDA for repeated displays might thus reflect faster accumulation of evidence toward the decision threshold (e.g., parallel matching against both templates, rather than serial comparisons, or simply expedited matching) and/or a reduction of the amount of evidence required for a decision, as the target location (in repeated displays) is reliably predicted by the distractor context, which would increase the certainty that the selected item is the target (see Chen et al., 2021a; Sewell et al., 2018). More detailed (mathematical) modeling of decision making would be required to understand the dynamics at this postselective stage. Importantly, however, with regard to the question at issue in the present study: the CDA effects were comparable between the two learning conditions, that is: postselective target analysis is equally efficient whether search is guided by unimodal or crossmodal distractor-target contextual memories.

In contrast to the PCN and CDA effects, which both correlated significantly with the RT contextual-cueing effect, the motor-related ERPs appeared to contribute little, if anything: the response-locked LRP did not differ significantly between repeated and nonrepeated displays, which is consistent with Schankin and Schubö (2009, 2010); and they were not systematically correlated with the RT cueing effect—in either of the two contextual learning conditions. This suggests that, at least in the paradigms implemented and investigated in the present study (which differ in several respects from the “standard”, T vs. L’s search paradigm), there is not strong, or in Bayesian terms: inconclusive, evidence of the role of motor-related processes in accounting for the behavioral contextual-cueing effects.

Concerning possible limitations of our search tasks, we note that the visual displays differ between cross-modal and uni-modal conditions, which may have impacted on the ease with which participants searched the displays in the two conditions. Specifically, in order to induce context-based predictions, we had to make the visual context constant by presenting the visual items at the full set of 8 locations (cross-modal condition), while there were only 4 predictive visual items in the uni-modal condition. Thus, there were differences in the number of visual items (8 vs. 4) in the cross-modal and uni-modal conditions. Accordingly, visual search may have been more efficient with larger numbers of homogeneous distractors (e.g., Geyer et al., 2007; Wang & Theeuwes, 2020). Related to this is the possibility that the presence of more—distractor—filler items increased the amplitude of the PCN component in the cross-modal condition (Drisdelle

et al., 2020). Thus, more efficient search due to larger numbers of visual items may have confounded practice-related improvements of search from repeated contexts. However, what is at odds with this proposal is that we did not find reliable effects involving the factor Context in relevant behavioral and electrophysiological measures. There were no significant differences between the uni-modal and cross-modal conditions in either the RTs or error rates when comparing these measures in separate analyses for only repeated displays or only nonrepeated displays (all $ps > .11$, $d_zs < .35$, $BF_{10}s < 0.7$). Likewise, separate analyses of PCN amplitudes obtained from repeated and nonrepeated displays and a comparison of these measures between the uni-modal and cross-modal condition did not find a reliable effect (both $ps > .72$, $d_zs < .07$, $BF_{10}s < 0.23$). Moreover, the above cited studies (Drisdelle et al., 2020; Wang & Theeuwes, 2020) used larger numbers of (≥ 16) filler items for inducing search-related changes, while in the present study the difference in visual items between the uni-modal and cross-modal conditions was only small (at 4 items). Moreover, the uni-modal condition also used circular elements at non-Gabor stimulus locations (i.e., these elements did not contain Gabor gratings; cf. Figures 1 and 2), which would at least equate the number of visual circular stimuli between the uni-modal and cross-modal conditions. Worth mentioning is also that our study had a particular focus on relative ERP (and behavioral) effects, analyzing sensory and motor ERP waveforms in the critical repeated conditions always with reference to nonrepeated baseline displays. In doing so, unsystematic variability relating, e.g., to display design, would be effectively canceled out between the repeated and nonrepeated displays. Given these considerations, we believe that it is unlikely that the differences in the number of visual elements contributed to the pattern of highly comparable contextual-cueing effects found in the present uni-modal and cross-modal conditions. However, future work will be necessary to provide a stronger case for this particular claim.

5 | CONCLUSION

By tracking a series of ERP components (in addition to RTs) that reflect attentional and response-related processes in visual-search tasks, the present study provides new insights into the cognitive processes that are facilitated by unimodal and crossmodal contextual memory. We found that both uni- and crossmodal contexts afford more efficient guidance of attention, in terms of both the attentional engagement and speed with which selective attention is shifted toward the target location (or, at least, side) with repeated as compared to nonrepeated

distractor-target contexts. Also, focal-attentional processing of the selected target was facilitated by both uni- and crossmodal contexts, whereas response-related processes contributed little (if anything) to the behavioral contextual cueing effect. Overall, our new results provide *little* evidence at variance with the (parsimonious) notion of a single, supramodal mechanism underlying contextual cueing of visual search, whether the cues are visual themselves or, respectively, tactile.

ACKNOWLEDGMENT

This work was supported by German Research Foundation (DFG) grants GE 1889/5-1, awarded to TG, and SH166/7-1 to ZS. We thank Gizem Vural for her help with data collection. We thank Shaoyang Tsai for technical support. Open access funding enabled and organized by Projekt DEAL.

AUTHOR CONTRIBUTIONS

Siyi Chen: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; validation; visualization; writing – original draft. **Zhuanghua Shi:** Conceptualization; funding acquisition; methodology; project administration; resources; supervision; writing – review and editing. **Artyom Zinchenko:** Conceptualization; data curation; methodology; software; validation; writing – review and editing. **Hermann J. Müller:** Conceptualization; data curation; resources; supervision; validation; writing – review and editing. **Thomas Geyer:** Conceptualization; data curation; funding acquisition; methodology; project administration; resources; supervision; validation; writing – review and editing.

DATA AVAILABILITY STATEMENT

The data supporting the findings of the study are available in the Open Science Framework repository at <https://osf.io/mgyh9/>.

ORCID

Siyi Chen  <https://orcid.org/0000-0002-3090-6183>

REFERENCES

- Assumpção, L., Shi, Z., Zang, X., Müller, H. J., & Geyer, T. (2015). Contextual cueing: Implicit memory of tactile context facilitates tactile search. *Attention, Perception & Psychophysics*, 77(4), 1212–1222. <https://doi.org/10.3758/s13414-015-0848-y>
- Assumpção, L., Shi, Z., Zang, X., Müller, H. J., & Geyer, T. (2018). Contextual cueing of tactile search is coded in an anatomical reference frame. *Journal of Experimental Psychology: Human Perception and Performance*, 44(4), 566–577. <https://doi.org/10.1037/xhp0000478>
- Böcker, K. B., Brunia, C. H., & Cluitmans, P. J. (1994). A spatio-temporal dipole model of the readiness potential in humans. II. Foot movement. *Electroencephalography and Clinical Neurophysiology*, 91(4), 286–294. [https://doi.org/10.1016/0013-4694\(94\)90192-9](https://doi.org/10.1016/0013-4694(94)90192-9)
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436. <https://doi.org/10.1163/156856897x00357>
- Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, 13(1), 99–108. <https://doi.org/10.1080/13506280500165188>
- Brunia, C. H., & Van den Bosch, W. E. (1984). Movement-related slow potentials: I. A contrast between finger and foot movements in right-handed subjects. *Electroencephalography and Clinical Neurophysiology*, 57(6), 515–527. [https://doi.org/10.1016/0013-4694\(84\)90087-7](https://doi.org/10.1016/0013-4694(84)90087-7)
- Chen, S., Shi, Z., Müller, H. J., & Geyer, T. (2021a). Multisensory visuo-tactile context learning enhances the guidance of uni-sensory visual search. *Scientific Reports*, 11, 9439. <https://doi.org/10.1038/s41598-021-88946-6>
- Chen, S., Shi, Z., Müller, H. J., & Geyer, T. (2021b). When visual distractors predict tactile search: The temporal profile of cross-modal spatial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(9), 1453–1470. <https://doi.org/10.1037/xlm0000993>
- Chen, S., Shi, Z., Zang, X., Zhu, X., Assumpção, L., Müller, H. J., & Geyer, T. (2020). Crossmodal learning of target-context associations: When would tactile context predict visual search? *Attention, Perception & Psychophysics*, 82(4), 1682–1694. <https://doi.org/10.3758/s13414-019-01907-0>
- Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4(5), 170–178. [https://doi.org/10.1016/s1364-6613\(00\)01476-5](https://doi.org/10.1016/s1364-6613(00)01476-5)
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36(1), 28–71. <https://doi.org/10.1006/cogp.1998.0681>
- Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, 10(4), 360–365. <https://doi.org/10.1111/1467-9280.00168>
- Chun, M. M., & Nakayama, K. (2000). On the functional role of implicit visual memory for the adaptive deployment of attention across scenes. *Visual Cognition*, 7(1–3), 65–81. <https://doi.org/10.1080/135062800394685>
- Coles, M. G. (1989). Modern mind-brain reading: Psychophysiology, physiology, and cognition. *Psychophysiology*, 26(3), 251–269. <https://doi.org/10.1111/j.1469-8986.1989.tb01916.x>
- Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience*, 16(6), 1000–1009. <https://doi.org/10.1162/0898929041502733>
- Drisdelle, B. L., Corriveau, I., Fortier-Gauthier, U., & Jolicoeur, P. (2020). Effects of task-irrelevant or filler items on brain mechanisms of visual spatial attention. *Psychophysiology*, 57(11), e13644. <https://doi.org/10.1111/psyp.13644>
- Eimer, M., & Coles, M. G. H. (2003). The lateralized readiness potential. In M. Jahanshahi & M. Hallett (Eds.), *The Bereitschaftspotential: Movement-related cortical potentials* (pp. 229–248). Springer US. https://doi.org/10.1007/978-1-4615-0189-3_14
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28(1), 1–11. <https://doi.org/10.3758/BF03203630>

- Geyer, T., Baumgartner, F., Müller, H. J., & Pollmann, S. (2012). Medial temporal lobe-dependent repetition suppression and enhancement due to implicit vs. explicit processing of individual repeated search displays. *Frontiers in Human Neuroscience*, 6, 272. <https://doi.org/10.3389/fnhum.2012.00272>
- Geyer, T., Müller, H. J., Assumpcao, L., & Gais, S. (2013). Sleep-effects on implicit and explicit memory in repeated visual search. *PLoS One*, 8(8), e69953. <https://doi.org/10.1371/journal.pone.0069953>
- Geyer, T., Müller, H. J., & Krummenacher, J. (2007). Cross-trial priming of element positions in visual pop-out search is dependent on stimulus arrangement. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 788–797. <https://doi.org/10.1037/0096-1523.33.4.788>
- Geyer, T., Zehetleitner, M., & Müller, H. J. (2010). Contextual cueing of pop-out visual search: When context guides the deployment of attention. *Journal of Vision*, 10(5), 20. <https://doi.org/10.1167/10.5.20>
- Giesbrecht, B., Sy, J. L., & Guerin, S. A. (2013). Both memory and attention systems contribute to visual search for targets cued by implicitly learned context. *Vision Research*, 85, 80–89. <https://doi.org/10.1016/j.visres.2012.10.006>
- Goujon, A., Didierjean, A., & Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. *Trends in Cognitive Sciences*, 19(9), 524–533. <https://doi.org/10.1016/j.tics.2015.07.009>
- Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55(4), 468–484. [https://doi.org/10.1016/0013-4694\(83\)90135-9](https://doi.org/10.1016/0013-4694(83)90135-9)
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.
- Hout, M. C., & Goldinger, S. D. (2012). Incidental learning speeds visual search by lowering response thresholds, not by improving efficiency: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 38(1), 90–112. <https://doi.org/10.1037/a0023894>
- Jeffreys, H. (1961). *Theory of probability*. Clarendon Press.
- Johnson, J. S., Woodman, G. F., Braun, E., & Luck, S. J. (2007). Implicit memory influences the allocation of attention in visual cortex. *Psychonomic Bulletin & Review*, 14(5), 834–839. <https://doi.org/10.3758/bf03194108>
- Kasper, R. W., Grafton, S. T., Eckstein, M. P., & Giesbrecht, B. (2015). Multimodal neuroimaging evidence linking memory and attention systems during visual search cued by context. *Annals of the New York Academy of Sciences*, 1339, 176–189. <https://doi.org/10.1111/nyas.12640>
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Kunar, M. A., Flusberg, S., Horowitz, T. S., & Wolfe, J. M. (2007). Does contextual cuing guide the deployment of attention? *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 816–828. <https://doi.org/10.1037/0096-1523.33.4.816>
- Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 1000–1014. <https://doi.org/10.1037//0096-1523.20.5.1000>
- Mazza, V., Turatto, M., Umiltà, C., & Eimer, M. (2007). Attentional selection and identification of visual objects are reflected by distinct electrophysiological responses. *Experimental Brain Research*, 181(3), 531–536. <https://doi.org/10.1007/s00221-007-1002-4>
- Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, 35(1), 99–115. <https://doi.org/10.1111/1469-8986.3510099>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. <https://doi.org/10.1163/156856897X00366>
- Peterson, M. S., & Kramer, A. F. (2001). Attentional guidance of the eyes by contextual information and abrupt onsets. *Perception & Psychophysics*, 63(7), 1239–1249. <https://doi.org/10.3758/bf03194537>
- Preston, A. R., & Gabrieli, J. D. (2008). Dissociation between explicit memory and configural memory in the human medial temporal lobe. *Cerebral Cortex*, 18(9), 2192–2207. <https://doi.org/10.1093/cercor/bhm245>
- Schankin, A., & Schubö, A. (2009). Cognitive processes facilitated by contextual cueing: Evidence from event-related brain potentials. *Psychophysiology*, 46(3), 668–679. <https://doi.org/10.1111/j.1469-8986.2009.00807.x>
- Schankin, A., & Schubö, A. (2010). Contextual cueing effects despite spatially cued target locations. *Psychophysiology*, 47(4), 717–727. <https://doi.org/10.1111/j.1469-8986.2010.00979.x>
- Schlagbauer, B., Mink, M., Müller, H. J., & Geyer, T. (2017). Independence of long-term contextual memory and short-term perceptual hypotheses: Evidence from contextual cueing of interrupted search. *Attention, Perception & Psychophysics*, 79(2), 508–521. <https://doi.org/10.3758/s13414-016-1246-9>
- Sewell, D. K., Colagiuri, B., & Livesey, E. J. (2018). Response time modeling reveals multiple contextual cuing mechanisms. *Psychonomic Bulletin & Review*, 25(5), 1644–1665. <https://doi.org/10.3758/s13423-017-1364-y>
- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, 12(11), 411–417. <https://doi.org/10.1016/j.tics.2008.07.006>
- Smyth, A. C., & Shanks, D. R. (2008). Awareness in contextual cuing with extended and concurrent explicit tests. *Memory & Cognition*, 36(2), 403–415. <https://doi.org/10.3758/MC.36.2.403>
- Töllner, T., Conci, M., Rusch, T., & Müller, H. J. (2013). Selective manipulation of target identification demands in visual search: The role of stimulus contrast in CDA activations. *Journal of Vision*, 13(3), 23. <https://doi.org/10.1167/13.3.23>
- Töllner, T., Gramann, K., Müller, H. J., Kiss, M., & Eimer, M. (2008). Electrophysiological markers of visual dimension changes and response changes. *Journal of Experimental Psychology: Human Perception and Performance*, 34(3), 531–542. <https://doi.org/10.1037/0096-1523.34.3.531>
- Töllner, T., Rangelov, D., & Müller, H. J. (2012). How the speed of motor-response decisions, but not focal-attentional selection, differs as a function of task set and target prevalence. *Proceedings of the National Academy of Sciences of the United States of America*, 109(28), E1990–E1999. <https://doi.org/10.1073/pnas.1206382109>
- Töllner, T., Zehetleitner, M., Krummenacher, J., & Müller, H. J. (2011). Perceptual basis of redundancy gains in visual pop-out search. *Journal of Cognitive Neuroscience*, 23(1), 137–150. <https://doi.org/10.1162/jocn.2010.21422>
- Tseng, Y.-C., & Li, C.-S. R. (2004). Oculomotor correlates of context-guided learning in visual search. *Perception & Psychophysics*, 66(8), 1363–1378. <https://doi.org/10.3758/BF03195004>

- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428(6984), 748–751. <https://doi.org/10.1038/nature02447>
- Wang, B., & Theeuwes, J. (2020). Saliency determines attentional orienting in visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, 46(10), 1051–1057. <https://doi.org/10.1037/xhp0000796>
- Wascher, E., & Wauschkuhn, B. (1996). The interaction of stimulus- and response-related processes measured by event-related lateralizations of the EEG. *Electroencephalography and Clinical Neurophysiology*, 99(2), 149–162. [https://doi.org/10.1016/0013-4694\(96\)95602-3](https://doi.org/10.1016/0013-4694(96)95602-3)
- Wolfe, J. M. (2019). Visual attention: The multiple ways in which history shapes selection. *Current Biology*, 29(5), R155–R156. <https://doi.org/10.1016/j.cub.2019.01.032>
- Wolfe, J. M. (2021). Guided search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, 28, 1060–1092. <https://doi.org/10.3758/s13423-020-01859-9>
- Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, 1(3), 1–8. <https://doi.org/10.1038/s41562-017-0058>
- Woodman, G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception & Psychophysics*, 72(8), 2031–2046. <https://doi.org/10.3758/APP.72.8.2031>
- Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, 400(6747), 867–869. <https://doi.org/10.1038/23698>
- Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 29(1), 121–138. <https://doi.org/10.1037//0096-1523.29.1.121>
- Woodman, G. F., & Vogel, E. K. (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin & Review*, 15(1), 223–229. <https://doi.org/10.3758/PBR.15.1.223>
- Zinchenko, A., Conci, M., Hauser, J., Müller, H. J., & Geyer, T. (2020). Distributed attention beats the down-side of statistical context learning in visual search. *Journal of Vision*, 20(7), 4. <https://doi.org/10.1167/jov.20.7.4>
- Zinchenko, A., Conci, M., Töllner, T., Müller, H. J., & Geyer, T. (2020). Automatic (mis) guidance of visuo-spatial attention by acquired scene memory: Evidence from an N1pc polarity reversal. *Psychological Science*, 31(12), 1531–1543. <https://doi.org/10.1177/0956797620954815>
- Zivony, A., Allon, A. S., Luria, R., & Lamy, D. (2018). Dissociating between the N2pc and attentional shifting: An attentional blink study. *Neuropsychologia*, 121, 153–163. <https://doi.org/10.1016/j.neuropsychologia.2018.11.003>

How to cite this article: Chen, S., Shi, Z., Zinchenko, A., Müller, H. J., & Geyer, T. (2022). Cross-modal contextual memory guides selective attention in visual-search tasks. *Psychophysiology*, 59, e14025. <https://doi.org/10.1111/psyp.14025>