Contextual cueing in co-active visual search: Joint action allows acquisition of task-irrelevant context

Xuelian Zang[1] · Artyom Zinchenko[2] · Jiao Wu[1] · Xiuna Zhu[2] · Fang Fang[3] · Zhuanghua Shi[2]

Abstract

Repeatedly presenting a target within a stable search array facilitates visual search, an effect termed *contextual cueing*. Previous solo-performance studies have shown that successful acquisition of contextual memories requires explicit allocation of attentional resources to the task-relevant repeated contexts. By contrast, repeated but task-irrelevant contexts could not be learned when presented together with repeated task-relevant contexts due to a blocking effect. Here we investigated if such blocking of context learning could be diminished in a social context, when the task-irrelevant context is task-relevant for a co-actor in a joint action search mode. We adopted the contextual cueing paradigm and extended this to the co-active search mode. Participants learned a context-cued subset of the search displays (color-defined) in the training phase, and their search performance was tested in the transfer phase, where previously irrelevant and relevant subsets were swapped. The experiments were conducted either in a solo search mode (Experiments 1 and 3) or in a co-active search mode (Experiment 2). Consistent with the classical contextual cueing studies, contextual cueing was observed in the training phase of all three experiments. Importantly, however, in the "swapped" test session, a significant contextual cueing effect was manifested only in the co-active search mode, not in the solo search mode. Our findings suggest that social context may widen the scope of attention, thus facilitating the acquisition of task-irrelevant contexts.

Keywords Co-active visual search · Contextual cueing · Joint action · Task-irrelevant context · Contextual learning · Memory

Introduction

Spatial environment consists of complex configurations of objects that we need to navigate daily to find certain items while ignoring others. Such configurations of objects are not randomly organized, but contain a certain regularity and consistency. The human visual system has developed an ability to employ these statistical regularities in an effective goal-directed manner, allowing past experience to optimize task performance (see Kersten et al., 2004). For instance, one can more likely detect a mailbox in a front yard than in a dining room (see Biederman, 1972; Conci & Müller, 2014; Davenport & Potter, 2004; Palmer, 1975). In lab settings, the role of learned contexts is typically demonstrated in a visual search paradigm, where search becomes more efficient over time when the target item is presented within spatially invariant search layouts – an effect that has been referred to as *contextual-cueing* (Chun & Jiang, 1998).

In a typical contextual-cueing task, observers search for a target letter "T" embedded in a set of distractor letters "L" (as depicted in an example in Fig. 1A). Unbeknown to the study participants, half of the trials contain repeated displays, while in the other half, the spatial arrangement of the distractors surrounding the target is randomly generated on each trial. The typical finding in this paradigm is that participants' search performance improves for repeated relative to random, non-repeated displays (Chun & Jiang, 1998; Sisk et al., 2019).

Xuelian Zang and Artyom Zinchenko contributed equally to this work.

✉ Xuelian Zang
 zangxuelian@gmail.com

✉ Fang Fang
 ffang@pku.edu.cn

✉ Zhuanghua Shi
 strongway@psy.lmu.de

[1] Center for Cognition and Brain Disorders, Affiliated Hospital of Hangzhou Normal University, Hangzhou 310015, China

[2] General and Experimental Psychology, Ludwig-Maximilians-Universität München, 80802 Munich, Germany

[3] School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China
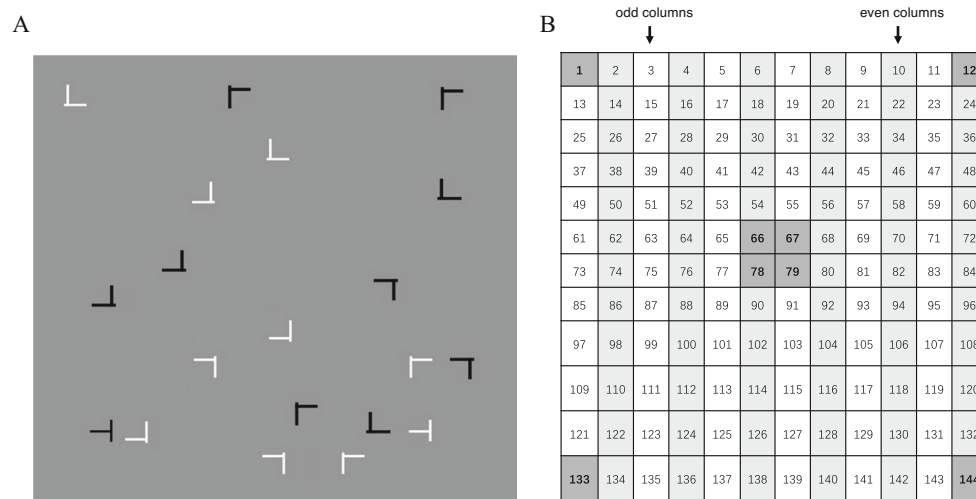
A

B

Fig. 1 (**A**) A sample of the search displays. (**B**) A spatial matrix with 144 cells was used in the experiment. One group (the black/white group) of search items was randomly distributed across the positions with an odd number, whereas items from the other group were randomly distributed across the positions with an even number. The eight dark grey shading cells located in the center and the corner were excluded for the target presentation. The grid, number, and grey shading were invisible during the actual experiment

Moreover, the ability to discriminate between repeated and non-repeated displays is typically at chance level even after a long period of learning (see Chun & Jiang, 1998; Jiang et al., 2019; Zang et al., 2017; but also see Annac et al., 2019; Kroell et al., 2019; Vadillo et al., 2016 for a different opinion), suggesting that implicit *context* memory of repeated target-distractor configurations can *cue* attention to the target location, enabling more efficient visual search.

The role of attention in context learning

The contextual cueing effect has been consistently associated with improved guidance of attention and context-related facilitation of attentional selection, as evidenced by studies using event-related potentials (ERPs) of encephalography (Johnson et al., 2007; Schankin & Schubö, 2009, 2010; Zinchenko, Conci, Töllner, et al., 2020b). For example, in a recent work, Zinchenko, Conci, Töllner, et al. (2020b) used electroencephalography (EEG) to show that learned contexts can aid attentional processing at early 80 ms (N1pc) post display onset (see also Summerfield et al., 2011). Furthermore, Johnson et al. (2007) showed that repeated, as compared to non-repeated, contexts elicit enhanced amplitudes of early attention-specific posterior-contralateral-negativity (PCN, also known as N2pc; Luck & Hillyard, 1994; Töllner et al., 2015), a negative-going ERP component that is used to track the allocation of attention. The enhanced PCN for repeated relative to novel contexts indicates superior attentional target selection with repeated contexts (see also Schankin & Schubö, 2009). The magnitude of the behavioral contextual cueing effect is tightly linked to the neurophysiological correlates of

attentional selection, which speaks in favor of an attention-related origin of the contextual cueing effect (but see Kunar et al., 2007). Collectively, these studies showed that information encoded into episodic long-term memory can capture attention and guide visual search (see also Nickel et al., 2020).

Although it has been proposed consistently that repeated contexts boost selective attention and subsequently guide visual search, the effect seems to work the other way around too: the availability of selective attentional resources can influence acquisition of invariant spatial configurations (e.g., Jiang & Chun, 2001; Jiang & Leung, 2005). For instance, the contextual cueing effect was overall weaker and developed slower when the search display was randomly divided into two subsets by color or size relative to the homogeneous display (Conci & von Mühlenen, 2011). This result suggested that segmentation of a single display into various subsets increases competition for attentional resources (i.e., each segment attracts attention) and in turn weakens contextual learning. Further, segmentation of search displays could block context acquisition, as shown by Jiang and Chun (2001), who manipulated the task relevance (task-relevant vs. -irrelevant) of two color subsets (a repeated and a non-repeated) in a display, and observed a contextual cueing effect only when the repeated context was task-relevant, not when it was task-irrelevant (ignored). Jiang and Leung (2005) took this experiment one step further by introducing a subsequent transfer phase, where the colors of previously relevant and irrelevant items were swapped after the initial training phase. The authors observed improved search time performance for the previous task-irrelevant but now relevant repeated context when it was paired with non-repeated configurations (i.e., the ignored-old condition), but not

when it was paired with configurations that were repeated too (i.e., the both-old condition). The authors proposed that when both task-relevant and -irrelevant contexts are invariant and presented together, the task-relevant context blocks the learning of the task-irrelevant context, consistent with an associative blocking effect in learning (Kamin, 1969).

Note that the observed blocking effect in Jiang and Leung (2005) could be explained by task-irrelevant feature suppression. In more detail, it has been shown previously that participants are able to filter out consistently task-irrelevant features (e.g., a task-relevant subset of items being consistently black has been shown to suppress a task-irrelevant white subset strongly; see Gaspelin et al., 2015, 2017; Vatterott & Vecera, 2012). Thus, the blocking effect shown in previous studies (e.g., Jiang & Chun, 2001; Jiang & Leung, 2005) can at least partially be explained by the suppression of task-irrelevant color, given that the task-relevant and -irrelevant colors were never changed for a given observer and the visual system could adapt to ignore a certain color in a top-down manner (Vatterott & Vecera, 2012). In line with this hypothesis, Geyer et al. (2010) showed that alternating task-relevant and -irrelevant colors across trials also resulted in a significant learning effect for unattended items. Taken together, top-down allocation of attentional resources appears to be essential for the acquisition of contextual regularities, and/or for the retrieval of contextual memories. In other words, task-irrelevant context is likely to be learned when attention is shared across the task-relevant and -irrelevant features.

Shared focus of attention is very common in social interactions. For instance, agents performing joint-action tasks are able to form shared representations, which greatly strengthens attentional processes of both co-actors (see below; Sebanz et al., 2006; Szymanski et al., 2017; Vesper et al., 2017). However, the effect of joint attention under social conditions has been seldomly investigated in the previous literature on acquisition and expression of context memories. Therefore, the present study focused on the joint attention induced by a co-active search (a social context) and its role in context learning and expression.

Attention and memory in joint tasks

Joint attention is an important component of social cognition. Agents involved in a sharing task/event are cognizant of each other's task/goal, and quasi-automatically form shared representations of the task (Sebanz et al., 2006; Vesper et al., 2017). A basic mechanism underlying these shared representations is that one's attention can be distributed to where the co-actors are attending (i.e., joint attention). Take a football game as an example, players would search for and attend to the moving ball as well as the intentions of other players who may catch or pass the ball. The performance of the whole team could be promoted by the shared context knowledge, such as team members' roles and understanding of teammates' intentional movements (e.g., response directions, rhythm, strategies etc.).

The influence of joint attention on the joint-task performance has been demonstrated in studies where co-actors concentrate on different (and potentially conflicting) properties of the same stimulus. For instance, Böckler et al. (2012) showed that participants' performance was significantly slower when pairs of participants had to focus on different features (local, global) of a Navon stimulus, suggesting that the co-actor's attentional focus induced a conflict where prepotent responses were actually incorrect (e.g., when the co-actor's global stimulus identity interfered with the participant's task-relevant, local stimulus identity). Importantly, this effect disappeared in a solo condition without a co-actor even when local and global stimulus features remained incongruent (Böckler et al., 2012). In a different study, Sebanz et al. (2005) asked pairs of co-actors to perform two different tasks on a stimulus (a color task and a hand-pointing task), and found the response time was prolonged in the joint relative to single actor condition, despite the fact that participants could not observe each other's actions directly. Together, these studies showed that co-actors in a joint task share each-other's task representation and attend jointly to tasks for both co-actors. In other words, participants' performance in a social context is influenced by their co-actors' task-representation, which could be achieved via modulation of attention.

Apart from modulation of attention, joint task performance could also facilitate participants' encoding of information and memory performance (e.g., Verga & Kotz, 2017). For instance, Eskenazi et al. (2013) asked participants to perform a categorization task either alone or in pairs. The main task was then followed by a surprise free-recall test where participants had to recall items from the categorization task. The results from the recall test revealed that participants performed better for their own items as well as for those items that were relevant for the co-actor (but irrelevant for themselves) in the previous joint task as compared to the individual task alone. Importantly, the performance was much poorer for items that were not relevant for both participants in the joint task condition. Interestingly, participants still remembered their co-actor's items, even when there was a financial incentive to concentrate exclusively on their own items. This suggests that having to process a task together with a co-actor may increase the relevance of the co-actor's subset of items and thus facilitate implicit acquisition of such invariant contexts. Note, however, that Eskenazi et al. (2013) has explored the influence of joint action on explicit memory performance, and adopted a relatively easy joint action task where one participant categorized a word and the other participant was simply observing. This leaves the open question of whether a similar phenomenon would be observed in implicit learning, such as contextual

cueing, in a co-active search mode, in which both observers have their own target. We hypothesize that if the task-relevant context blocks learning of the self-irrelevant but co-actor-relevant context (Jiang & Leung, 2005), swapping the subsets would show no contextual cueing for the previously task-irrelevant context. By contrast, if the co-active search mode leads to a shared focus of attention, one may observe additional contextual cueing for the co-actor-relevant context.

In a recent study, Sakata et al. (2021) attempted to answer this question using a joint contextual cueing paradigm, in which pairs of *participants* were instructed to simultaneously search for a single target. The cueing effect was observed earlier in the co-active search condition (starting from Epoch 1) relative to the solo condition (Epoch 5). In another experiment of that work, the displays were split into two color-defined subsets, and the two co-actors had to search for a target within their respective subsets. A crucial manipulation was introduced when one subset of items was repeated and the other non-repeated (analogous to the attended-old and ignored-old condition in Jiang & Leung, 2005). As compared to the baseline condition (both subsets were non-repeated contexts), the results revealed contextual cueing facilitation when participants searched within the repeated color-defined subset of items (i.e., attended-old) rather than within the non-repeated subset (i.e., ignored-old). The authors then concluded that joint action does not facilitate context learning. However, Sakata et al. (2021) could not effectively rule out joint action facilitation in implicit contextual learning since participants who searched within the non-repeated subset may still have (i) learned their partners' relevant context (i.e., the partner's target-distractors association) or (ii) formed an "ignored-context" association (associate one's own target and partner's repeated distractor context). Importantly, both of these potential associations could only be expressed when the ignored-context becomes task-relevant (as proposed in Jiang & Leung, 2005). Unfortunately, such a possibility was not accounted for in Sakata et al. (2021). In addition, similar to previous studies (Jiang & Chun, 2001; Jiang & Leung, 2005) that fixed the color of the task-relevant and -irrelevant subsets of search items for each individual participant, Sakata et al. (2021) also pre-assigned the colors of these subsets throughout the whole experiment. Thus, the lack of co-active search benefit in Sakata et al. (2021) could also be a result of perceptual filtering bias (Gaspelin et al., 2015, 2017; see also Geyer et al., 2010; Vatterott & Vecera, 2012), which outpaces any contextual learning of task-irrelevant context under joint attention.

To explore the role of joint-task performance in contextual learning, we conducted three contextual cueing experiments: Experiments 1 and 3 in a solo search mode, and Experiment 2 in a co-active search mode. Experiment 3 was a replication of Experiment 2 in the solo condition. In all three experiments, task-relevant and -irrelevant subsets of items (note that task-irrelevant subsets were relevant for the co-actor in the co-

active search condition) were used to examine contextual cueing effects. Importantly, we dynamically adjusted the color of task-relevant (and irrelevant) subsets on a trial-by-trial basis, such that perceptual filtering (Gaspelin et al., 2015, 2017; Vatterott & Vecera, 2012) of one color was not possible. We hypothesized that if the lack of learning of task-irrelevant context in previous studies is due to the perceptual filtering bias (e.g., swapping colors across trials), then lifting such a bias in the solo-condition would result in the learning of a task-irrelevant context in this condition. In addition, in each experiment we explicitly tested the task-irrelevant context in a transfer session, where it became task-relevant. If participants could learn task-irrelevant but co-actor relevant subsets in the learning phase, we would expect a significant cueing effect in the transfer phase in the joint-task, but not in the solo condition.

Experiment 1

Method

Participants

Sixteen naive volunteers from LMU Munich (14 females; mean age ± standard error: 25.38 ± 4.65 years) took part in Experiment 1, and were paid for their participation. All of them reported normal or corrected-to-normal visual acuity, and gave written informed consent prior to the experiment. Based on effect size measures provided in previous studies (Geringswald et al., 2015; Zang et al., 2015; Zang et al., 2018; Zellin et al., 2011; Zinchenko et al., 2018), our sample size was appropriate to detect an f(U) effect size of 0.816 with 85% power ($\eta_p^2 = 0.4$, groups = 2, number of measurements = 4), given an alpha level of .05 and a non-sphericity correction of 1. The Ethics Committee of the Department of Psychology of LMU Munich approved the current study.

Apparatus and stimuli

The experiment was conducted in a quiet, dim cabin (26.5 cd/m^2). All stimuli were presented on a 21-in. CRT monitor, which was set at a viewing distance of 57 cm from a fixed chin rest. The monitor refresh rate was 100 Hz. Stimulus presentation and response recording were controlled by Matlab (Mathworks, Natick, MA, USA) programs with the Psychtoolbox (Brainard, 1997; Pelli, 1997).

Each search display comprised 20 items (1.0° × 1.0° in visual angle) presented on a grey background (RGB value = [128, 128, 128], see Fig. 1A). The display had two subsets of items (ten items/subset): the black subset (RGB value = [0, 0, 0], 0.26 cd/m^2) and the white subset (RGB value = [255, 255, 255], 97.74 cd/m^2). In each subset, there was a T-shaped

target and nine L-shaped distractors. All L-shaped distractors had a small offset of 0.1° at the line junctions, so they resembled the target "T." In each display, "L"s featured in one of four orthogonal rotations (0°, 90°, 180°, and 270°), while "T"s were rotated either 90° or 270° clockwise. Stimuli were placed in an invisible 12 × 12 square grid (see Fig. 1B). The items from one subset were randomly arranged in the cells with odd column numbers, while the other subset items were allocated randomly in the cells with even column numbers. The two subsets (in odd or even columns) were randomly assigned to the black and white color for both old and new configurations, and the color for each subset was balanced across each configuration. Targets were uniformly distributed except for the center four cells and the corner four cells (see Fig. 1B, marked with a dark grey color).

Design and procedure

The experiment comprised three sessions: a training (25 blocks), a transfer (five blocks), and a recognition (one block) session. Each block contained 24 trials, with half repeated and half random displays presented in a random order. Each display had two targets (one black and one white) and 18 distractors (nine for each color set). Participants were instructed to respond to only one target, with a text cue of the target color "Black" or "White" presented at the start of each trial. The target-relevant color was balanced within each block, which means participants should respond to the white target in half of the trials and to the black target in the other half of the trials. For the repeated displays, the locations and orientations of distractors in both the task-relevant and -irrelevant groups, together with the locations of both targets, were kept constant and repeated once per block; for the random displays, the locations and orientations of distractors in both groups varied randomly. Note that the possible locations of the targets for both repeated and random configurations were fixed during the whole experiment to rule out any positional learning. The target orientations appeared randomly to the left or the right and were balanced across the whole experiment (for both repeated and non-repeated configurations) to rule out potential confounding of response learning on the target feature (i.e., orientation).

During the training session, each trial began with a cue ("Black" or "White") presented for 500 ms indicating the target-relevant color (see Fig. 2), and participants had to search for and identify that color target. Then, a fixation cross was shown for 500–700 ms randomly in the center of the display, followed by a search display presented for a maximum of 3 s or until the participant's response. This maximum of 3 s was selected based on past studies. Participants were required to discriminate the orientation of the "T" and to respond as fast and accurately as possible by pressing a key on the keyboard. Four numeric keys were used: "1" for the white

target orienting (pointing) to the left, and "3" for the white target orienting to the right, "8" for the black target orienting to the left, and "0" for the black target pointing to the right. After the response, the next trial started after an inter-trial interval (ITI) of 1,000–1,200 ms.

During the transfer session, the procedure was the same as in the training session. However, unbeknown to participants, the precue for a given repeated search configuration was switched from the training to the transfer session (from "Black" to "White" and vice versa), while the search configuration remained unchanged. Thus, we aimed to observe whether the subjects had learned the task-irrelevant display configurations in the training session. Finally, during the recognition session, a set of 12 repeated configurations from the training session and a set of 12 newly generated configurations were presented. Participants were instructed to respond whether a given configuration had been presented in the training session or not (if yes, press "1," if no, press "3"). The display was presented until a response, or for a maximum of 5 s.

Prior to the formal experiment, participants practiced the task with a block of 24 trials in order to become familiar with the task. The design of the practice session was analogous with the first block of the formal experiment, but its configurations were never reused in the formal experiment. Participants were required to achieve a correct rate above 85%. An extra practice block would be administered if they did not reach the accuracy threshold (maximum of two practice blocks across all participants).

Results

Error rates

The overall mean error rate was 4.54% (4.59% and 4.32% of the training and test phases, respectively). In order to increase statistical power, blocks were grouped into six epochs, each containing five consecutive blocks, with Epochs 1–5 corresponding to the training session and Epoch 6 corresponding to the transfer session. For statistical analyses, the Greenhouse-Geisser correction was applied in case of violation of the sphericity assumption. Error rates during the training session were further submitted to a 2 × 5 repeated-measures ANOVA with factors Context (2, repeated vs. non-repeated) × Epoch (5, Epochs 1–5). Both main effects were significant: Epoch [*Epoch 1* = 7.71%, *Epoch 5* = 3.02%; $F(2.04, 30.54) = 9.56$, $p < .001$, $\eta_p^2 = .39$] and Context [repeated = 3.81%, non-repeated = 5.38%; $F(1, 15) = 21.66$, $p = .001$, $\eta_p^2 = .59$], while the interaction between Context and Epoch was not significant [$F(4, 60) = .66$, $p = .63$, $\eta_p^2 = .04$]. In other words, the mean error rates were lower for the repeated relative to the random configurations, and Epoch 5 had 4.69% fewer errors as compared to Epoch 1. The error rates
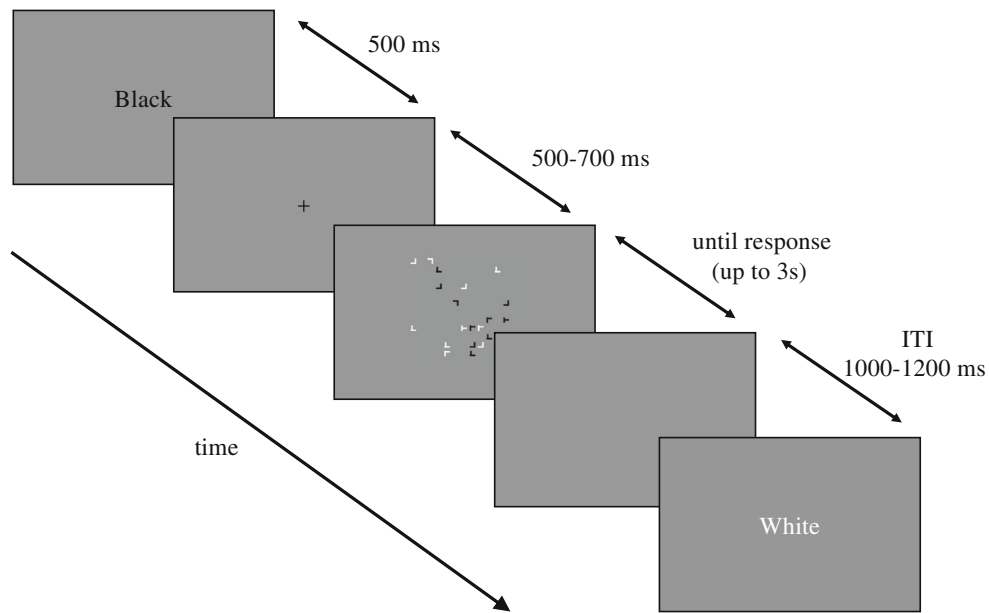
Fig. 2 Schematic illustration of the experimental paradigm. Two targets, black and white, were presented among the black and white distractors. In each trial, the participant was shown the word "Black" or "White" for the task set. The search display was presented for 3 s or until a response was made

in the transfer session were not significant between the repeated and non-repeated conditions, $t(15) =1.04$, $p = .31$, Cohen's $d = .37$, $BF_{10} = .63$. Mean error rates were 4.79% and 3.85% of the repeated and non-repeated display, respectively. Taken together, these results suggested participants improved their search performance with practice and with learning of repeated spatial context during the training session. However, the superior accuracy with the repeated display disappeared in the transfer session when the black and white contexts were swapped for the task relevance.

Mean response time (RT)

All the error trials and outliers (trials with response times (RTs) below 200 ms or above 2.5 standard deviations of the mean) were excluded from further analysis (1.46% trials were excluded). The overall mean RTs, after removal of error and outlier trials, are depicted in Fig. 3. A 2×5 repeated-measures ANOVA on mean RTs in the training session with the factors Configuration (repeated vs. random) and Epoch (1–5) revealed a significant main effect of Configuration [$F(1, 15) = 8.98$, $p = .009$, $\eta_p^2 = .37$] (see Fig. 3). The RTs were overall 87 ms faster for the repeated display (1,216 ms) than for the random display (1,303 ms), manifesting a robust contextual cueing. There was also a significant main effect of Epoch [$F(4, 60) = 17.15$, $p < .001$, $\eta_p^2 = .53$], with a 158 ms faster mean RT in Epoch 5 (1,189 ms) as compared to Epoch 1 (1,347 ms), showing a general procedural learning over time. The Configuration × Epoch interaction was also significant [$F(4, 60) = 3.38$, $p = .0155$, $\eta_p^2 = .18$], suggesting the development of contextual cueing effect over the course of training.

Post hoc tests revealed that the contextual effects were significant from Epoch 3 onward (mean cueing: 96, 105, and 146 ms for Epochs 3–5, respectively) [Epochs 3, $t(15) = 2.27$, $p = .039$, Cohen's $d = .57$; Epoch 4, $t(15) = 3.01$, $p = .009$, Cohen's $d = .75$; Epoch 5, $t(15) = 4.95$, $p < .001$,
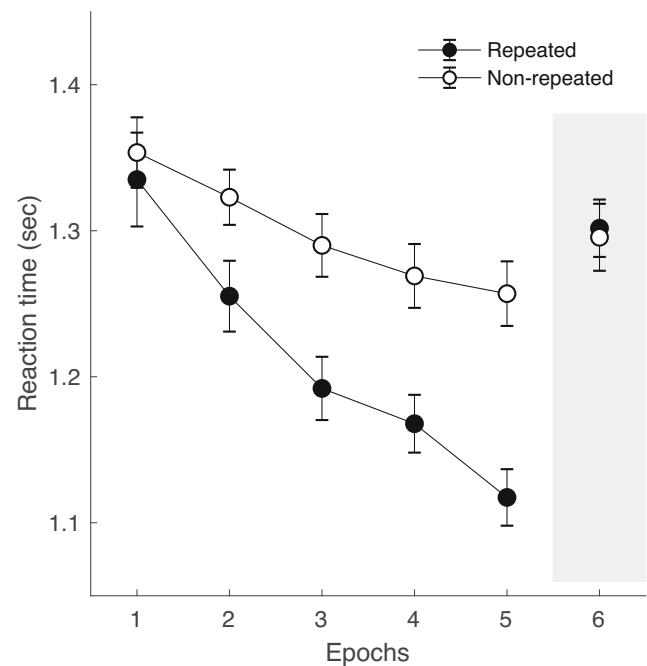

Fig. 3 Mean reaction times (RTs) with associated within-subject standard errors are plotted as a function of Epoch, separated for Configuration in Experiment 1. Each epoch contains five consecutive blocks. Epochs 1–5 are from the training session and Epoch 6 (shaded) the transfer session. The RTs for the repeated configuration are shown with the dark-filled circle lines, and for the random configuration with the open-circle lines

Cohen's $d = 1.24$], but not for Epoch 1 [$t(15) = .40$, $p = .70$, Cohen's $d = .10$, $BF_{10} = .27$] and Epoch 2 [$t(15) = 2.03$, $p = .061$, Cohen's $d = .51$, $BF_{10} = 1.29$]. Taken together, the results indicated both a procedure learning and a contextual learning, and an upward trend of contextual cueing effect over learning.

In the transfer session, the task cue was switched from the previous relevant to irrelevant subset with the same repeated displays. A direct comparison between RTs of the repeated versus random displays revealed no cueing effect [$t(15) = .16$, $p = .88$, Cohen's $d = .04$, $BF_{10} = .26$, mean RTs were 1,302 and 1,297 ms of repeated and non-repeated configuration respectively]. A further comparison of the contextual cueing from the last epoch in the training session (146 ms) to the transfer epoch (-5 ms) revealed a significant effect, $t(15) = 3.56$, $p = .003$, Cohen's $d = .89$, $BF_{10} = 15.50$, suggesting that the learned contextual cueing effect in the last epoch of the training session decreased significantly to the transfer epoch. In addition, participants' mean RTs for the non-repeated contexts were comparable between the last training epoch and the transfer epoch [$t(16) = .16$, $p = .88$, Cohen's $d = .04$, mean of 1,262.2 and 1,297.2 ms, respectively].

Taken together, the results suggest that the task-irrelevant subset that was repeatedly presented during the training session hasn't been learned, and no contextual facilitation was observed for the task-irrelevant subset in the transfer session.

Recognition

Participants' overall mean hit rate (mean = 43.75%, SD = 18.13%) was numerically lower than the mean false alarm rate (mean = 45.83%, SD = 18.26%), but the recognition sensitivity (d') was relatively small (mean = .07, SD = .78) and statistically indistinguishable from 0, $t(15) = .34$, $p = .74$, Cohen's $d = .09$, $BF_{10} = .27$. These results revealed no significant explicit memory of the contextual learning, implying that the spatial context was acquired implicitly.

Discussion

Experiment 1 showed that repeated spatial configurations were learned and established a robust contextual cueing effect even when we increased the number of search distractors (18 L-shape distractors and additional, irrelevant T-shape distractor). However, this contextual learning was mainly based on the task-relevant subset that contains the same color as the target. The task-irrelevant subset, by contrast, was not acquired. These results are consistent with the associative blocking account discussed above (Jiang & Leung, 2005; Kamin, 1969). In short, this account argues that invariant but task-irrelevant context is not acquired when presented

simultaneously with task-relevant invariant configurations. In other words, relevance-induced saliency of one invariant configuration can hinder acquisition of context memories for less salient (i.e., irrelevant) contexts.

Importantly, the current results may also extend previous findings to the condition where the two invariant contexts (relevant, irrelevant) are each associated with a unique target. Note that the experimental design in Jiang and Leung (2005) contained two invariant contexts and a single target, so that both the relevant and irrelevant contexts competed for the association with the target item. By contrast, our current findings show that associative blocking may hinder contextual learning in general, without the necessity to compete for a single target.

To conclude, Experiment 1 serves as a baseline condition to investigate whether co-working would modulate the contextual learning. To answer this question, in a follow-up experiment, we instructed two participants to search for different targets in the same display; however, one searched for the white target while the other searched for the black (and the other way around, randomly). Thus, the task-relevant subset for one participant is a task-irrelevant subset for their partner and vice versa.

Experiment 2

Method

Participants

Sixteen pairs of participants, ten pairs (i.e., 20 participants) from LMU Munich, Germany (12 females; mean age: 20.54 ± 3.15 years), and six pairs (i.e., 12 participants) from Hangzhou Normal University, China (11 females; mean age: 19.75 ± 1.36 years) took part in the experiment, and were paid for their participation. All of them had normal or corrected-to-normal visual acuity, and gave written informed consent before the experiment started. The experiment was approved by the Ethics Committee of the Department of Psychology of LMU Munich, Germany, and the Institutes of Psychological Sciences in Hangzhou Normal University, China.

Design and procedure

Experiment 2 aimed to investigate whether there is a mutual influence between paired subjects. The experiment design was essentially identical to that of Experiment 1, except that there were two participants who performed the task simultaneously during the whole experiment (including practice, training, transfer, and recognition sessions). Importantly, the pairs of participants were strangers before the experiment, and they
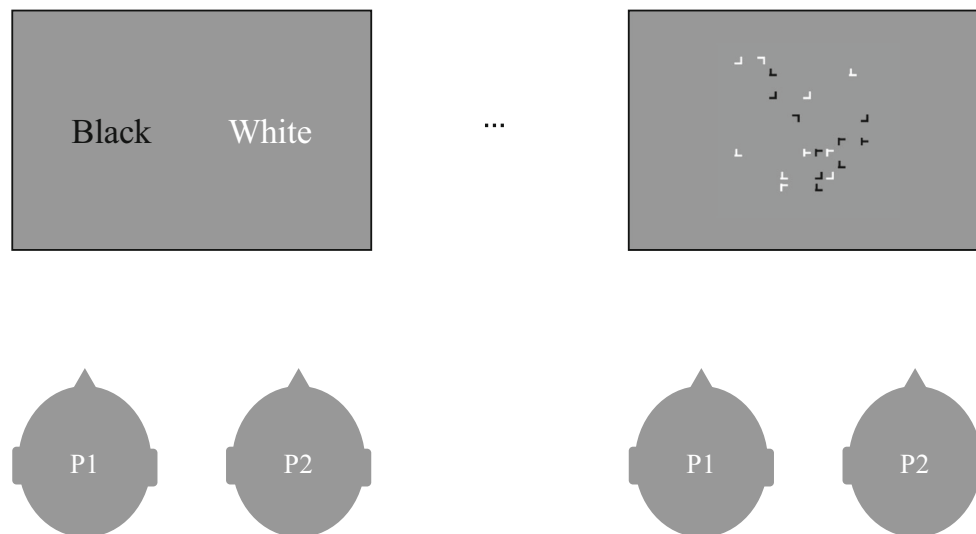
Fig. 4 An example of a cue display and the search display for paired subjects. In this example, participant P1 saw a "Black" cue on the left half of the screen, so black became their task-relevant color in this trial. Likewise, participant P2 searched for the white target

were required to search for different (black or white) targets. Thus, two cues were shown in each trial (see Fig. 4). Participants sat in front of the monitor and responded with the same keyboard. At the beginning of each trial, they received a pair of English words (see Fig. 4), and the word presented on their side indicated their target color for the trial. The left observer was instructed to respond with the "1" and "3" keys for the target pointing to the left and the right, respectively, whereas the right observer responded with the "8" and "0" keys for the target pointing to the left and right, respectively. Both of them were instructed to use their right hand to respond. Consequently, the task-relevant target for one participant was the task-irrelevant target for the other. The visual display was presented until both participants made responses.

Results

Error rates

Since the participant's side (sitting left or right) and the stimuli color (white or black) were random in our design, these factors were not included in the following analysis. The overall mean error rate and mean outlier (defined as RTs of the trial below 200 ms and longer than 2.5 times standard deviation of the mean response time) were low: mean error rate 3.22%, mean outlier 0.82%.

Repeated-measures ANOVA for mean error rates with factors Context (2, repeated vs. non-repeated) × Epoch (5, Epochs 1–5) showed a significant main effect of Epoch [*Epoch 1* = 5.16%, *Epoch 5* = 2.03%; $F(2.06, 63.91) = 8.39$, $p < .01$, $\eta_p^2 = .21$] but not of Context [repeated = 2.97%, non-repeated = 3.24%; $F(1, 31) = 1.0$, $p = .33$, $\eta_p^2 = .031$] and

Context × Epoch interaction [$F(4, 124) = .38$, $p = .83$, $\eta_p^2 = .012$], suggesting the accuracy was increased with general procedural learning. Those trials with erroneous responses and outliers were excluded for the following RT analyses. The mean error rates in the test session were not significantly different between the repeated (2.45%) and non-repeated (2.65%) configurations, $t(31) = .40$, $p = .69$, Cohen's $d = .07$, $BF_{10} = .20$.

Mean RT

Figure 5 depicts the mean RT as a function of Epoch, separated for different contexts. The RT data were submitted to a 2 × 5 repeated-measures ANOVA with Configuration (repeated vs. random) and Epoch (1–5) as factors, which revealed a significant main effect of Configuration [$F(1, 31) = 14.76$, $p < .01$, $\eta_p^2 = .32$], with response being faster for the repeated configurations (1,255 ms) relative to the random configurations (1,321 ms). The main effect of Epoch was also significant [$F(2.41, 74.62) = 91.74$, $p < .01$, $\eta_p^2 = .75$]. Mean RT was 242 ms faster in Epoch 5 (1,200 ms) as compared to Epoch 1 (1,442 ms), suggesting a general speeding-up of task performance over time. We also found a significant Configuration × Epoch interaction [$F(4, 124) = 4.72$, $p < .01$, $\eta_p^2 = .13$], implying that the contextual cueing effect developed over the course of training. Post hoc tests showed that the contextual effect was significant from Epoch 3 onward [Epoch 3, $t(31) = 3.59$, $p = < .01$, Cohen's $d = .63$, mean cueing effect of 81.6 ms; Epoch 4, $t(31) = 4.12$, $p < .01$, Cohen's $d = .74$, mean effect of 76.6 ms; Epoch 5, $t(31) = 5.19$, $p < .01$, Cohen's $d = .92$, mean of 102.7 ms], but not in Epoch 1 [$t(31) = .96$, $p = .34$, Cohen's $d = .17$, $BF_{10} = .29$] and Epoch 2 [$t(31) = 1.92$, $p = .06$, Cohen's $d = .34$, $BF_{10} = .96$].
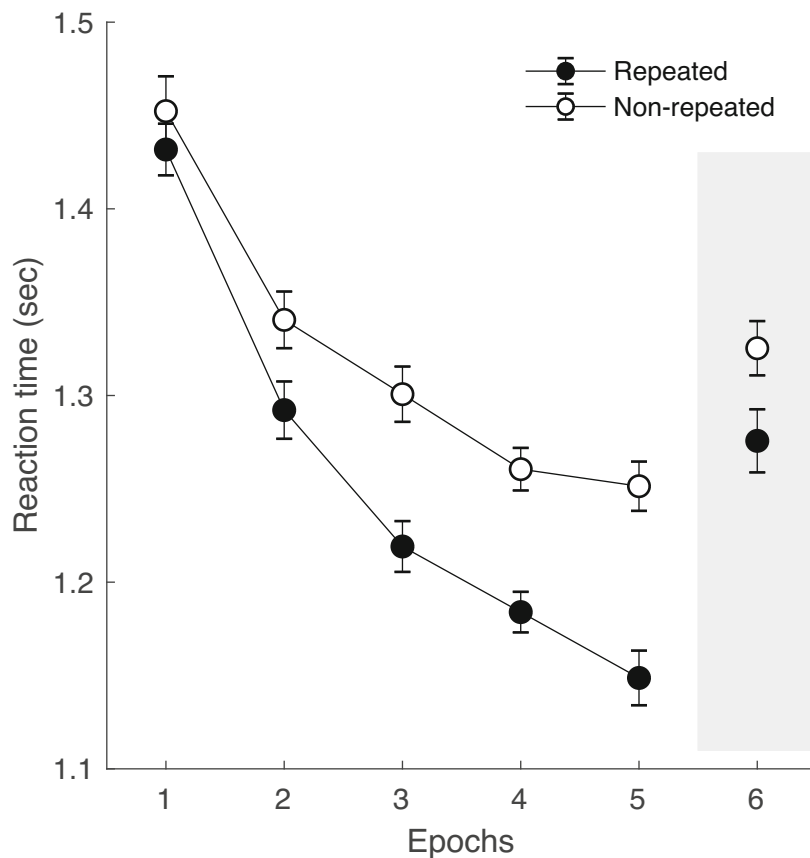
Fig. 5 Mean reaction times (RTs) and their associated within-subject standard errors are shown as a function of Epoch and Configuration in Experiment 2. Similar to Experiment 1, Epochs 1–5 were from the training session and Epoch 6 from the transfer session. The mean RTs for repeated configurations were indicated by dark-filled lines, whereas mean RTs for the non-repeated configurations were indicated by open-circle lines

Paired t-test for the RTs in the transfer session revealed a significant contextual cueing effect [$t(31) = 2.59$, $p = .02$, Cohen's $d = .46$]. The response was 49.6 ms faster in the repeated condition (1,276 ms) as compared to the random condition (1,325 ms), showing a small but clear contextual cueing. When comparing the contextual cueing from the last epoch in the training session to the transfer epoch, the cueing effect was numerically smaller in the transfer session (49.6 and 102.7 ms, respectively), but this difference was only marginally significant [$t(31) = 1.85$, $p = .05$, Cohen's $d = .33$, $BF_{10} = .86$]. Interestingly, participants' mean RTs in the transfer epoch were significantly longer relative to the last epoch of the training (i.e., Epoch 5) not only for the repeated displays [$t(31) = 5.42$, $p < .01$, mean difference of 127 ms] but also the non-repeated displays [$t(31) = 3.35$, $p < .01$, mean difference of 74 ms], suggesting that the color swap resulted in overall slower responses.

Note that contextual benefit in the transfer session might largely come from those fast trials in which observers had some additional exploration time to explore the co-actor's subset of items. To rule out this possibility, we ran a group analysis to examine whether "faster" participants from the training session had a larger contextual

cueing effect in the transfer phase compared to their "slower" co-actors. In the analysis, participants were classified into a "faster" and a "slower" group by their general mean reaction times during training. That is, participants with a faster mean reaction time in a pair were to the "faster" group; while the other participants were distributed to the "slower" group. The result revealed a comparable contextual cueing effect between two groups, [$t(30) = .94$, $p = .36$, $BF_{10} = .47$] in the transfer session (see Fig. S1 in the Online Supplementary Material (OSM)). We further used an alternative classification, a "faster" co-actor who responded faster in more than half of the trials relative to a "slower" co-actor. The result revealed the same comparable contextual cueing between two groups in the transfer session, $t(30) = .67$, $p = .51$, $BF_{10} = .40$. Taken together, the analyses confirmed that a faster response, together with the resulting prolonged exposure time, could not be the origin of contextual transfer effect.

Recognition

The recognition data for one participant were excluded for the analysis as the participant reported all the displays to be new

(i.e., continuously press one response key). The analysis of the rest of the data showed the same mean hit and false alarm rates (*mean* = 48.66%, *SD* = 14.6% and *mean* = 48.66%, *SD* = 19.96%, respectively). The mean recognition sensitivity *d'* (*mean* = .018, *SD* = .57) was not significantly deviated from 0 [$t(30) = .17$, $p = .86$, Cohen's $d = .031$, $BF_{10} = .19$], suggesting participants did not explicitly remember the learned repeated displays.

Discussion

The main finding of Experiment 2 was that the task-irrelevant part of the repeated configuration during training was also learned in the joint task condition. The results suggest that the social context motivated participants to allocate some attention to the task-irrelevant context that was relevant for their co-actors, which resulted in improved performance during the transfer phase when that task-irrelevant context became task-relevant. Apart from social context, one difference between Experiments 1 and 2 is the instructions. While only one cue (White or Black) was presented prior to each trial in Experiment 1, two cues were presented prior to each trial in Experiment 2 (White and Black). Thus, it is possible that the second, task-irrelevant word could already prime participants' attention to the task-irrelevant color and thus induce learning of the irrelevant subset of items. Additionally, while participants in Experiment 1 were seated in the center in front of a computer, participants in Experiment 2 had to share one monitor and this could have affected their performance. To rule out those nuisance factors that might contribute to the transfer effect, we conducted a further control in Experiment 3.

Experiment 3

Experiment 3 was identical to Experiment 2 except that only one participant was involved in the experiment. Sixteen participants (13 females; mean age: 21.44 ± 1.75 years) from Hangzhou Normal University took part in the experiment, and were paid for their participation. Half of the participants sat on the left chair, and followed the cue on the left side, while the other half of the participants sat on the right chair, and followed the cue on the right side. Participants' seating positions (left vs. right) were counterbalanced during the experiment.

Results and discussion

Error rates and outliers

The overall ratio of error responses and outlier trials (RT below 200 ms or above 2.5 standard deviations from the mean)

was low 4.90% (5.09% and 3.96% of the training and test phases, respectively). Error rates in the training session were submitted to a 2 × 5 repeated-measures ANOVA with factors Context (2, repeated vs. non-repeated) × Epoch (5, Epochs 1–5). Both main effects were significant: Epoch [*Epoch 1* = 8.75%, *Epoch 5* = 2.97%; $F(1.97, 29.47) = 8.94$, $p < .001$, $\eta_p^2 = .37$] and Context [repeated = 4.25%, non-repeated = 5.94%; $F(1, 15) = 12.04$, $p < .01$, $\eta_p^2 = .45$], while the interaction of Context × Epoch was not significant [$F(4, 60) = 5.61$, $p = .53$, $\eta_p^2 = .72$]. The pattern remained the same as in Experiment 2. The mean error rates were lower for the repeated relative to the random configurations, and Epoch 5 had 5.78% fewer errors as compared to Epoch 1. The error rates did not differ between the repeated (3.65%) and non-repeated (4.27%) configurations in the test session, $t(15) = .69$, $p = .50$, Cohen's $d = .17$, $BF_{10} = .31$. Taken together, these results suggested that participants improved their search performance with practice and with learning of repeated spatial context during the training session, while the accuracy difference disappeared in the test session when the task set was switched.

Mean RT

All the error and outlier trials were excluded from further analysis. The overall mean RTs are depicted in Fig. 6. A 2 × 5 repeated-measures ANOVA on mean RTs from the training session with the factors Configuration (repeated vs. random)
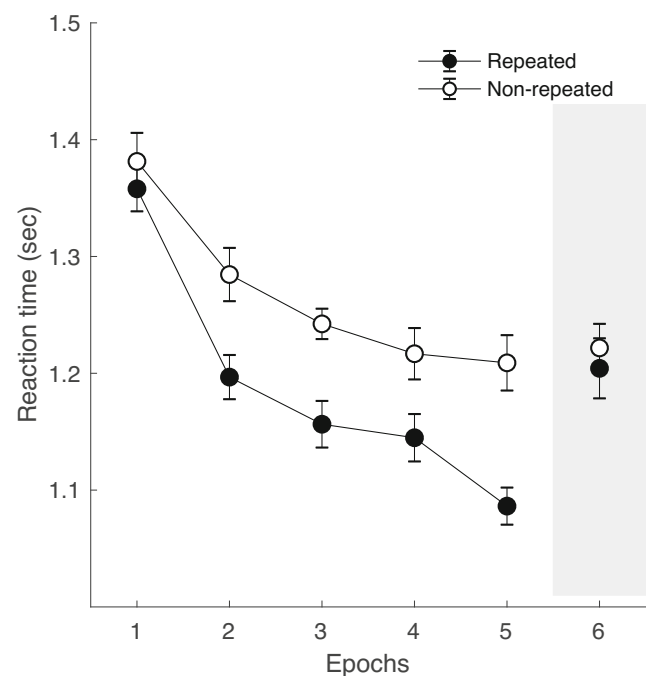

Fig. 6 Mean reaction times and within-subject standard errors are shown as a function of Epoch, separated for the repeated and non-repeated configurations in Experiment 3. Epochs 1–5 were from the training session and Epoch 6 (shaded) from the transfer session

and Epoch (1–5) revealed a significant main effect of Configuration [$F(1, 15) = 12.05$, $p < .01$, $\eta_p^2 = .45$]. The RTs were overall 78 ms faster for repeated (1,188 ms) than random displays (1,267 ms), showing a robust contextual cueing effect. There was also a significant main effect of Epoch [$F(4, 60) = 31.04$, $p < .01$, $\eta_p^2 = .67$], with 222 ms faster mean RT in Epoch 5 (1,148 ms) as compared to Epoch 1 (1,370 ms), showing a general procedural learning over time. The Configuration × Epoch interaction was also significant [$F(4, 60) = 3.33$, $p = .02$, $\eta_p^2 = .18$], showing the development of contextual cueing effect over the course of the training. Post hoc tests showed that the contextual effect was significant from Epoch 2 onward [Epoch 2: $t(15) = 2.69$, $p = .02$, Cohen's $d = .67$, $CC = 88$ ms; Epoch 3: $t(15) = 3.19$, $p < .01$, Cohen's $d = .80$, $CC = 86$ ms; Epoch 4: $t(15) = 2.70$, $p < .01$, Cohen's $d = .68$, $CC = 72$ ms; Epoch 5: $t(15) = 3.76$, $p < .01$, Cohen's $d = .94$, $CC = 123$ ms], but not in Epoch 1: $t(15) = 1.01$, $p = .33$, Cohen's $d = .25$, $BF_{10} = .40$. Taken together, the results were indicative of both a procedure learning and a contextual learning, as well as an upward trend of contextual cueing effect over time, which was similar to the finding of Experiment 2.

What differs from Experiment 2 is the results of the transfer session. A direct comparison between RTs of the repeated versus non-repeated configurations in the transfer session revealed no contextual cueing effect, $t(15) = .75$, $p = .47$, Cohen's $d = .19$, $BF_{10} = .33$. The mean RTs were comparable, 1,204 and 1,222 ms, for the repeated and non-repeated configuration, respectively. The mean RTs of the non-repeated context were comparable to the mean RTs in the last training epoch [$t(15) = .41$, $p = .69$, Cohen's $d = .10$, mean of 1,209 and 1,222 ms, respectively]. By contrast, the mean RTs of the repeated context increased significantly from the last training epoch to the transfer epoch [$t(15) = 3.77$, $p < .01$, Cohen's $d = .94$, mean of 1,086 and 1,209 ms, respectively]. The reduction of the contextual cueing was also confirmed from the comparison between the last epoch in the training session and the transfer epoch, showing a significant decreased cueing effect from the training to the transfer epoch, $t(15) = 2.22$, $p = .04$, Cohen's $d = .56$, mean of 123 ms and 17 ms, respectively. These results suggested that a significant cueing effect from the training diminished to non-significance in the transfer session when the task set was switched, which was in contrast to the finding in Experiment 2, when the paired participants performed the tasks together. Thus, the null finding of the transfer effect from the solo participant effectively ruled out the transfer effect obtained in Experiment 2 was due to the shared cues or other nuisance factors.

Recognition

Participants' overall mean hit rate (*mean* = 52.36%, *SD* = 12.59%) was numerically lower than the mean false alarm rate (*mean* = 53.32%, *SD* = 13.98%), but the recognition sensitivity (*d'*) was relatively small (*mean* = .031, *SD* = .50) and not significantly deviated from 0, $t(15) = .25$, $p = .81$, Cohen's $d = .06$, $BF_{10} = .26$. Again, here we showed the same implicit learning.

Omnibus analysis

The above analyses were done separately for individual experiments. In order to have a more complete view on the role of joint attention in contextual cueing, we further conducted an omnibus analysis between the co-active and the solo groups. We pooled Experiments 1 and 3 together as the solo group, and Experiment 2 as the co-active group. Both groups had the same number of participants (N = 32).

As reported in previous contextual cueing studies, there were two types of individuals: learners and non-learners. Non-learners are participants whose contextual cueing effect was < 0 in the last epoch of the training session. It has been shown that up to one-third of all participants may fail to obtain any contextual cueing effects (Zellin et al., 2014; Zinchenko, Conci, Hauser, et al., 2020a). For studies focusing on the transfer effect (e.g., Zellin et al., 2014), comparison among conditions is often done for the learners only, given that only learners exhibit contextual cueing in the training session. Similarly, here we selected learners from the co-active and solo groups and compared their transfer effects. We used the last epoch of the training session, Epoch 5, to classify the learner and non-learner. There were seven non-learners out of the 64 participants (four from the solo group, and three from the co-active group). Given that the number of non-learners was small, excluding those non-learners had little impact on the RT patterns and statistical results we reported in the previous section (for comparison, we also included analysis of the training session).

In the training session (Epochs 1–5), mean RTs were submitted to a 2 × 5 × 2 repeated-measures ANOVA with Context (repeated vs. non-repeated) and Epoch (1–5) as within-subject factor, and Search Mode (solo vs. co-active) as a between-group factor. The results showed significant main effects of Context [$F(1, 55) = 52.90$, $p < .01$, $\eta_p^2 = .49$, mean cueing effects of 94 ms and 84 ms of solo and co-active mode, respectively] and Epoch [$F(2.50, 137.48) = 111.44$, $p < .01$, $\eta_p^2 = .67$, mean procedural learning effect of 217 ms], but not of the Search mode [$F(1, 55) = .31$, $p = .58$, $\eta_p^2 < .01$], although the mean RTs of the co-active search mode were numerically slower than that of the solo mode (29 ms). The Context × Epoch interaction was also significant [$F(1, 55) = 12.35$, $p < .01$, $\eta_p^2 = .18$]. This is because contextual cueing effects increased with the progress of the experiments (see Figs. 3, 5 and 6). All the other interactions did not reach significance (all ps > .05).

In the transfer session (Epoch 6), repeated-measures ANOVA for mean RTs were applied with Context (repeated vs. non-repeated) as within-subject factor, and Search mode (solo vs. co-active) as between-subject factor. The results failed to reveal any significant main effects [Context: $F(1, 55) = 3.10$, $p = .08$, $\eta_p^2 = .05$; Search mode: $F(1, 55) = .21$, $p = .65$, $\eta_p^2 < .01$]. Interestingly, the interaction between Context and Search reached significance [$F(1, 55) = 4.87$, $p = .03$, $\eta_p^2 = .08$], which was mainly driven by a significant contextual cueing effect in the co-active search mode in Experiment 2 [$t(28) = 2.73$, $p = .01$, Cohen's $d = .51$, mean cueing effect of 56 ms], but it was non-significant in the solo mode Experiments 1 and 3 [$t(27) = .33$, $p = .75$, Cohen's $d = .06$, mean cueing effect of -6 ms].

To further identify contextual cueing differences from the training to the transfer session under the co-active and solo conditions, participants' mean contextual cueing effects in the last training Epoch 5 and the transfer Epoch 6 were submitted to a repeated-measures ANOVA with Epoch (5 vs. 6) as a within-subject factor and Search mode (solo vs. co-active) as a between-subject factor. The results showed a significant main effect of Epoch [$F(1, 55) = 32.15$, $p < .01$, $\eta_p^2 = .08$], and the interaction between Epoch and Search mode [$F(1, 55) = 5.44$, $p = .02$, $\eta_p^2 = .09$], but not the main effect of Search mode [$F(1, 55) = .63$, $p = .43$, $\eta_p^2 = .01$]. The mean cueing effects of Epochs 5 and 6 were 153 ms and -6.3 ms for the solo group, and 125 ms and 56 ms for the co-active group, respectively. Taken together, these results corroborate our conclusion that participants learned only the task-relevant context under the solo condition. Under the co-active search mode, by contrast, participants learned both the task-relevant and irrelevant contexts, but the amount of contextual cueing effect for the irrelevant context was weaker than that of the relevant context.

In order to examine whether target color switching from the training to the transfer session would induce any general cost on mean RTs, a further repeated-measures ANOVA was applied to the mean RTs of the *non-repeated* displays with Epoch (5 vs. 6) as within-subject factor, and Search mode (solo vs. co-active) as between-subject factor. The main factor of Epoch was significant [$F(1, 55) = 5.28$, $p = .03$, $\eta_p^2 = .09$], and the interaction between Epoch and Search mode was also significant [$F(1, 55) = 5.09$, $p = .03$, $\eta_p^2 = .09$]. However, the main effect of Search mode was not significant [$F(1, 55) = .21$, $p = .65$, $\eta_p^2 < .01$], which suggests that the RT pattern was comparable between the solo and coactive groups. The significant interaction was mainly caused by a significant larger RT in the transfer relative to the training session in the co-active group [$t(28) = 3.05$, $p < .01$, mean of 1,317 and 1,251 ms, respectively] but not in the solo group [$t(27) = -.03$, $p = .98$, mean of 1,259 and 1,258 ms, respectively]. This finding is consistent with the work of Vaskevich and Luria (2018), who showed that processing of blocks of mixed repeated

and non-repeated displays is slowed relative to processing of non-repeated displays alone. In other words, these findings imply that participants in the co-active but not in the solo group did learn co-actor relevant displays during the initial learning phase and treated contexts in the transfer block as a mixture of repeated and non-repeated displays.

To summarize, the omnibus analysis showed comparable response speeds and contextual learning in the initial training session. When the target context was swapped in the following transfer session, the contextual cueing effect in the co-active search mode remained significant and larger than that in the solo mode, confirming our conclusion that the co-active search mode activates joint attention to the co-actor-relevant but task-irrelevant context.

General discussion

The present study of three experiments examined the effect of joint action on context learning in visual search. Participants searched for a white or a black target (the color randomly assigned on each trial) within subsets of white and black distractors either in a solo search mode (Experiments 1 and 3) or in a co-active search mode (Experiment 2). Critically, all items (black and white) in the repeated displays (50% of all trials) kept their locations constant. In the transfer session, unknown to the participants, task sets of the relevant and irrelevant contexts were swapped and participants searched for previously irrelevant targets among distractors. We observed robust contextual cueing effects in the training session in all three experiments. Importantly, the contextual cueing diminished immediately in the transfer session of both solo search experiments, indicating that the task-irrelevant context was ignored in the training session of the solo condition. By contrast, a significant contextual transfer effect was observed in the joint-action search mode, suggesting the search for the task-irrelevant target could also be facilitated, probably because the task-irrelevant but co-actor's relevant context could also be learned in the co-active search mode. An equally plausible possibility is that participants learned an association between task-relevant context and task-irrelevant target.

The robust contextual cueing effect in the training session is consistent with previous studies where search items were grouped by color (e.g., Conci & von Mühlenen, 2011; Geyer et al., 2010), and studies with only a subset of search items being task-relevant and repeated (e.g., Geyer et al., 2010; Jiang & Chun, 2001; Jiang & Leung, 2005). The results of the transfer session in the *solo search* mode were also comparable to the previous study by Jiang and Leung (2005), showing no acquisition of the task-irrelevant context. It should be noted that in contrast to previous studies that used fixed color-subset assignments (Jiang & Chun, 2001; Jiang & Leung, 2005), the task-relevant subset of search items varied

randomly on a trial-by-trial basis in the present study. The color swapping was aimed at preventing any potential reliance purely on the color of search items (e.g., concentrating on white items only) and was used to weaken the possibility of perceptual filtering of task-irrelevant information. Thus, this manipulation tested whether associative blocking in previous works is at least partially caused by the perceptual filtering of task-irrelevant contexts (Gaspelin et al., 2015, 2017; Vatterott & Vecera, 2012). The current findings to some extent discard the perceptual filtering account, since participants in the solo condition showed no reliable post-transfer cueing effect, despite the randomized assignment of task-relevant colors on every trial. Nevertheless, it remains possible that some residual perceptual filtering may still remain effective in the solo-condition since participants could adjust the filtering template based on the pre-cue information on every trial (Conci & von Mühlenen, 2011; Zang et al., 2016). To summarize, the results of Experiments 1 and 3 replicated (most of the) previous findings on a lack of transfer effect for the task-irrelevant context and served as a baseline for further comparison with the co-active search mode.

The lack of the transfer effect in Jiang and Leung (2005) was explained via the "associative blocking" mechanism, according to which a salient cue (i.e., task-relevant repeated subsets) blocks an association with a less salient cue (task-irrelevant repeated subset). In other words, the coexistence of the salient and non-salient contexts inhibits the learning of the latter (Endo & Takeda, 2004; Geyer et al., 2021; Kamin, 1969). The associative blocking could account for the lack of post-transfer context effect in the two solo experiments in the current work. It should be noted that in contrast to previous studies that used a single target (Jiang & Chun, 2001; Jiang & Leung, 2005), the present study used two targets (one relevant and one irrelevant). The lack of context memory for the task-irrelevant target-context association extends the associative blocking account in the solo search to a condition where both task-relevant and -irrelevant contexts had their own respective targets relative to a single common target in Jiang and Leung (2005). That is, the task-relevant association blocked the acquisition of not only the association between task-irrelevant distractors and relevant target (as in Jiang & Leung, 2005), but also the irrelevant target-distractor association in solo tasks. Most importantly, we showed for the first time that the task-irrelevant context could be acquired in the co-active search mode (Experiment 2), which suggests that associative blocking was somehow weakened in the co-active search mode, enabling acquisition of a task-irrelevant context.

There are a number of potential reasons for the observed task-specific activation patterns in the transfer phase of the single and joint-search experiment. One possibility could be related to the widening of attention in social context (e.g., joint search in the current study). Social context of the joint task could extend the scope of attention (see Böckler et al., 2012;

Sebanz et al., 2003) and additionally encompass task-irrelevant but co-actor-relevant search items that were thus conjointly learned in the co-active search task. That is, participants may have allocated certain attentional resources to task-irrelevant subsets since they were actually relevant for the co-actor. This notion is consistent with a recent finding by Zinchenko, Conci, Hauser, et al. (2020a), who showed that a broader scope of attention could facilitate updating of learned context memories. In that study, the target was relocated to a new location after a reliable cueing effect was established in the initial learning phase. As a result of the relocation, there was an expected reduction in the strength of the cueing effect in participants from the *focused* search mode group, but not in the group with an induced *global* attentional state who showed an advantage even for target-relocated repeated contexts. Furthermore, extending the scope of attention to a task-irrelevant context has also been reported in a recent study from Zang et al. (2021). In that design, search displays were presented for short enough durations (300 ms) to impede the perceptual segmentation of the display into relevant and irrelevant subsets. In this case, task-irrelevant contexts have been learned already in the solo search mode, probably due to the expanded focus of attention (Zang et al., 2021). Collectively, these studies support the idea that the improved context learning in the joint task in Experiments 2 could be a result of socially induced widening of attention to both self-relevant and co-actor relevant contexts during the learning phase, thus yielding a reliable cueing effect in the subsequent transfer phase.

One may argue that participants in the co-active task may learn to associate the task-relevant context with both the task-relevant and -irrelevant targets (i.e., the associative learning account). This way, in the transfer session, the learned associations of previously relevant (but now-irrelevant) context and now-relevant target could still guide participants' attention, resulting in a reliable contextual cueing in this session of the joint task, even when the colors were swapped. On the one hand, this possibility may seem less likely since previous studies have shown that participants are unable to associate a single context with multiple target locations even when the search configurations were not divided into color-specific subsets of items (Zellin et al., 2011; Zellin et al., 2014; Zinchenko, Conci, Hauser, et al., 2020a). Additionally, the associative learning hypothesis would imply that participants not only search for a target within a task-irrelevant context, but also learn to differentiate between the irrelevant target and the irrelevant distractors. This is a demanding additional process and should have resulted in some performance cost, which we did not observe. On the other hand, it is also important to note that Zellin and colleagues alternated presentation of a single invariant context with variable target locations across subsequent blocks, and never paired invariant context with two targets simultaneously within a single trial. Additionally, it is

possible that the joint-action condition reduced the effect of colors during visual search, and participants did not need to differentiate the task-irrelevant target and task-irrelevant distractors. Therefore, the associative learning account (i.e., associating the task-relevant context with both the task-relevant and -irrelevant targets) still remains a possibility and should be addressed in future work. Finally, joint task performance, as any social interaction, may also involve changes in the level of arousal, social desirability, sense of cooperation or competition. All these factors are rooted in the joint task performance and could further influence attention and memory-related processes. However, the current work cannot address/control for potential influences of the mentioned factors. Therefore, future studies should take this point into account and examine what specific factors of joint task performance modulate context memories.

It should be noted that a recent study by Sakata et al. (2021) also attempted to examine the role of joint action in context-based memory, and the results of that study may first seem at odds with the current findings. However, there are critical differences between the two studies' goals and designs. The current study tested if social context would not only result in context memory for task-relevant configurations, but if it would also build additional contextual associations for co-actor-relevant items in a joint visual search paradigm. By contrast, Sakata et al. (2021) examined whether social context would enhance the association of a co-actor's repeated distractors and self-relevant targets, specifically when the self-relevant contexts were non-repeated. Thus, we have adopted exclusively the "both-old" design, while Sakata et al. (2021) concentrated on the "ignored-old" condition in the joint action paradigm (see Jiang & Leung, 2005, for the crucial distinction between the two conditions). The lack of the joint action effect in the work of Sakata could stem from the fact that self-relevant and co-actor-relevant subsets were segmented by color, and it is difficult to associate a relevant target and an irrelevant context across color-defined subsets. For instance, imagine a pair of people shopping together, one searching for a bottle of shampoo, while the other one is simultaneously searching for a body lotion. When the layout of the shampoo shelf is random (analogous to Sakata's design), searching for the target shampoo would unlikely be facilitated by the adjacent body lotion shelf with a constant layout. However, when both the shampoo and the body lotion shelves keep item layouts fixed (analogous to our design), a possible joint action facilitation may occur: the pair may widen the scope of attention, associate configurations of items on both shelves with the two targets, and, over time, become more efficient regardless of the current target (shampoo or the body lotion).

Another critical difference is that the target colors in the current work were randomly assigned on a trial-by-trial basis, while Sakata and colleagues fixed the target-distractor colors

for a given observer. As was discussed earlier, fixing the color of the to-be-ignored subsets could potentially strengthen the perceptual filtering (Gaspelin et al., 2015, 2017; Vatterott & Vecera, 2012). Thus, in the two-target task of Sakata et al. (2021), co-actors may have filtered out the irrelevant color subset in order to boost the performance. In other words, it is possible that co-actors in that work might have attended to the same display in parallel, but had little to no joint attention (Carpenter & Call, 2013). As the authors put it "... just knowing whether the co-actor responded or not is insufficient to facilitate the visuospatial learning …" (Sakata et al., 2021, page 10). By contrast, random swapping of colors in the current work may have somewhat limited the color-based filtering in the joint task condition. Taken together, it is possible that making both colors task-relevant in an unpredictable manner could have weakened the associative blocking, but only and specifically in combination with the social context of the joint task.

To summarize, the current work is an inspiring demonstration of the potential influence of a joint task in context learning. We provide important evidence that joint task performance forms a common contextual representation of the environment, where even task-irrelevant context can be acquired via co-active joint attention.

Acknowledgements

References

Annac, E., Pointner, M., Khader, P. H., Müller, H. J., Zang, X., & Geyer, T. (2019). Recognition of incidentally learned visual search arrays is supported by fixational eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(12), 2147–2164. https://doi.org/10.1037/xlm0000702

Böckler, A., Knoblich, G., & Sebanz, N. (2012). Effects of a coactor's focus of attention on task performance. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1404. https://doi.org/10.1037/a0027523

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*(4043), 77–80. https://doi.org/10.1126/science.177.4043.77

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Carpenter, M., & Call, J. (2013). How joint is joint attention of apes and human infants? In J. Metcalfe & H. S. Terrace (Eds.), *Agency and joint attention* (pp. 49–61). New York, NY: Oxford University Press. https://dx.doi.org/10.1093/acprof:oso/9780199988341.003.0003

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

Conci, M., & Müller, H. J. (2014). Global scene layout modulates contextual learning in change detection. *Frontiers in Psychology, 5*, 89. https://doi.org/10.3389/fpsyg.2014.00089

Conci, M., & von Mühlenen, A. (2011). Limitations of perceptual segmentation on contextual cueing in visual search. *Visual Cognition, 19*(2), 203–233. https://doi.org/10.1080/13506285.2010.518574

Davenport, J. L., & Potter, M. C. (2004). Scene Consistency in Object and Background Perception. *Psychological Science, 15*(8), 559–564. https://doi.org/10.1111/j.0956-7976.2004.00719.x

Endo, N., & Takeda, Y. (2004). Selective learning of spatial configuration and object identity in visual search. *Perception & Psychophysics, 66*(2), 293–302. https://doi.org/10.3758/BF03194880

Eskenazi, T., Doerrfeld, A., Logan, G. D., Knoblich, G., & Sebanz, N. (2013). Your words are my words: Effects of acting together on encoding. *The Quarterly Journal of Experimental Psychology, 66*(5), 1026–1034. https://doi.org/10.1080/17470218.2012.725058

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2015). Direct Evidence for Active Suppression of Salient-but-Irrelevant Sensory Inputs. *Psychological Science, 26*(11), 1740–1750. https://doi.org/10.1177/0956797615597913

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2017). Suppression of overt attentional capture by salient-but-irrelevant color singletons. *Attention, Perception, & Psychophysics, 79*(1), 45–62. https://doi.org/10.3758/s13414-016-1209-1

Geringswald, F., Herbik, A., Hofmüller, W., Hoffmann, M. B., & Pollmann, S. (2015). Visual memory for objects following foveal vision loss. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(5), 1471. https://doi.org/10.1037/xlm0000112

Geyer, T., Seitz, W., Zinchenko, A., Müller, H. J., & Conci, M. (2021). Why Are Acquired Search-Guiding Context Memories Resistant to Updating? *Frontiers in Psychology, 12*, 564. https://doi.org/10.3389/fpsyg.2021.650245

Geyer, T., Shi, Z., & Müller, H. J. (2010). Contextual cueing in multiconjunction visual search is dependent on color-and-configuration-based intertrial contingencies. *Journal of Experimental Psychology: Human Perception and Performance, 36*(3), 515. https://doi.org/10.1037/a0017448

Jiang, Y. V., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology Section A, 54*(4), 1105–1124. https://doi.org/10.1080/713756001

Jiang, Y. V., & Leung, A. W. (2005). Implicit learning of ignored visual context. *Psychonomic Bulletin & Review, 12*(1), 100–106. https://doi.org/10.3758/BF03196353

Jiang, Y. V., Sisk, C. A., & Toh, Y. N. (2019). Implicit guidance of attention in contextual cueing: Neuropsychological and developmental evidence. *Neuroscience & Biobehavioral Reviews, 105*, 115–125. https://doi.org/10.1016/j.neubiorev.2019.07.002

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

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Cambell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296) New York, NY: Appleton-Century-Crofts.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annual Review of Psychology, 55*(1), 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Kroell, L. M., Schlagbauer, B., Zinchenko, A., Müller, H. J., & Geyer, T. (2019). Behavioural evidence for a single memory system in contextual cueing. *Visual Cognition, 27*(5-8), 551–562. https://doi.org/10.1080/13506285.2019.1648347

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. 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. https://doi.org/10.1037/0096-1523.20.5.1000

Nickel, A. E., Hopkins, L. S., Minor, G. N., & Hannula, D. E. (2020). Attention capture by episodic long-term memory. *Cognition, 201*, 104312. https://doi.org/10.1016/j.cognition.2020.104312

Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition, 3*(5), 519–526. https://doi.org/10.3758/BF03197524

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

Sakata, C., Ueda, Y., & Moriguchi, Y. (2021). Learning of spatial configurations of a co-actor's attended objects in joint visual search. *Acta Psychologica, 215*, 103274. https://doi.org/10.1016/j.actpsy.2021.103274

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

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences, 10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition, 88*(3), B11–B21. https://doi.org/10.1016/S0010-0277(03)00043-X

Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: corepresenting stimulus-response mappings. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1234. https://doi.org/10.1037/0096-1523.31.6.1234

Sisk, C. A., Remington, R. W., & Jiang, Y. V. (2019). Mechanisms of contextual cueing: A tutorial review. *Attention, Perception, & Psychophysics, 81*(8), 2571–2589. https://doi.org/10.3758/s13414-019-01832-2

Summerfield, J. J., Rao, A., Garside, N., & Nobre, A. C. (2011). Biasing Perception by Spatial Long-Term Memory. *The Journal of Neuroscience, 31*(42), 14952. https://doi.org/10.1523/JNEUROSCI.5541-10.2011

Szymanski, C., Pesquita, A., Brennan, A. A., Perdikis, D., Enns, J. T., Brick, T. R., . . . Lindenberger, U. (2017). Teams on the same wavelength perform better: Inter-brain phase synchronization constitutes a neural substrate for social facilitation. *NeuroImage, 152*, 425-436. https://doi.org/10.1016/j.neuroimage.2017.03.013

Töllner, T., Conci, M., & Müller, H. J. (2015). Predictive distractor context facilitates attentional selection of high, but not intermediate and low, salience targets. *Human Brain Mapping, 36*(3), 935–944. https://doi.org/10.1002/hbm.22677

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2016). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review, 23*(1), 87–102. https://doi.org/10.3758/s13423-015-0892-6

Vaskevich, A., & Luria, R. (2018). Adding statistical regularity results in a global slowdown in visual search. *Cognition, 174*, 19–27. https://doi.org/10.1016/j.cognition.2018.01.010

Vatterott, D. B., & Vecera, S. P. (2012). Experience-dependent attentional tuning of distractor rejection. *Psychonomic Bulletin & Review, 19*(5), 871–878. https://doi.org/10.3758/s13423-012-0280-4

Verga, L., & Kotz, S. A. (2017). Help me if I can't: Social interaction effects in adult contextual word learning. *Cognition, 168*, 76–90. https://doi.org/10.1016/j.cognition.2017.06.018

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., . . . Nijssen, S. R. (2017). Joint action: mental representations, shared information and general mechanisms for coordinating with others. *Frontiers in Psychology, 7*, 2039. https://doi.org/10.3389/fpsyg.2016.02039

Zang, X., Assumpção, L., Wu, J., Xie, X., & Zinchenko, A. (2021). Task-Irrelevant Context Learned Under Rapid Display Presentation: Selective Attention in Associative Blocking. *Frontiers in Psychology, 12*, 675848–675848. https://doi.org/10.3389/fpsyg.2021.675848

Zang, X., Geyer, T., Assumpção, L., Müller, H. J., & Shi, Z. (2016). From foreground to background: How task-neutral context influences contextual cueing of visual search. *Frontiers in Psychology, 7*, 852. https://doi.org/10.3389/fpsyg.2016.00852

Zang, X., Jia, L., Müller, H. J., & Shi, Z. (2015). Invariant spatial context is learned but not retrieved in gaze-contingent tunnel-view search. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 807. https://doi.org/10.1037/xlm0000060

Zang, X., Shi, Z., Müller, H. J., & Conci, M. (2017). Contextual cueing in 3D visual search depends on representations in planar-, not depth-defined space. *Journal of Vision, 17*(5), 17–17. https://doi.org/10.1167/17.5.17

Zang, X., Zinchenko, A., Jia, L., Assumpção, L., & Li, H. (2018). Global Repetition Influences Contextual Cueing. *Frontiers in Psychology, 9*, 402. https://doi.org/10.3389/fpsyg.2018.00402

Zellin, M., Conci, M., von Mühlenen, A., & Müller, H. J. (2011). Two (or three) is one too many: testing the flexibility of contextual cueing with multiple target locations. *Attention, Perception, & Psychophysics, 73*(7), 2065. https://doi.org/10.3758/s13414-011-0175-x

Zellin, M., von Mühlenen, A., Müller, H. J., & Conci, M. (2014). Long-term adaptation to change in implicit contextual learning. *Psychonomic Bulletin & Review, 21*(4), 1073–1079. https://doi.org/10.3758/s13423-013-0568-z

Zinchenko, A., Conci, M., Hauser, J., Müller, H. J., & Geyer, T. (2020a). Distributed attention beats the down-side of statistical context learning in visual search. *Journal of Vision, 20*(7), 4–4. https://doi.org/10.1167/jov.20.7.4

Zinchenko, A., Conci, M., Müller, H. J., & Geyer, T. (2018). Predictive visual search: Role of environmental regularities in the learning of context cues. *Attention, Perception, & Psychophysics, 80*(5), 1096–1109. https://doi.org/10.3758/s13414-018-1500-4

Zinchenko, A., Conci, M., Töllner, T., Müller, H. J., & Geyer, T. (2020b). Automatic Guidance (and Misguidance) of Visuospatial Attention by Acquired Scene Memory: Evidence From an N1pc Polarity Reversal. *Psychological Science, 31*(12), 1531–1543. https://doi.org/10.1177/0956797620954815