Spatial predictive context speeds up visual search by biasing local attentional competition

Spatial predictive context biases local competition

Floortje G. Bouwkamp, Floris P. de Lange, Eelke Spaak

Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, 6525 EN, the Netherlands

Corresponding author: floortje.bouwkamp@donders.ru.nl

The authors declare no competing financial interests
We thank Jorie van Haren for his help with collecting the data

Abstract

The human visual system is equipped to rapidly and implicitly learn and exploit the statistical regularities in our environment. Within visual search, contextual cueing demonstrates how implicit knowledge of scenes can improve search performance. This is commonly interpreted as spatial context in the scenes becoming predictive of the target location, which leads to a more efficient guidance of attention during search. However, what drives this enhanced guidance is unknown. First, it is unclear whether improved attentional guidance is enabled by target enhancement or distractor suppression. Second, it is unknown whether the entire scene (global context) or more local drives this phenomenon. In the present MEG experiment, we leveraged Rapid Invisible Frequency Tagging (RIFT) to answer these two outstanding questions. We found that the improved performance when searching implicitly familiar scenes was accompanied by a stronger neural representation of the target stimulus, at the cost specifically of those distractors directly surrounding the target. Crucially, this biasing of local attentional competition was behaviorally relevant when searching familiar scenes, indicating that it is the local, and not global, spatial context that is modulated, culminating in a search advantage for familiar scenes. Taken together, we conclude that implicitly learned spatial predictive context improves how we search our environment by sharpening the attentional field.

Introduction

Locating a relevant item in a crowded visual field, also known as *visual search*, is a critical skill guiding our behavior in daily life. Prior experience yields a vast amount of knowledge on how items tend to co-occur in scenes. Knowledge of these relations can narrow down the search space by generating expectations about *what* to expect, and *where* to expect it (Võ et al., 2019; Peelen et al., 2023). The implicit and automatic learning of regularities is called *statistical learning*. This is a very useful tool in investigating how we learn from our environment, and exploit that knowledge in our interaction with the world.

Within the domain of visual search, a classic example of statistical learning is *contextual cueing*. Chun & Jiang (1998) discovered that participants are markedly faster in finding a target in search scenes that are repeated, compared to search scenes that are novel. After (implicit) learning, the predictive relationship between spatial context and target location is exploited, leading to more efficient guidance of spatial attention when searching these scenes (Goujon et al., 2015; Jiang et al., 2019; Sisk et al., 2019). This improved attentional guidance can be interpreted as the result of an optimized attentional priority map (Sisk et al., 2019; Wolfe, 2021), but how exactly these learned target-distractor relations alter what we prioritize, isan open question.

Firstly, more efficient guidance of selective attention can be achieved by either enhancing the target, or suppressing distractors, and there is behavioral evidence implicating both (Ogawa et al., 2007; Makovski and Jiang, 2010). It remains unclear, however, whether these effects are the consequence of the target getting relatively higher attentional priority, or whether distractors are suppressed. This is a relevant distinction, as target enhancement and distractor suppression are thought to rely on distinct (neural) mechanisms (Noonan et al., 2016; Slagter et al., 2016; Wöstmann et al., 2022).

Secondly, there is debate on the extent of context involved in contextual cueing. Some have argued that only *local* context directly surrounding the target is learned, since repeating only such local context produced a similar effect to classical contextual cueing (Brady and Chun, 2007). However, this does not hold when presentation time is limited (Xie et al., 2020), and there has been evidence that distractor context *unrelated* to the target is also learned (Beesley et al., 2015). The question thus remains whether local and/or global context is learned in contextual cueing (Goujon et al., 2015).

We set out to investigate how spatial predictive context modulates attentional guidance during visual search at the stimulus level: does contextual learning involve attentional enhancement of the target and/or a suppression of distractors? And does contextual cueing mainly involve local or global

context? For this, we leveraged Rapid Invisible Frequency Tagging (RIFT) (Zhigalov et al., 2019; Drijvers et al., 2021; Minarik et al., 2022; Seijdel et al., 2023). The distinct advantage of *rapid* tagging is that it is invisible to the human eye, and therefore does not interfere with perceptual processes, while signal-to-noise ratio in neural (EEG/MEG) recordings is still very high (Minarik et al., 2022). Importantly, the RIFT signal is modulated by selective attention (Zhigalov et al., 2019). This enables us to measure both stimulus processing, and the attentional modulation thereof, with high temporal and stimulus precision.

In brief, we found improved visual search of repeated compared to novel scenes. Neurally, we found target enhancement and suppression of distractors near the target, effectively biasing local attentional competition. Specifically this biasing of local attentional competition improved behavioral performance when searching repeated scenes. We conclude that spatial predictive context enhances attentional guidance by sharpening the attentional field.

Methods and materials

Data and code availability. All data and code used for stimulus presentation and analysis are freely available on the Donders Repository. The link to the repository will be accessible to reviewers and be publicly available upon publication.

Participants. Thirty-six adults (23 identifying as female and 13 identifying as male, age 18-49 years, M = 28.95, SD = 8.72) participated in the experiment. All were recruited via the Radboud University participant database and informed consent was given before the start of the experiment. The study was approved by the local ethics committee (CMO 2014/288; CMO Arnhem-Nijmegen, the Netherlands) and was conducted in compliance with these guidelines. None reported a history of neurophysiological disorders and all had normal or corrected-to-normal vision. Four participants were excluded due to insufficient data quality (see MEG analyses). The final sample therefore consisted of 32 participants.

Stimulus material. Search scenes consisted of scenes with 8 stimuli: one target letter T and seven distractor L shapes, with a small (10%) offset in the line junction to increase search difficulty (Jiang and Chun, 2001). Stimuli measured 2.5° x 2.5° and were displayed as mid grey on a dark grey background. Distractor shapes were rotated randomly with a multiple of 90°. The target was tilted either to the left or to the right. Each scene had a fixation dot at the center (outer white diameter 8 pixels, inner white diameter 4 pixels). Stimuli were placed on a 7x5 grid spanning the screen from -11° to 11° horizontally and -7.5° to 7.5° vertically, excluding the center position. To prevent collinearity the stimuli were jittered with ± 0.5°(Chun and Jiang, 1998b). To ensure search scenes were approximately equal in difficulty, we generated search scenes abiding multiple restrictions. The target was always placed between 6° and 10° of eccentricity, the mean distance between target and distractors was kept between 9° and 11°, no more than one quadrant was empty, and in quadrants with >2 stimuli, local crowding was prevented. Additionally, target locations across trials was evenly distributed over the four quadrants. Three types of stimuli within these scenes were tagged with unique frequencies (see Frequency tagging): the target, a distractor near (D*near*) the target and a distractor further (D*far*) from the target. The far standing distractor was always placed at the opposite position of the target relative to the center of the screen. The distractor closest to the target was selected as Dnear, with a minimum distance of *less* than 4° from the target, while being minimally 4° from the center of the screen.

Experimental conditions. Each block consisted of 32 trials and half of the trials each block were newly created (*New)*, while the other half of trials were the same 16 search scenes, rendering them

Old from block 2 onwards (figure 1B). In these Old displays, both (jittered) location and orientation of distractors were repeated, as well as target location. Target orientation was always randomized (tilted left or right) to prevent any learning of the correct response in Old trials. Practice trials consisted of a separate set of scenes without any repetitions.

Frequency tagging. To achieve Rapid Invisible Frequency Tagging, we used a PROPixx DLP LED projector (VPixx Technologies Inc., Daint-Bruno-de-Montarville, Canada). This monitor can interpret four quadrants of the screen in three color channels of the GPU screen buffer, as twelve separate smaller grayscale frames, which it then projects in rapid succession. This leads to a twelvefold increase of the 120 Hz presentation rate, yielding 1440 Hz. Three stimuli (figure 1C) were tagged with either 60, 64 or 68 Hz, and stimulus-frequency mapping was counterbalanced across subjects. Color values of the stimuli oscillated between 0 (black) and 255 (white), at the assigned frequency by multiplying the max luminance value with a sinusoid of the relevant frequency. Other stimuli were kept constant at the mean value of 127. All stimuli therefore appeared as mid grey on the screen. To prevent straining the eyes, the background was kept at dark grey (value = 43). No frequency tagging was applied in the final two blocks, and all stimuli were kept constant at the mean value of 127.

Experimental procedure. After verbal instructions and signing informed consent, participants were seated in the dimly-lit magnetically shielded MEG room wearing a neck brace for head stabilization. Stimuli were presented on a screen at approximately 85 cm distance, using Matlab 2018b (Mathworks Inc, Natrick, USA) and the Psychophysics Toolbox, (Brainard, 1997; Pelli, 1997). The experiment started with a 9 point calibration session for the eye tracker, which was used to monitor fixation of the participant throughout the experiment. Participants were introduced to the task and instructed to report the orientation of the Target (either tilted left or right) with a left or right button press.

Furthermore, it was stressed that participants had to remain their fixation at the center of the screen. Each trial (figure 1A) started with a fixation period of one second, followed by the search scene until the subject pressed a button to indicate their response. Maximum response time was 2.5 seconds. At the end of each trial they received feedback (0.5 seconds) on performance with a symbol: V for correct, X for incorrect and O for being too late. Participants practiced this task for 64 trials. If either performance or fixation was insufficient, they practiced for another 64 trials. Then, the main experiment started, consisting of 26 blocks offered in sets of two blocks of 32 trials each. In between these 'subject blocks' of 64 trials, participants received feedback on performance (% correct and average response time in seconds) and were offered a short break. When ready for the next block, their head position was realigned if necessary before starting. At the end of the main

experiment participants performed a short recognition task where they had to judge for 32 search scenes (16 Old, 16 New) whether this scene was familiar or not. Responses in the recognition block were self-paced (no timeouts). The entire experiment lasted approximately 1.5 hours.

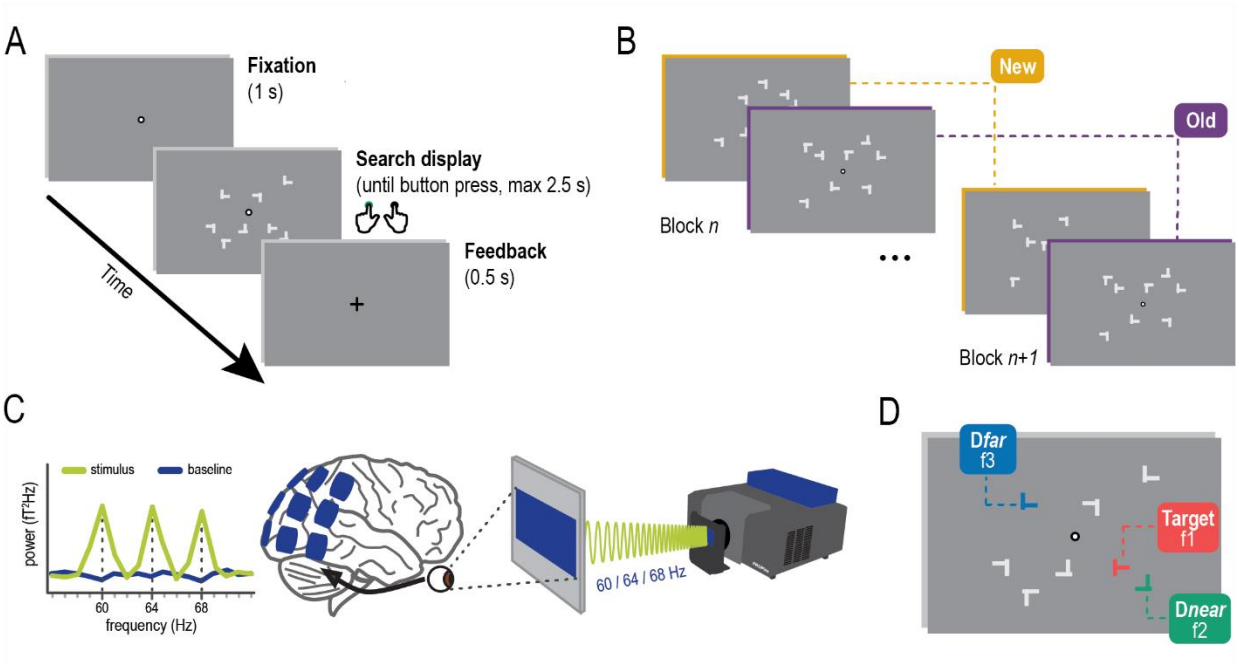


Figure 1. Paradigm. A) Task. Participants engaged in a visual search task where they had to locate a target letter T embedded amongst distractor L-shapes and report its orientation with a left or right button press. B) Manipulation. Half of all trials consisted of search scenes that were repeated every block, rendering them 'Old'. The other half of the trials consisted of unique and newly created search scenes. C) RIFT. The projector allows rapid tagging of stimuli at our frequencies of interest. This tagging signal can be identified in the power spectrum as sharp peaks at the frequencies used (image adapted from Seijdel et al 2023) D) Tagging search scenes. Within each search scene three types of stimuli were tagged at a unique frequency: the target, a near-standing distractor and a far-standing distractor. Assignment of frequencies to stimuli was counterbalanced across participants. Colors of the different stimulus types are for illustration purposes only.

Behavioral analyses. Analyses and visualization of behavioral data was done using R (R Core Team, n.d.) and packages ggplot2 (Wickham, 2016), raincloud plots (Allen et al., 2021), Effsize (Torchiano, Marco, 2016) and BayesFactor (Morey and Rouder, 2022) for R. Reaction time was our primary, and accuracy our secondary variable of interest. Only trials were a response was given in time were included in the analyses (only 1 subject was too late on 2 trials), trials with an response given before 100 ms were regarded as accidental presses and excluded (3.85% of trials). Reaction time analyses were performed on correct trials only. We improved normality of the reaction times by taking the log10 of the raw values. Plots and statistical tests are based on these log transformed reaction

times, however we report raw reaction times in the result section. Plots over experiment time of both reaction times and accuracy are smoothed across neighboring blocks. Statistical assessment was done on the latter part of the experiment (blocks 9-26, unsmoothed, selected a priori) when learning is assumed to have taken place. We directly contrasted conditions by means of a paired sample t-test. We report Cohen's d for effect size and additionally quantified the relative evidence for the alternative against the null hypothesis using Bayes factors. Accuracy on the recognition task at the end of the experiment was pitted against the null hypothesis of random guessing (50% accuracy) with a one sample t-test.

MEG acquisition. During the main experiment MEG activity was recorded using a 275-channel axial gradiometer CTF MEG system (CTF MEG systems, Coquitlam, Canada) situated in a magnetically shielded room (MSR). An online 300hz low pass filter was used and data was digitized with a sampling rate of 1200 Hz. The head position of participants was monitored in real time using markers on the nasion, and left and right periauricular points (Stolk et al., 2013). Participants were asked to readjust their head position when the deviation from the original starting point was larger than 5 mm. Additionally participants' gaze was monitored using an SR Research Eyelink 1000 eye tracker and responses were collected using a MEG-compatible button box.

MEG preprocessing. MEG data was preprocessed and analyzed using the Fieldtrip toolbox (Oostenveld et al., 2011) in a Matlab environment (Mathworks, version 2018b). The data was segmented into trials and preprocessed by first applying third order gradient denoising and general demeaning of the data. Subsequently, trials affected by high noise or muscle movement were removed using a semi-automated method by visually identifying unique high variance trials. Then, the data was downsampled to 400 Hz and cleaned of any residual artifacts due to eye movements, heartbeat or other sources of noise using Independent Component Analysis (Bell and Sejnowski, 1995; Jung et al., 2000). During recording, the MEG system was affected by an occasional drifting high-frequency artifact caused by external RF-noise entering the MSR. This artifact was detected in six of the 36 acquired datasets, and in three of these it drifted through our frequency range of interest during a subset of trials. All six datasets were successfully cleaned in the preprocessing stage with an in house procedure (Principal Component Analysis applied with the artifact as a reference signal). Together these preprocessing procedures lead to removing on average 6.48% of trials.

MEG analyses. For the assessment of general tagging we used all tagged trials, independent of condition. We discarded trials that were shorter than 800 ms to ensure a wider time window to analyze. Data was cut to an a priori defined time window of interest (0.2 – 0.8 s post stimulus onset)

and a baseline window (-0.5 – 0 s pre stimulus onset). As tagged stimuli varied in location per search scene, we applied trial-based spectral analyses. Power was estimated using a boxcar window focused on the narrow band responses at our three tagging frequencies of interest. Data was zero-padded to 1 second in order to obtain a frequency spacing of 1 Hz. We additionally approximated planar gradiometer data by converting our axial gradiometer data to orthogonal planar gradiometer pairs, and computing power separately before combining them by averaging. This enables a more straightforward interpretation of the MEG data, as planar gradient maxima are thought to be centrally located above their neuronal source (Bastiaansen and Knösche, 2000). Power during the window of interest was expressed as decibels relative to the baseline period. Based on this data, we selected the twenty sensors within occipital-parietal region that were most responsive to average power at the tagged frequencies per individual subject (results section, figure 3B). We then inspected the condition-averaged power spectrum (frequency-resolved from 52 to 76 Hz) at the selected sensors per subject. This led to the exclusion of 4 subjects that had very noisy power spectra, with no clear peaks at the frequencies of interest and/or additional peaks at neighboring frequencies. This yielded our final sample size of 32 subjects. For our main analyses we subsequently selected only correct (target orientation task) trials from the latter part of the experiment (a priori decision, latter 2/3 of all tagged blocks), and considered the initial 8 blocks learning blocks (Chun and Jiang, 1998b). We separated the trials into our Old and New condition. Because trials varied in length we analyzed the data both aligned to stimulus onset (stimulus-locked) and aligned to the response (response-locked). We subjected the data to spectral analyses, both averaged over a priori defined time windows of interest (0.2 - 0.8 s post-stimulus or –0.6 - 0 pre-response) following the procedure described above, and in a time-resolved manner (–0.5 - 0.8 s stimulus-locked or –0.8 – 0.5 response-locked). Time-resolved power was estimated using a 250 ms sliding window (min. 15 cycles) in steps of 50 ms, after zero-padding the data to 2 seconds. Baseline normalization was done within condition. Statistical analyses of the power values per stimulus type was done by contrasting the target to both distractor types, with a one-tailed paired sample t-test, testing for the increase in power of target over distractors. Additionally we tested the difference between the two distractor types with a two-tailed paired sample t-test. Statistical analyses on the time-resolved power was done with non-parametric cluster based permutation tests (Maris and Oostenveld, 2007). For this, only the time windows post-stimulus and pre-response, taking temporal leakage into account, were considered (0.15 - 0.6 s stimulus-locked or –0.6 – 0.15 response-locked). Clustering threshold was set to 0.1, the number of randomizations was set to 10,000, and the test statistic was based on a paired sample t-test. Again we contrasted target to both distractors using a one-tailed test, while contrasting the distractors using a two-tailed test. For the analyses of

behavioral relevance we used the trial-based stimulus-locked power of Old and New trials and correlated the difference between target and each distractor to the behavior (log transformed reaction times) on these trials within subject. Behavioral relevance was assessed not only across all trials, but also specifically across the 16 displays within the Old condition. For these analyses, both power and reaction times on Old trials were averaged across trials per display (16 in total). We then, for both analyses, computed the Pearson correlation values across displays, per subject, and tested the distribution of these correlations over subjects against null with a one sample t-test.

Results

Participants were seated in the MEG while engaged in a visual search task. Typically for contextual cueing, half of all trials were search scenes randomly generated per trial, labeled as 'New'. The other half were search scenes that were repeated every block, which we labeled as 'Old'. To assess the mechanisms underlying enhanced attentional guidance by spatial predictive context in Old trials compared to New trials, we tagged the Target and two distractors in each search display using Rapid Invisible Frequency Tagging (RIFT). By tagging both a distractor *near* the target (D*near*) and a distractor *far* from the Target (D*far*), we were able to investigate the roles of local and global context in contextual cueing.

Contextual cueing: improved search performance on Old trials compared to New trials

For all analyses, both behavioral and MEG, we focused on the latter two thirds of the experiment (blocks 9 and onward, selected a priori), after participants putatively learned the predictive spatial context within Old trials.

Accuracy. Overall accuracy on the task was high (Mean 85.90% ± 6.08% SD; range: 68.78% - 95.87%). There was no difference in accuracy between Old and New scenes in the latter part of the experiment (New: 87.22% ± 6.03% and Old: 87.57% ± 5.88%, difference 0.35% ± 3.47% , $t(31) = -.57$, $p = .57$, $d = .06$, $BF_{10} = .22$; Supplementary Figure 1), allowing us to focus solely on reaction times.

Reaction times. As the distribution of reaction times is typically non-normal, all analyses are based on log transformed data. For interpretability, we will report raw RT values in the text. Overall search speed was well within the time limit of 2.5 seconds (1.21 s ± 0.43 s. Focusing on the latter part of the experiment, participants were faster in finding the target on Old trials compared to New scenes (New: 1.19 s ± 0.41 s and Old: 1.14 s ± 0.41 s, difference: 50 ms ± 60 ms, $t(31) = 4.87$, $p = 3.14 \times 10^{-6}$, $d = .33$, $BF_{10} = 733$). We thereby replicated the classic contextual cueing effect (Figure 2).

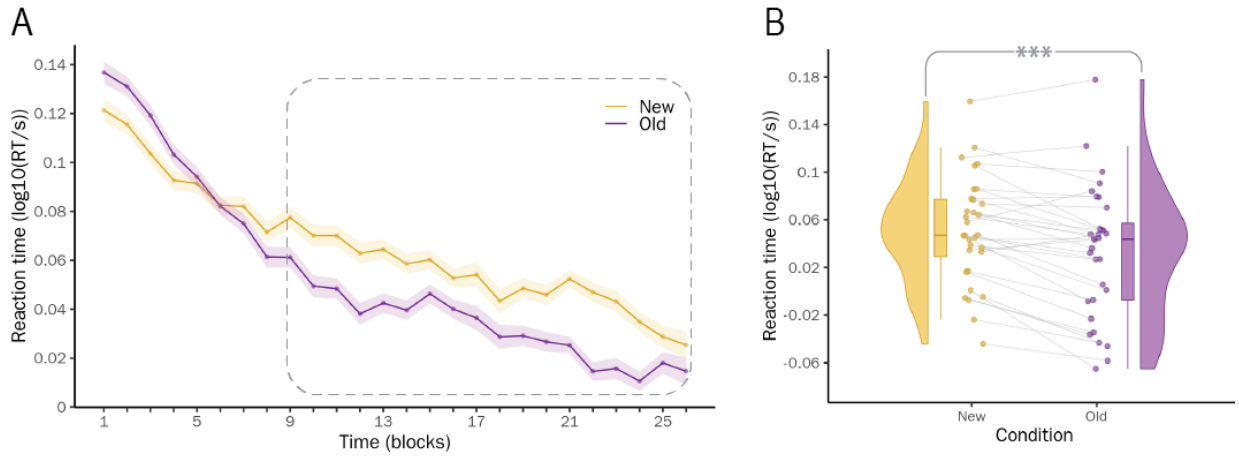


Figure 2. Reaction times. (A) Reaction time (smoothed across neighbors, by taking the mean of block N, N − 1 and N + 1) plotted over the time course of the experiment (shading indicates within-participant corrected standard error of the mean). Dotted box indicates the trials used contrasting Old to New scenes, where contextual learning is assumed to have taken place. (B) Reaction times on New versus Old trials on learned trials. Dots are individual participants. The boxplot indicates the interquartile range (IQR, the box), the median (bar on box) and the minimum and maximum value within ±1.5 x IQR (whiskers).

Identifiable power peaks at all tagged frequencies

As a first step, we calculated power at our frequencies of interest pooled over conditions and participants. When normalizing power with a baseline period and plotting the spectra (Figure 3A), we observed clear peaks at the tagged frequencies of interest, with no apparent differences in power between the three frequencies. Additionally, we observed no such peaks for trials where the search scenes were not tagged (Fig 3A, blue line). This steady state visual evoked response was, as expected, localized at posterior sensors, covering primary visual cortex (Figure 3B). We can therefore conclude that RIFT can be used to reliably tag several relatively small (~2.5 degrees) non-foveated stimuli simultaneously on the screen, at variable locations in retinotopic visual space.

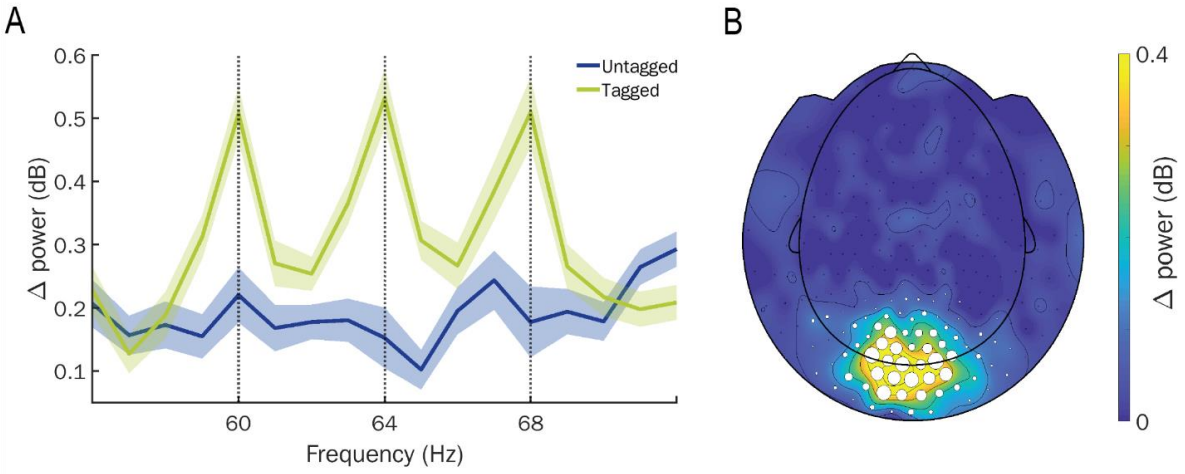


Figure 3. Power spectrum and topography of RIFT. A) The power spectrum shows clear peaks at the frequencies of interest, which are absent on trials of the last two blocks when the search scenes were not tagged. The topography of (planar) power at the frequencies of interest indicates that tagging power is mostly seen in the posterior occipital region. White circles indicate sensors that have been selected for the analyses, with the size indicating for how many participants a sensor was used.

Stronger power for target compared to distractors when searching Old scenes

For our main analyses, we calculated power at the tagged frequencies separately for trials where participants were searching Old and New scenes. Since visual search trials vary in length, we analyzed these trials both locked to stimulus onset and locked to the response. We analyzed tagging power values both averaged over a priori defined time windows of interest (0.2 to 0.8 s post-stimulus, and −0.8 to −0.2 s pre-response), as well as in a time-resolved manner using cluster-based permutation tests. As we are specifically interested in attentional competition between the tagged stimuli during visual search, we contrasted target power with power for both types of distractors (D*far* and D*near*) and additionally tested the difference between local and global context (D*near* versus D*far*).

First, we looked at regular visual search trials without spatial predictive context (New trials). We found no statistically significant differences for any of our comparisons in New trials (Figure 4A): neither between Target and distractors (Target > D*near* stimulus-locked t(31) = −1.12, p = .135, d = 0.12, BF_{10} = .34 and response-locked t(31) = −0.82, p =.210, d = .08, BF_{10} = .26; Target > D*far* stimulus-locked t(31) = −0.14, p = .444, d = .19, BF_{10} = .19 and response-locked t(31) = −0.09, p = .463, d = .02, BF_{10} = .19) nor between distractors (D*near* <> D*far* stimulus-locked t(31) = −1.15, p = .259, d = .11, BF_{10} = .35 and response-locked t(31) = −0.84, p = .406, d = .08, BF_{10} = .26). Time-resolved results also did not reveal any significant cluster-based differences between the stimulus types during New trials (Figure 4B), neither stimulus-locked nor response-locked.

Next, we looked at visual search on Old trials, with established spatial predictive context. A priori time window analysis (Figure 4C) revealed stronger power at the Target compared to distractors. Compared to D*near* this stronger power of the Target was evident in the data locked to the response ($t(31) = -2.18$, $p = .019$, $d = .25$, $BF_{10} = 1.48$), and marginally in the data locked to stimulus onset ($t(31) = -1.66$, $p = .053$, $d = .16$, $BF_{10} = .65$). Compared to D*far,* stronger power at the target was only present in the response-locked data ($t(31) = -1.70$, $p = .049$, $d = .19$, $BF_{10} = .68$), but not in stimulus-locked data ($t(31) = -.82$, $p = .209$, $d = .09$, $BF_{10} = .26$). There was no difference between the distractor types (stimulus-locked: ($t(31) = -1.10$, $p = .278$, $d = .14$, $BF_{10} = .33$, response-locked: $t(31) = -0.75$, $p = .461$, $d = 0.09$, $BF_{10} = .24$).

Time-resolved analysis of Old trials (Figure 4D) further identified a positive cluster, especially comparing power at the target to power at the near standing distractor (D*near*; cluster-based permutation test $p = .018$). This cluster was identified in the early part of the response-locked window, from 550 to 250 ms before the response. There were no significant cluster-based differences between the target and the far-standing distractor, nor when contrasting the two distractor types in the response-locked window. There were additionally no significant clusters identified between any of the stimulus types during the stimulus-locked window.

We thus found evidence that spatial predictive context within the Old trials enhances the Target compared to distractors. This biased competition appears strongest locally, between the target and the near-standing distractor specifically. However, we must note that there was no direct evidence for a difference in power between the two distractor types.

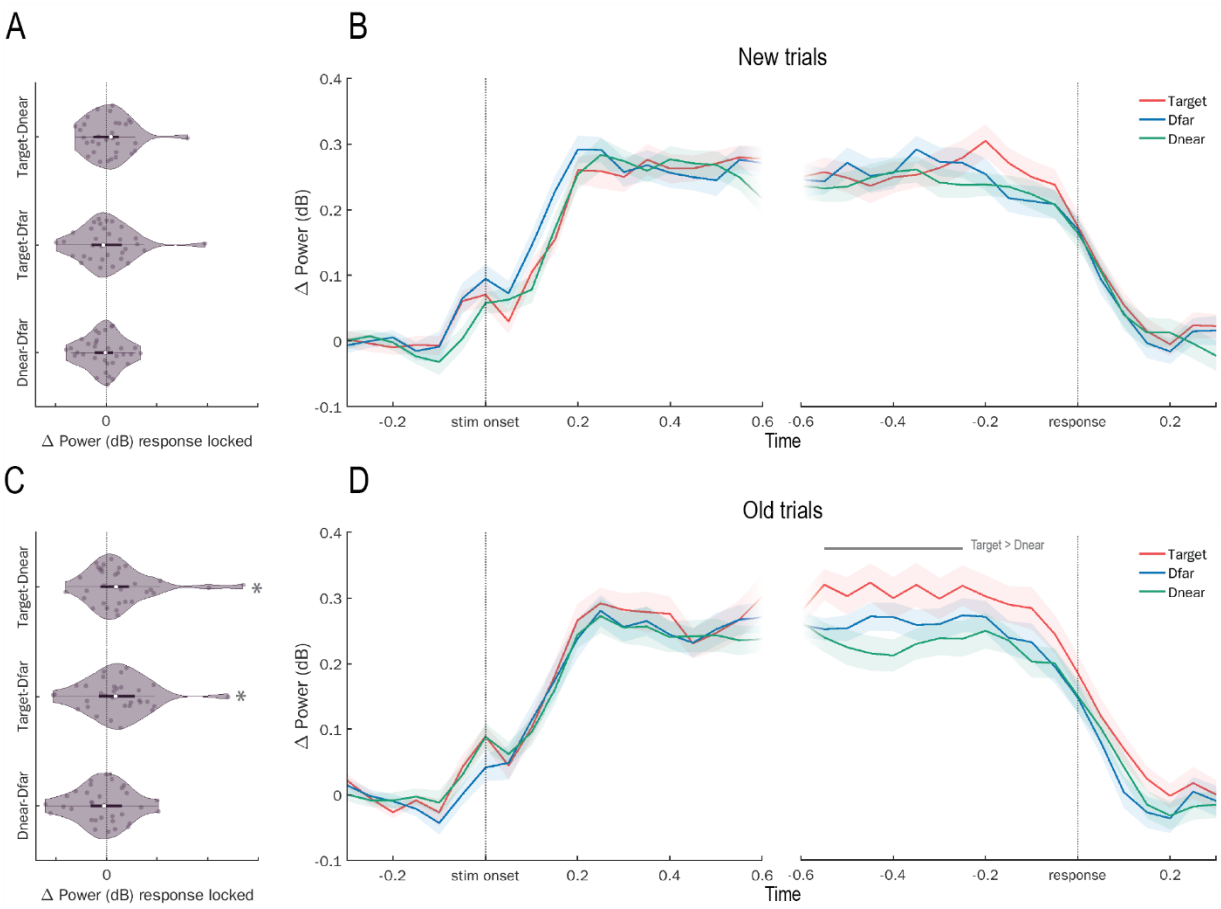


Figure 4. Power Old and New trials. (A) Response-locked power New trials averaged over a priori time window (−0.6 to −0.2 s before response), contrasting the stimuli types. Dots are individual participants, the white dot reflects the median, the dark bar covers the interquartile range (IQR) with the whiskers indicating the maximum and minimum value within ±1.5 IQR. (B) Time-resolved power at the tagged frequencies in New trials. Colors indicate stimulus type. Shading indicates within-participant corrected standard error of the mean.
(C) Response-locked power Old trials averaged over time, contrasting the stimuli types. Stars indicate significant differences. (D) Time-resolved power at the tagged frequencies in Old trials. Grey bar above plots indicates the significant cluster.

Target enhancement and distractor suppression in Old compared to New scenes bias local attentional competition

To see the direct impact of spatial predictive context on each stimulus type in isolation, we directly contrasted the power at each stimulus between Old and New scenes. We found evidence for higher target-related power in Old compared to New scenes, specifically when the data was aligned to the response (response-locked: t(31) = 2.32, p = .027, d = .15, BF_{10} = 1.91, stimulus-locked (t(31) = 1.06, p = .296, d = .09 , BF_{10} = .32), but we found no evidence for lower distractor-related power in Old compared to New scenes (response-locked: D*near*: t(31) = −0.39, p = .69, d = .04, BF_{10} = .20,

D*far*: t(31) = −0.13, *p* = .90, *d* = .01, BF_{10} = .19; stimulus-locked: D*near*: t(31) = 0.06, *p* = .96, *d* = .01, BF_{10} = .19, D*far*: t(31) = −0.14, *p* = .89, *d* = .02, BF_{10} = .19).

Next, we wanted to understand how the differences we find between Old and New trials, interacts with the attentional competition between stimuli during visual search. We therefore directly contrasted power on Old and New trials and subsequently compared the stimulus types (i.e., testing differences-of-differences, in essence testing the interaction between stimulus type and condition). We found a difference between Target and both distractors when the data was aligned to the response (Target−D*near* response-locked: t(31) = −1.74, *p* = .046, *d* = .50, BF_{10} = .73 and stimulus-locked: t(31) = −0.84, *p* = .204, *d* = .22, BF_{10} = .26; Target−D*far* response-locked: (t(31) = −1.98, *p* = .028, *d* = .43, BF_{10} = 1.06 and stimulus-locked: t(31) = 0.86, *p* = .197, *d* = .17, BF_{10} = .27), while there was no difference between the distractors (response-locked: t(31)= −0.15, *p* = .881, *d* = .04, BF_{10} = .19 and stimulus-locked: t(31)= −0.14, *p* = .887, *d* = .04, BF_{10} = .19).

When analyzing these differences in attentional competition between Old and New trials in a time-resolved manner (Figure 5B), we found a significant cluster only when contrasting Target to D*near* (*p* = .038) from 350 ms to 450 ms before the response, and not contrasting it to D*far*, nor when contrasting the two distractor types. There were again no significant differences during the stimulus-locked window. Visual inspection of figure 5A further supports these cluster based statistical results: target enhancement specifically co-occurs with distractor suppression of D*near* (*p* = .04). Given the apparent time course of this effect, it is likely that the analyses in the a priori defined time window (-0.8 to -0.2 s pre-response) were not optimally sensitive to detect it, while the cluster-based analysis did identify a significant difference.

Taking these results together, we conclude that the main difference between Old and New scenes is an enhancement of the target, and this target enhancement biases specifically local attentional competition between the target and directly surrounding distractors, which are subsequently suppressed.

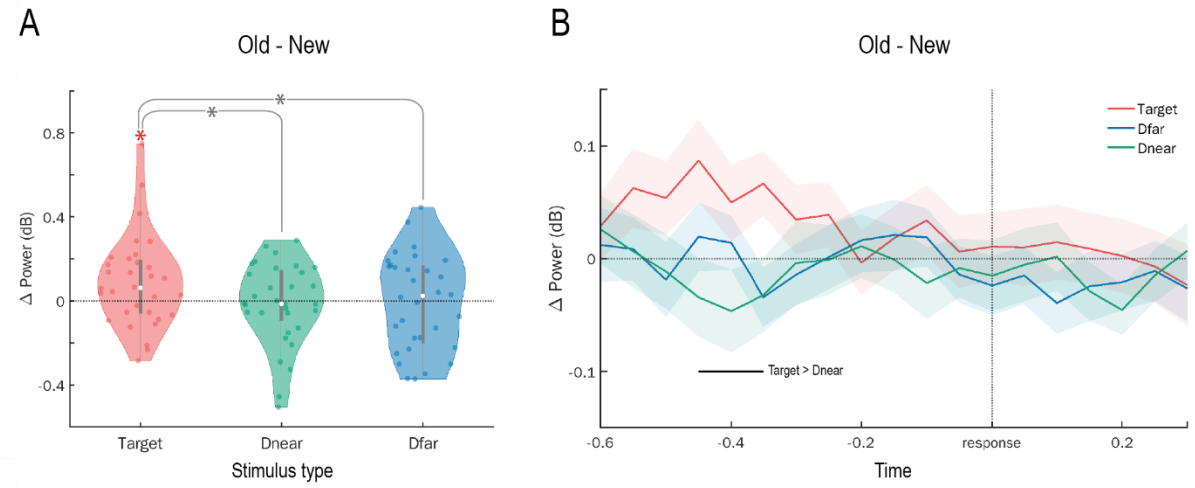


Figure 5. Response-locked power per stimulus type contrasting Old and New trials. (A) Response-locked power averaged across time, contrasting Old and New trials per stimulus type. Colored dots are individual participants, the white dot reflects the median, the dark bar covers the interquartile range (IQR) with the whiskers indicating the maximum and minimum value within ±1.5 IQR. Colored star indicates significant difference between Old and New trials. Grey stars indicate a significant difference of Old minus New power values between the stimuli types. (B) Time-resolved power. Shading indicates within-participant corrected standard error of the mean, the bar indicates the significant cluster.

Both local and global competition are behaviorally relevant, but only local context is learned

Subsequently, we investigated whether this competition between target and distractors during visual search is meaningful for behavior. To do so, we correlated, within each subject, differences in power between these stimulus types with reaction times, across trials. To remove the overall effect of condition, we analyzed the New and Old trials separately (see Figure 6). Since response-locked data does not yield trial-by-trial variance which can be related to variance in reaction times, we only analyze stimulus-locked power here.

On New trials, the difference between stimulus-locked Target power and power at D*far* was negatively correlated with reaction times, (mean $r = -.03$, $t(31) = -2.15$, $p = .039$, $d = .38$, $BF_{10} = 1.41$), while the difference between Target and D*near* was not (mean $r = -.005$, $t(31) = -0.44$, $p = .666$, $d = .08$, $BF_{10} = .21$). We found a similar pattern within Old trials, with again the difference between Target and D*far* being predictive of search speed (mean $r = -.04$, $t(31) = -2.68$, $p = .012$, $d = .47$, $BF_{10} = 3.88$), and not the difference between Target and D*near* (mean $r = -.01$, $t(31) = -0.67$, $p = .505$, $d = .12$, $BF_{10} = .23$). We thus found a subtle, yet reliable effect of *global* competition: The difference in power between target and distractors is meaningful for general visual search

performance, as participants were faster in finding the target when target power was stronger compared to power at far-standing distractors. This was true both for New and for Old trials.

We next sought to remove this general visual search effect, and isolate specifically what is being *learned* in contextual cueing. Old trials, by definition, consisted of repeating the same spatial configuration, or 'display'. This allowed us to remove general trial-by-trial variance and analyze the stimulus-locked power and reaction times averaged across Old trials per display (16 in total). We found that the power difference between Target and D*near* was correlated with behavioral performance (mean $r = -.11$, $t(31) = -2.48$, $p = .019$, $d = .44$, $BF_{10} = 2.58$), while this was not the case for the difference between Target power and D*far* (mean $r = -.08$, $t(31) = -1.37$, $p = 0.179$, , $d = .24$, $BF_{10} = .44$). Thus, when target power was high compared to power at the near-standing distractor specifically, participants were faster in finding the target in Old scenes.

From this we conclude that, when there is more power for the target compared to distractors people become faster in finding it. In general visual search, this holds specifically when target power is increased compared to distractors further away (presumably reflecting a general attentional drift effect, independent of learning). However, when we zoom in on the actual impact of learning within Old trials, we find that specifically competition between target and the near-standing distractor is impacting search performance. This finding of behavioral relevance of *local* attentional competition between target and near-standing distractors dovetails with our results on stimulus-specific power both within Old trials separately, and in the contrast between Old and New.

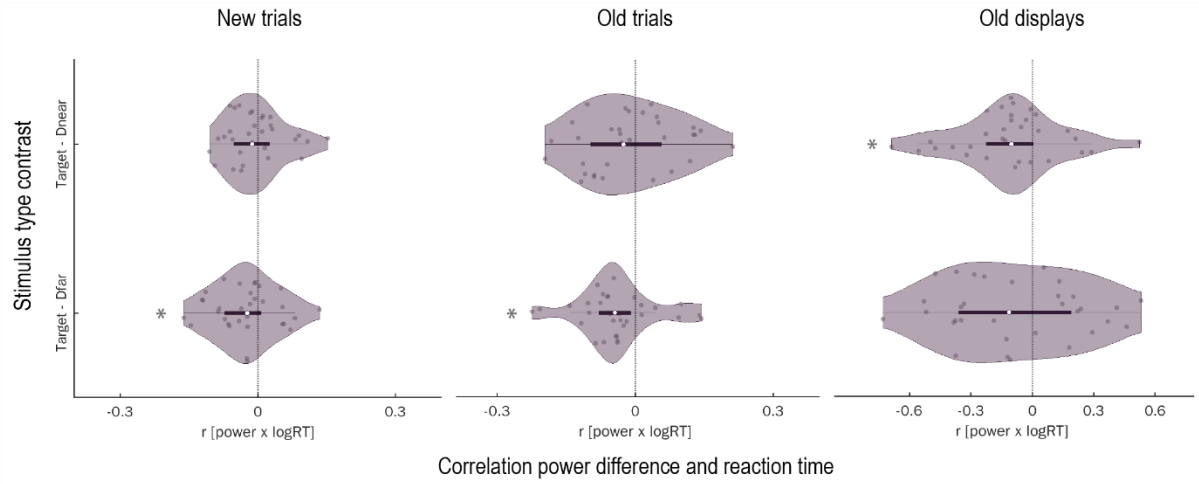


Figure 6. Behavioral relevance of stimulus-specific power. Distribution of r values expressing the relationship between stimulus-specific power and behavioral performance (logRT) per participant. Top plots contrast Target power to D*near* power, bottom plots contrast Target Power to D*far* power. From left to right the results for New trials, Old trials and Old trials averaged per display. Darker dots are individual participants, the white dot reflects the median and the grey bar indicates the interquartile range. Stars indicate that the r values are significantly different from zero.

No evidence of explicit knowledge of Old scenes

Recognition task. With a mean accuracy of 51.46% (±11.05%), participants were at chance level in correctly identifying Old and New displays in the recognition task (Supplementary Figure 2, t(31) = 0.75, *p* = .459, *d* = .13 BF_{10} = .25). Recognition accuracy was unrelated to whether participants were able to exploit contextual cueing behaviorally as we find no relationship between size of contextual effect and recognition accuracy (r=0.06, t(30) = 0.34, *p*= .734). We thus find no evidence indicating explicit knowledge of old scenes.

Discussion

Our study had two main objectives. Firstly, we aimed to investigate the neural mechanism underlying enhanced attentional processing of visual search when spatial context becomes predictive of target location. We utilized Rapid Invisible Frequency Tagging (RIFT) to track attentional processing of visual search scenes, and measure enhancement or suppression at the stimulus level. Secondly, by tagging two types of distractors, we probed the extent of the context involved in this improved attentional guidance. More specifically, we set out to test whether the entire context, including more distant distractors, or mainly the local context (i.e. distractors directly surrounding the target) play a role. Our findings revealed that visual search with predictive context (repeated scenes, referred to as "Old"), exhibited stronger Target-related power, especially compared to a distractor near that target. Moreover, contrasting predictive search to regular visual search without predictive context (referred to as "New") reveals that both Target enhancement and distractor suppression underly this biased local attentional competition between the Target and directly surrounding distractors. Crucially, we found that for visual search in general, the difference between target and the far standing distractor was reflected in behavioral performance. However, when investigating Old displays specifically, we observed that biasing local competition, between the target and the near standing distractor, is responsible for the accelerated search performance in contextual cueing. The implications of these findings are discussed below.

Our results indicate that the improved search performance in Old scenes is accompanied by stronger target power relative to distractors. This means that spatial predictive context creates a shift in shared attentional resources. More attention is allocated to the target, at the cost of the distractors, specifically those directly surrounding the Target. These results may explain several reported behavioral findings, such as contextual costs when the target in Old scenes is swapped with a distractor (Makovski and Jiang, 2010), and slower dot detection at learned distractor locations (Ogawa et al., 2007). Interestingly, our prediction would be that these impairments would only hold for local context and not for distractors further away, which has not been tested. Our findings additionally shed light on what underlies the enhanced N2pc found when there is spatial predictive context during visual search (Schankin and Schubö, 2009). The N2pc is an event-related potential (ERP) component believed to be related to attentional allocation. It relies on lateralized activity and therefore has limited granularity, and whether it reflects target selection, filtering out distractors, or both is under debate (Stoletniy et al., 2022). By leveraging RIFT, we were able to demonstrate that attentional allocation to the target within Old scenes is enhanced by biasing local competition in favor of the target, and at the cost of local distractors. Notably, Luck et al. (1997) demonstrated that

increasing local competition enhances the N2pc. This finding strongly supports our interpretation that spatial predictive context improves attentional guidance in visual search by biasing this local competition, which would indeed lead to an enhanced N2pc.

Moreover, the comparison between Old and New scenes revealed a simultaneous occurrence of target enhancement and distractor suppression. Previous research has implicated both in visual search, also in more natural settings (Hickey et al., 2009; Seidl et al., 2012). It has been argued that these processes rely on different mechanisms (Hickey et al., 2009; Noonan et al., 2016; Slagter et al., 2016; Wöstmann et al., 2022). However, distractor suppression has been mainly established using exogenous cues or obvious manipulations (e.g., cues indicating distractor location, cues consistently in the same hemifield, etc.) that can induce strategic and therefore 'proactive' distractor suppression (Wöstmann et al., 2022; Ferrante et al., 2023). Here we demonstrate that distractor suppression can also operate when strategic suppression based on explicit knowledge is highly unlikely. We demonstrate a tight coupling between target enhancement and distractor suppression during visual search, indicative of an attentional field with surround suppression (Müller et al., 2005; Hopf et al., 2006; Störmer and Alvarez, 2014).

We found that the repetition of search scenes primarily affects the processing of spatial context directly surrounding the target. This finding supports the notion that local context plays a critical role in contextual cueing (Brady and Chun, 2007; Goujon et al., 2015; Sisk et al., 2019). Previous research has demonstrated that merely repeating the local context is sufficient to produce similar contextual cueing to repeating the entire spatial array (Brady and Chun, 2007). However, this mainly indicates that the visual system will exploit what can be learned from our environment, it does not prove global context to be redundant. In our study, we directly demonstrate the importance of local context, even when global context is available. We speculate that the relationship between the attentional field and learning in contextual cueing is bi-directional: the attentional field strengthens the learning of local context and this learned local context subsequently sharpens the attentional field, ultimately sidelining global context entirely. This implies a need for sufficient time to process the local context around the target during visual search, before this bi-directional effect can come into play. Interestingly, Xie et al. (2020) demonstrated that only repeating local context does *not* yield the same contextual cueing effect as repeating the entire search array when presentation time is very short. Further research is necessary to fully understand the role of global context as it clearly plays a role in natural scenes that evoke more global processing (Brockmole et al., 2006).

Importantly, we were able to demonstrate the behavioral relevance of target power compared to the far-standing distractor on all *trials*, both Old and New. This may serve as evidence of efficient versus

inefficient visual search (Tseng and Li, 2004; Schlagbauer et al., 2017), where higher power at the target relative to the distractor at the opposite position of the visual field indicates that the random starting point of search was immediately in the right direction. Our analyses of Old *displays*, instead, show the behavioral advantage of Target power compared to the near standing distractor. This highlights what is actually learned, i.e. the local context. This means that spatial predictive context does not simply speed up visual search by biasing towards the target, but fundamentally changes the process by biasing local attentional competition.

The time course of biased local attentional competition during visual search with spatial predictive context reveals two important aspects. First, when aligning to stimulus onset we did not observe any apparent differences between Old and New trials. This suggests that very early differences between Old and New scenes (Chaumon et al., 2008) are not responsible for the behavioral advantage. Instead, we did find a difference between Old and New trials in our data aligned to the response, indicating the attentional benefit emerges over time as the search progresses (Kunar et al., 2008; Goujon et al., 2015). Similar findings have been reported in recent eye-tracking research, where the benefit of improved attentional guidance consistently occurs during the final 500 ms of the search. Interestingly, this timing was found independent of set size, which was argued to be evidence of the importance of local context in contextual cueing (Harris and Remington, 2020). Second, the difference between Old and New manifested relatively early, reaching its maximum at approximately 400 ms and waning well before the response. This supports the idea that spatial predictive context enhances attentional guidance, rather than facilitating the response itself (Goujon et al., 2015; Sisk et al., 2019).

Previous RIFT studies have predominantly involved low attentional competition, with spatial attention often being explicitly cued to a hemifield (Zhigalov et al., 2019; Seijdel et al., 2023) and a maximum of four, typically large, stimuli on the screen (Ferrante et al., 2023). In our visual search paradigm, stimuli were considerably smaller, spatial attention was more distributed, and competition was uncued, resulting in much subtler differences between stimuli. The fact that we were able to measure attentional competition and the biasing thereof by spatial predictive context highlights the strength of this innovative method.

Lastly, as expected, we established contextual cueing in participants' behavior, demonstrating a robust and reliable difference wherein they were faster in locating a target on Old trials compared to New trials. Notably, participants performed at chance level when tested on their memory of these Old scenes. Moreover, participants who demonstrated larger contextual cueing benefits, did not have

higher accuracy rate in the recognition task. Therefore, we conclude that our results provide no evidence of explicit knowledge acquisition of spatial predictive context.

In conclusion, our findings provide compelling evidence that spatial predictive context improves attentional guidance during visual search. It does so by enhancing the target and suppressing directly surrounding distractors, sharpening the attentional field. This leads to biased attentional competition enabling improved search performance. These results thus further our understanding of how humans can quickly and implicitly learn from their environment and exploit this learning to their benefit.

Acknowledgements

We would like to thank Jorie van Haren for his help with collecting the data.

Funding

This work was supported by The Netherlands Organisation for Scientific Research (NWO Research Talent grant 406.18.508 awarded to FGB, and Vidi grant 452-13-016 awarded to FPdL) and the EC Horizon 2020 program (ERC starting grant 678286 awarded to FPdL)

References

Allen M, Poggiali D, Whitaker K, Marshall TR, Van Langen J, Kievit RA (2021) Raincloud plots: a multi-platform tool for robust data visualization. Wellcome Open Res 4:63.

Bastiaansen MCM, Knösche TR (2000) Tangential derivative mapping of axial MEG applied to event-related desynchronization research. Clin Neurophysiol 111:1300–1305.

Beesley T, Vadillo MA, Pearson D, Shanks DR (2015) Pre-exposure of repeated search configurations facilitates subsequent contextual cuing of visual search. J Exp Psychol Learn Mem Cogn 41:348–362.

Bell AJ, Sejnowski TJ (1995) An Information-Maximization Approach to Blind Separation and Blind Deconvolution. Neural Comput 7:1129–1159.

Brady TF, Chun MM (2007) Spatial constraints on learning in visual search: Modeling contextual cuing. J Exp Psychol Hum Percept Perform 33:798–815.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Brockmole JR, Castelhano MS, Henderson JM (2006) Contextual cueing in naturalistic scenes: Global and local contexts. J Exp Psychol Learn Mem Cogn 32:699–706.

Chaumon M, Drouet V, Tallon-Baudry C (2008) Unconscious associative memory affects visual processing before 100 ms. J Vis 8:10.

Chun MM, Jiang Y (1998a) Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. Cognit Psychol 36:28–71.

Chun MM, Jiang Y (1998b) Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. Cognit Psychol 36:28–71.

Drijvers L, Jensen O, Spaak E (2021) Rapid invisible frequency tagging reveals nonlinear integration of auditory and visual information. Hum Brain Mapp 42:1138–1152.

Ferrante O, Zhigalov A, Hickey C, Jensen O (2023) Statistical Learning of Distractor Suppression Downregulates Prestimulus Neural Excitability in Early Visual Cortex. J Neurosci 43:2190–2198.

Goujon A, Didierjean A, Thorpe S (2015) Investigating implicit statistical learning mechanisms through contextual cueing. Trends Cogn Sci 19:524–533.

Harris AM, Remington RW (2020) Late guidance resolves the search slope paradox in contextual cueing. Psychon Bull Rev 27:1300–1308.

Hickey C, Di Lollo V, McDonald JJ (2009) Electrophysiological Indices of Target and Distractor Processing in Visual Search. J Cogn Neurosci 21:760–775.

Hopf J-M, Boehler CN, Luck SJ, Tsotsos JK, Heinze H-J, Schoenfeld MA (2006) Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. Proc Natl Acad Sci 103:1053–1058.

Jiang Y, Chun MM (2001) Selective attention modulates implicit learning. Q J Exp Psychol Sect A 54:1105–1124.

Jiang YV, Sisk CA, Toh YN (2019) Implicit guidance of attention in contextual cueing: Neuropsychological and developmental evidence. Neurosci Biobehav Rev 105:115–125.

Jung T-P, Makeig S, Humphries C, Lee T-W, McKeown MJ, Iragui V, Sejnowski TJ (2000) Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37:163–178.

Kunar MA, Flusberg SJ, Wolfe JM (2008) Time to guide: Evidence for delayed attentional guidance in contextual cueing. Vis Cogn 16:804–825.

Luck SJ, Girelli M, McDermott MT, Ford MA (1997) Bridging the Gap between Monkey Neurophysiology and Human Perception: An Ambiguity Resolution Theory of Visual Selective Attention. Cognit Psychol 33:64–87.

Makovski T, Jiang YV (2010) Contextual cost: When a visual-search target is not where it should be. Q J Exp Psychol 63:216–225.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190.

Minarik T, Berger B, Jensen O (2022) Optimal parameters for Rapid Invisible Frequency Tagging using MEG. Neuroscience. Available at: http://biorxiv.org/lookup/doi/10.1101/2022.12.21.521401 [Accessed June 8, 2023].

Morey RD, Rouder JN (2022) BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-4.4. Available at: https://CRAN.R-project.org/package=BayesFactor.

Müller NG, Mollenhauer M, Rösler A, Kleinschmidt A (2005) The attentional field has a Mexican hat distribution. Vision Res 45:1129–1137.

Noonan MP, Adamian N, Pike A, Printzlau F, Crittenden BM, Stokes MG (2016) Distinct Mechanisms for Distractor Suppression and Target Facilitation. J Neurosci 36:1797–1807.

Ogawa H, Takeda Y, Kumada T (2007) Probing attentional modulation of contextual cueing. Vis Cogn 15:276–289.

Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Comput Intell Neurosci 2011:1–9.

Peelen M, Berlot E, De Lange F (2023) Predictive processing of scenes and objects. PsyArXiv. Available at: https://osf.io/nuf59 [Accessed July 13, 2023].

Pelli DG (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spat Vis 10:437–442.

R Core Team (n.d.) R: A language and environment for statistical computing. Available at: http://www.R-project.org/.

Schankin A, Schubö A (2009) Cognitive processes facilitated by contextual cueing: Evidence from event-related brain potentials. Psychophysiology 46:668–679.

Schlagbauer B, Mink M, Müller HJ, Geyer T (2017) Independence of long-term contextual memory and short-term perceptual hypotheses: Evidence from contextual cueing of interrupted search. Atten Percept Psychophys 79:508–521.

Seidl KN, Peelen MV, Kastner S (2012) Neural Evidence for Distracter Suppression during Visual Search in Real-World Scenes. J Neurosci 32:11812–11819.

Seijdel N, Marshall TR, Drijvers L (2023) Rapid invisible frequency tagging (RIFT): a promising technique to study neural and cognitive processing using naturalistic paradigms. Cereb Cortex 33:1626–1629.

Sisk CA, Remington RW, Jiang YV (2019) Mechanisms of contextual cueing: A tutorial review. Atten Percept Psychophys 81:2571–2589.

Slagter HA, Prinssen S, Reteig LC, Mazaheri A (2016) Facilitation and inhibition in attention: Functional dissociation of pre-stimulus alpha activity, P1, and N1 components. NeuroImage 125:25–35.

Stoletniy AS, Alekseeva DS, Babenko VV, Anokhina PV, Yavna DV (2022) The N2pc Component in Studies of Visual Attention. Neurosci Behav Physiol 52:1299–1309.

Stolk A, Todorovic A, Schoffelen J-M, Oostenveld R (2013) Online and offline tools for head movement compensation in MEG. NeuroImage 68:39–48.

Störmer VS, Alvarez GA (2014) Feature-Based Attention Elicits Surround Suppression in Feature Space. Curr Biol 24:1985–1988.

Torchiano, Marco (2016) Effsize - a package for efficient effect size computation. Available at: https://zenodo.org/record/1480624 [Accessed June 1, 2023].

Tseng Y-C, Li C-SR (2004) Oculomotor correlates of context-guided learning in visual search. Percept Psychophys 66:1363–1378.

Võ ML-H, Boettcher SE, Draschkow D (2019) Reading scenes: how scene grammar guides attention and aids perception in real-world environments. Curr Opin Psychol 29:205–210.

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis, 2nd ed. 2016. Cham: Springer International Publishing : Imprint: Springer.

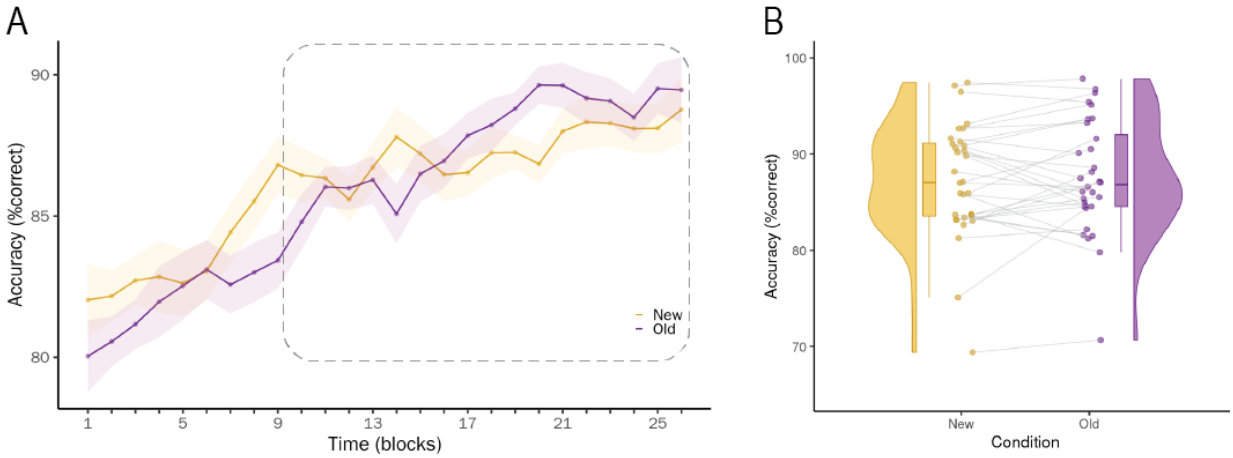
Wolfe JM (2021) Guided Search 6.0: An updated model of visual search. Psychon Bull Rev 28:1060–1092.

Wöstmann M, Störmer VS, Obleser J, Addleman DA, Andersen Søren K, Gaspelin N, Geng JJ, Luck SJ, Noonan MP, Slagter HA, Theeuwes J (2022) Ten simple rules to study distractor suppression. Prog Neurobiol 213:102269.

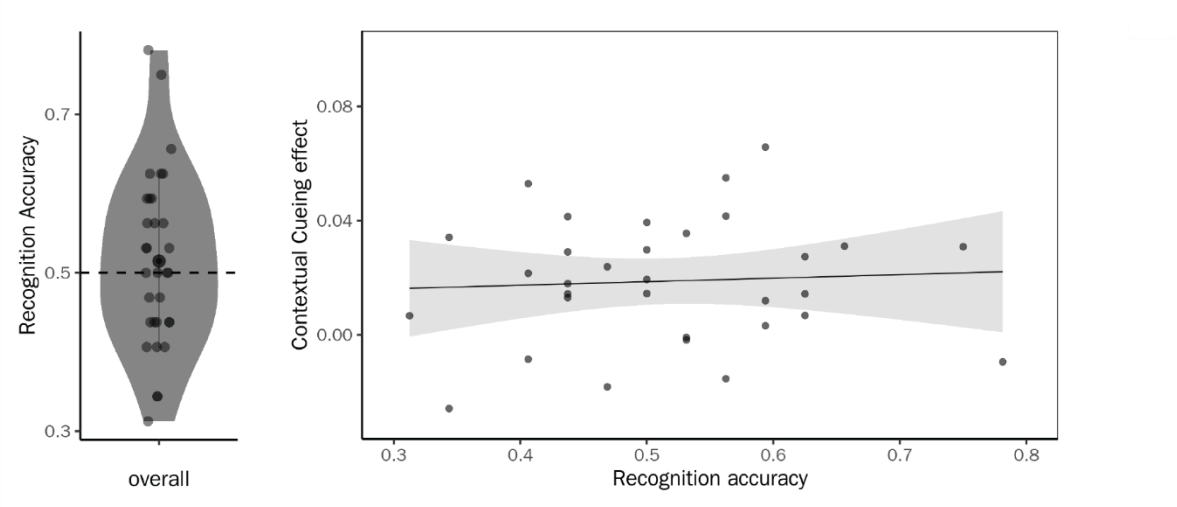
Xie X, Chen S, Zang X (2020) Contextual Cueing Effect Under Rapid Presentation. Front Psychol 11:603520.

Zhigalov A, Herring JD, Herpers J, Bergmann TO, Jensen O (2019) Probing cortical excitability using rapid frequency tagging. NeuroImage 195:59–66.

Supplementary materials



Supplementary figure 1. Accuracy. (A) Percentage correct (smoothed across neighbors, by taking the mean of block N, N − 1 and N + 1) plotted over the time course of the experiment (shading indicates within-participant corrected standard error of the mean). Dotted box indicates the trials used contrasting Old to New scenes, where contextual learning is assumed to have taken place. (B) Accuracy on New versus Old trials on learned trials. Dots are individual participants. The boxplot represents the median and interquartile range.



Supplementary figure 2. Recognition task. (A) Recognition accuracy (B) Contextual cueing (RTNew – RtOld) as a function of recognition accuracy. Dots in all panels represent individual participants.