Dissertation der Graduate School of Systemic Neurosciences der Ludwig-Maximilians-Universität München

Learning how: Procedural learning as a central element of visual statistical learning

Graduate School of Systemic Neurosciences
LMU Munich

submitted by Werner Seitz

21. April 2024

supervised by

Thomas Geyer

Date of submission: 21 April 2024

Date of oral examination: 10 February 2025

First reviewer: Prof. Dr. Thomas Geyer

Second reviewer: Prof. Dr. Christian Leibold

Third reviewer: Prof. Dr. Simone Schütz-Bosbach

Fourth reviewer: Dr. Virginia Flanagin

Index

Learning how: ‘Knowing that’ and ‘knowing how’ in statistical learning of contextual repetitions in a visual search task

General Introduction

Ryle (1945) famously made the distinction between ‘knowing that and knowing how’: For instance, knowing the rules and even having been provided with abundant instructions by an expert, does not make a good chess player. Conversely, a chess expert may not be able to explicate their method of success, while, notwithstanding, performing exceptionally (Ryle, 1945).

This principle holds true for many aspects of life, be it for playing chess, gardening or mathematics. However, there has been little research into the effects of ‘knowing how’ regarding our core cognitive functions, such as perception and, intricately related to it, selective attention.

Background and motivation

In fact, theories of visual search, the prevailing framework under which visual attention is investigated (Treisman & Gelade, 1980; Wolfe et al., 1989; Wolfe, 1994, Wolfe, 2021) predominantly focus on factors that allow to (learn to) “know that” something is the target given a set of stable visual objects and events in the sense that assessing the visuospatial properties of these (with the target covarying) items is akin to propositional statements whether an item should be considered as the target of the search (Chun & Jiang, 1999).

In contrast to this, an alternative, and so far, neglected account can be conceived, in which identifying the target is not a function of e.g., a process of elimination by performing computations in feature space (for early evidence of overt attention being placed not according to optimal feature selection, see e.g. see Zelinsky, 1996) but is instead a direct result of behavior which itself is intelligent (Ryle, 1945). Likewise, faster identification of the target in statistically informative environments of non-target, distractor, objects cannot just be thought of as a consequence of learned associations between the individual layout formed by the distractor elements and the individual target location (i.e., contextual-cueing effect; e.g., Chun & Jiang, 1998; Brady & Chun, 2007). Instead, an argument can be made that learning and refining the process of search itself, i.e. the ability to (learn to) “know how” the target can be searched for and found in an effective manner, may be a core mechanism in how we can make use of our environment for adaptive behavior.

The goal of this thesis is to demonstrate how procedural “knowing how” learning contributes, and moreover, proactively facilitates visual search, demonstrated on the example of contextual cueing. While it is quite well accepted that procedural learning aids performance in cognitive tasks (Fitts & Posner, 1967, Anderson, 1982), here, the argument is made that functions associated with attentional enhancement or attentional suppression (Geng, 2014) can be explained as an instance of procedural learning, without involving a rather “etheric”, intangible notion of attention (James, 1980). Further, learning a procedure of the task instead of learning propositional statements about the goal of the task (i.e. “meta information”, such as the computation of abstract conditional probability distributions) may often better reflect the cognitive processes (not) involved.

The Contextual Cueing task

The central paradigm utilized is that of Contextual Cueing (Chun & Jiang, 1998): In each trial, participants have to search for a target, the letter T, which is rotated by either 90° or 270°, hidden among a set of rotated L-letters, and have to indicate the T's (left vs. right) orientation by pressing one of two buttons, with the response verifying whether the target was indeed detected. Trials are organized into blocks, and in each block, target positions (but usually not their orientations) repeat. Unbeknownst to the participants, for half of those trials, not only the target position repeats, but also the (at the beginning of the search task randomly generated) configuration of L-letters. These are termed ‘repeated’ or ‘old’ contexts, in contrast to ‘non-repeated’ or ‘new’ contexts in which the L-configuration is generated anew for each block. The reason for keeping target positions constant in repeated and non-repeated contexts is to control for absolute target position learning (Geng & Behrman, 2005) and therefore allowing to investigate the influence of an invariant context on the search performance. Typically, the number of repeated and non-repeated displays is kept identical, and one block has usually 12, ranging from 4 (Schankin & Schubö, 2010; Sewell et al., 2019), 12 (Chun & Jiang, 1998), up to 18 (Jiang & Kramer, 2004) target locations for repeated, as well as for non-repeated displays. By keeping the target locations constant in non-repeated displays as well as repeated displays, the effect of context, independently of absolute target position learning (Shaw & Shaw, 1977; Shaw, 1978; Miller, 1988; Geng & Behrmann, 2005).

The central outcome of these studies is that a repeating context facilitates search in terms of reduced response times (Chun & Jiang, 1998, Jiang & Chun, 2003).

Furthermore, eye-tracking studies show that targets in repeated contexts are not only found faster, but also require fewer eye-movements (Peterson & Kramer, 2001; Tseng & Li, 2004). Also of note, a general decrease in reaction times and the number of fixations across blocks can be observed which is, however, more pronounced for repeated than for non-repeated displays, which can be well approximated by a power function, while the power function of repeated displays usually has a steeper slope (Jiang & Chun, 2003; Brooks et al., 2010). Power functions have long been recognized to characterize improvements in human performance as a consequence of practice in various tasks (Snoddy, 1926; Fitts & Posner, 1967; Newell & Roesenbloom, 1981, Anderson, 1982), particularly in association with proceduralization of a cognitive skill (Fitts & Posner, 1967; Anderson, 1982).

The predominant view of contextual cueing is that, through repeated encounters with a display layout, observers acquire a contextual memory that associates the invariant configuration of distractors with a certain target location. Upon re-encountering a repeated display, this memory is activated and guides attention towards the target location (Chun & Jiang, 1998). Specifically, this can happen either via the guidance of individual distractor items' location towards the target (e.g. see Jiang & Wagner, 2004, Experiment 1; for a theoretical account, see Brady & Chun, 2007), or, additionally, by the buildup of a 'scene memory' for repeated distractor locations (Beesley et al., 2015; Preuschhof et al., 2019), which subsequently facilitates target detection through suppression of (learned) distractor locations.

Aim and motivation

The central aim of this thesis is to test an alternative account of CC (which is not mutually exclusive with the previously introduced view): Instead of display-specific memories that guide attention, display-general procedural learning, which is a universal finding in virtually all pertinent studies (e.g., Jiang & Chun, 2003) may itself be not only the motor behind a general decrease in reaction times and the number of fixations, but also give rise to the CC effect itself. – Procedural learning refers to the observation that search speed improves, typically quite substantially, over the course of practice on the task, that is, across trial blocks. According to this view, observers learn to optimize the search procedure by which they scan displays. Importantly, as repeated displays are - by definition - encountered repeatedly, they weigh in stronger on (the development of) the search procedure as non-repeated displays. As a result, the search procedure will be optimized towards repeated displays, and consequently, performance on these displays will be better, compared to non-repeated displays, akin to the notion of overfitting in machine learning. Three studies were conducted to assess display generic learning in CC and whether this form of learning suffices to elicit a contextual facilitation effect (over and above that coming from display-specific memories).

Not surprisingly, the central measure of these studies are eye-movements. Eye-movements have long been recognized to reveal attentional processing, with fixations preferentially being placed at relevant locations and, crucially, in a task-dependent manner (Yarbus, 1967). Specifically, eye-movements have been the subject of several studies involving contextual manipulation, showing that repeated contexts require fewer eye-movements to identify the target (Peterson & Kramer, 2001), and that the length of the oculomotor scanpath is closer to the minimal scanpath length when compared against non-repeated displays (Brockmole & Henderson, 2006). Conversely, the number of fixations approaching the target monotonically does not differ between contexts, however the number of seemingly random fixations is reduced in repeated contexts, in the sense that fixations are not monotonically approaching the target (Tseng & Li, 2004): Tseng and Li (2004) computed distances between each fixation and the target location and based on the fixation that was maximally distant from the target, they classified fixations preceding and including the fixation at peak distance as random (or explorative) and subsequent fixations as monotonic. Taken together, these results were interpreted as evidence for the activation of a display-specific memory trace.

Somewhat at odds with these interpretations are studies that suggest that fixations not only serve the purpose of detecting a target, but also to acquire visual information in general (Zelinsky, 1996), to maximize task-relevant information (Najemnik & Geisler, 2005) or to avoid costs (Araujo et al., 2001). While observations such as these are consistent with the idea that there might be good and less optimal scanning strategies in which the strategy itself (i.e. the knowing how) is optimized/learned, so far no study has explicitly investigated this aspect of visual search. Paradoxically, the contextual cueing task, which is generally believed to illustrate display-specific top-down guidance, is also ideally suited to assess whether the procedure of scanning is optimized in a display-general manner as a result of some of them repeating.

To rephrase: The relationship between eye-movements and attention has been researched exhaustively in terms of how covert attention affects eye-movements (Sheliga et al., 1995; Hoffman, 2016). While it is certainly the case that shifts in attention elicit eye-movements, e.g., when attention is explicitly cued toward a particular location (Deubel & Schneider, 1996), so far the complement has received little to no attention, which is how a fixational policy can aid (pre-)attentional appraisal of an environment when there are no or only relatively weak cues to attention, that is, when fixations serve the purpose of collecting information and bringing relevant aspects of an environment into view in the first place. Given that in the classical contextual cueing task, an L-vs.-T letter search, each element is composed of identical features and that participants still have to search for the target even in repeated displays after ample exposure (e.g. see Peterson & Kramer, 2001), this task lends itself well to the assessment of oculomotor scanning strategies.

Studies

Study 1 investigates procedural learning in a Contextual Cueing task, by investigating whether not only the general improvement in visual scanning, as time progresses, but also the improved performance in repeated over non-repeated contexts is a function of procedural learning of the search itself: Since repeated displays have to be searched multiple times, while non-repeated displays are encountered only once, repeating contexts weight more strongly on the tuning of the search process. Since the search procedure is hence biased towards repeated displays, it follows that generic display scanning is adjusted toward and works more optimally on those repeated contexts.

This goes along with a reduced variability, in response to the general invariances of the task, and also particularly in response to the additional invariances provided by the overrepresentation of repeated displays for repeated contexts. As a result, the search procedure for each repeated display becomes increasingly similar.

The study investigated search behavior in terms of eye movements, and how the oculomotor scanpaths of participants unfold, conceptualized as a sequence of fixations over time. Besides replicating measures previously established in literature (a reduced reaction times, see Chun & Jiang, 1998; reduced number of fixations, see Peterson & Kramer, 2001; smaller scan-pattern-ratio, see Brockmole & Henderson, 2006 and Henderson et al., 1999) the central finding is an increase in similarity: Both, within participants (across each participant’s set of repeated and non-repeated displays, respectively) and within displays (i.e. across all participants for each individual repeated and each non-repeated display), similarity of scanpaths increases as the experiment progresses, and, importantly, similarity is higher for repeated than for non-repeated displays.

Having established that procedural learning takes place, study 2 aimed at answering two questions. Firstly, replicating the results from study 1 that procedural learning is an important source of the CC effect, while also examining the contribution of alternative, and not mutually exclusive, display-specific learning to search facilitation in repeated contexts. Secondly, after assessing the relation and interplay between procedural, that is, display-generic, and display-specific learning, the possible (evolutionary) advantages need to be explicated: By relying on a mostly procedural strategy, the perceptual system does not only save energy in terms of the costs of encoding each display, but also avoids interference from having to represent multiple similar memory traces.

To this end, oculomotor behavior was investigated before and after relocating the target to a new position of repeated displays (Manginelli & Pollmann, 2009): Relocation after having established a stable CC effect usually leads to a cost, i.e. the CC effect vanishes and only re-emerges slowly (Manginelli & Pollmann, 2009; Zellin et al., 2014). While CC re-emerges during the relocation phase in the study, the magnitude remains distinctly smaller than during the learning phase. Critically, observers’ gaze approaches previous target locations much closer in the relocation phase than the future, i.e., relocated, target position in the learning phase. Moreover, the closeness of fixations towards the previous target location explained the reaction time cost of relocation. These findings indicate that display-specific learning indeed plays a role in CC. Importantly however, previous targets on the left and on the right side were not approached equally closely, and reaction times, as well as the number of fixations, showed significantly different costs between different old-context displays (with original targets in left vs. right hemifields). Specifically, acquired visual (left-right) scanning strategies could explain the pattern of contextual cueing after target relocation. For instance, a target relocated from the left to the right hemifield produced a huge cost, amounting up to 4 additional fixations and additional 200 ms processing time. This contrasts with targets relocated from the right to the left display half, which even produced slight performance advantages. Taken together, these findings suggest that while traditional accounts of CC emphasize the acquisition of search-guiding LT-memory 'templates' that are specific to particular target-distractor contexts, contextual learning also tunes attentional (oculomotor) scanning routines to the prevailing statistical target-distractor regularities in the display arrangement encountered, yielding a context-unspecific LT 'proceduralization' of search: Although, on average, search incurred a cost in repeated display after target relocation, the amount differed, based on whether the target moved from the left to the right, or the right to the left half, before and after relocation, respectively. Thus, this study confirmed both mechanisms to contribute to initial contextual learning as well as the 'distraction' effect produced by re-location of the target to the opposite side of repeated-context displays. Hence, it can be theorized that guidance and misguidance of search by repeated contexts may involve two complementary LT mechanisms: procedural optimization of broad, i.e., display-generic, oculomotor scanning routines, and learning of where to expect the target to be located in specific repeated-context displays.

The third study aimed at investigating whether procedural learning alone can, in principle, give rise to the CC effect. In order to assess this hypothesis, a connectionist model was implemented. The advantage of a computational model is that it allows to analyze explicitly the aptness of procedural learning in isolation, in the strict absence of any configurational, i.e., display-specific, learning. Instead, the model was only able to learn the procedure and was never informed about the target's identity or its location.

Nevertheless, the model was able to improve its search performance and develop a CC effect. Moreover, a central bias (Clarke & Tatler, 2014) developed as an emergent phenomenon of learning the search procedure. The model outperforms previous models on various key metrics. This has implications for models of vision to incorporate procedural learning and procedural strategies, and also for the field of artificial intelligence in which acquiring procedural strategies that leverage the task structure can inspire novel learning mechanisms and datasets, as well as alleviate the need for annotated datasets.

Study 1 – Seitz et al. (2023)

Seitz, W., Zinchenko, A., Müller, H. J., & Geyer, T. (2023). Contextual cueing of visual search reflects the acquisition of an optimal, one-for-all oculomotor scanning strategy. *Communications Psychology*, *1*(1), 20.

Author contributions

WS and AZ conceived and designed the experiments; AZ performed the experiments; **WS** analyzed the data; HJM contributed to conceptual analysis and discussion; **WS**, HJM, TG wrote the paper.

communications
psychology

ARTICLE

https://doi.org/10.1038/s44271-023-00019-8 OPEN

Contextual cueing of visual search reflects the acquisition of an optimal, one-for-all oculomotor scanning strategy

Werner Seitz [1✉], Artyom Zinchenko[1], Hermann J. Müller[1,2] & Thomas Geyer[1,2,3]

Visual search improves when a target is encountered repeatedly at a fixed location within a stable distractor arrangement (spatial context), compared to non-repeated contexts. The standard account attributes this contextual-cueing effect to the acquisition of display-specific long-term memories, which, when activated by the current display, cue attention to the target location. Here we present an alternative, procedural-optimization account, according to which contextual facilitation arises from the acquisition of generic oculomotor scanning strategies, optimized with respect to the entire set of displays, with frequently searched displays accruing greater weight in the optimization process. To decide between these alternatives, we examined measures of the similarity, across time-on-task, of the spatio-temporal sequences of fixations through repeated and non-repeated displays. We found scanpath similarity to increase generally with learning, but more for repeated versus non-repeated displays. This pattern contradicts display-specific guidance, but supports one-for-all scanpath optimization.

---

[1] Department Psychologie, Ludwig-Maximilians-Universität München, Munich, Germany. [2] Munich Center for Neurosciences – Brain & Mind, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Munich, Germany. [3] NICUM - NeuroImaging Core Unit Munich, Ludwig-Maximilians-Universität München, Munich, Germany. ✉email: Werner.Seitz@campus.lmu.de

Visual search for a target object among nontarget, or distractor, objects can be facilitated by prior knowledge of the scene, including contextual long-term memory of co-occurring objects or the position of the target relative to an invariant (spatial) distractor arrangement (e.g., ref. [1]). Chun and Jiang[1], in their seminal study, presented participants with search arrays containing a target letter "T" among a set of distractor letters "L" (a relatively hard search task that affords little bottom-up or top-down guidance; cf[2].). Critically, in half of the trials, the spatial arrangements of the distractor and target stimuli were repeated (i.e., repeated, old contexts); in the other half, the distractor locations were generated anew on each trial, while keeping the target position constant (i.e., nonrepeated, new contexts). Thus, given that the absolute target positions were fixed in both types of trial, the only difference between them was whether or not the target location was predictable from the distractor context. The two sets of repeated and nonrepeated displays were presented randomly intermixed within each block of trials. Chun and Jiang[1] observed that the search reaction times (RTs) decreased with increasing block number for nonrepeated as well as repeated arrays, attributed to general procedural learning of how to perform the task. Critically, however, this practice-dependent improvement was larger for the repeated arrays – an effect Chun and Jiang[1] referred to as contextual cueing (CC). Search RT facilitation by repeated contexts typically emerges rapidly, after just a few (2–4) encounters of the same visual arrangement (e.g., ref. [3]), and it appears to be implicit (and automatic) in nature[4]. Further, in terms of oculomotor search performance, eye-tracking studies indicate that contextual learning leads to a reduction in the number of fixations required to reach the target in repeated, compared to nonrepeated, displays (e.g., ref. [5–9]).

One intriguing, and, as we will argue below, open question is how these savings in the number of fixations are actually produced in repeated search displays. The standard account attributes these savings to search being cued, or guided, more directly to the target location as a result of having acquired a (long-term) associative memory representation, or template, of a specific distractor-target arrangement. This template is activated upon re-encountering such an arrangement on a given trial, which then top-down increases the attentional priority of the target location (e.g., ref. [1]; for computationally explicit models, see, e.g., ref. [10,11]) – thus enhancing the target's potential to summon covert or overt attention. According to this account, the number of attention shifts required to detect a target in a repeated search array will decrease with increasing (re-)encounters of this array, due to the build-up of a search-guiding contextual memory template for this array (e.g., ref. [1]). Support for this comes from studies of contextual cueing that used fixation number as a dependent measure (e.g., ref. [5,6],). These studies showed that finding the target in repeated arrays requires overall fewer eye movements – though with the guidance effect emerging only after the first few fixations, suggesting that it may take some time for the template to come into play. In Tseng and Li's[6] terms, search may involve some 'inefficient', unguided scanning of the array until an informative constellation of distractors represented in the template is encountered. Activation of the template would then lead an effective, guided search phase: a relatively direct homing in of attention on the target location after a series of more exploratory fixations. Accordingly, the savings in the number of eye movements for repeated (vs. nonrepeated) displays would arise from later fixations in the saccadic scanpaths – perhaps with the template-based priority signal pointing to the target location growing increasingly stronger as oculomotor scanning approaches the target item[1]. This specific-template account is attractive, not least because it ties in seamlessly with the functional architecture assumed by general theories of search guidance, such as Guided Search[12–14].

However, there may be an alternative, more procedural account of contextual facilitation that does not rely on the notion that observers acquire memory representations that are specific to particular distractor-target arrangements – a conceptually new account that the present study set out to explore. In fact, procedural learning in CC paradigms is a universal finding in virtually all pertinent studies (for reviews see, e.g., ref. [15,16],): search speed improves, typically quite substantially, over the course of practice on the task, that is, across trial blocks[4]. Importantly, an improvement is evident for nonrepeated – as well as repeated – displays, which is generally attributed to procedural learning, which optimizes, or automatizes, performance through the development/refinement of a task-appropriate (search) settings, akin to the development of a skill (e.g., ref. [17–20],). Critically, though, the improvement is more marked for repeated (vs. nonrepeated) displays, which constitutes the contextual-facilitation effect.

Of note, the extant studies of contextual cueing have almost all examined the facilitation effect (in terms of RT, fixation/saccade number, etc. measures) across aggregated sets of repeated vs. nonrepeated displays. Accordingly, arguably little is known about how contextual facilitation comes about at the single-display level: Is it based on attentional guidance by specific LTM templates of spatial target-distractor relations in individual displays? Or is it due to the acquisition of more display-generic (i.e., relatively display-independent) scanning procedures that are mainly shaped by – and so best adjusted to – the set of repeated displays. On the latter hypothesis, what is optimized in procedural task learning may be a search strategy which is increasingly generic in the sense that it is applicable to all search displays, repeated and nonrepeated (rather than being specific for particular repeated displays). However, as a result of *statistical learning*, this strategy is more tuned to, and so more effective for, those displays that are encountered frequently (repeated displays), rather than displays searched only once (nonrepeated displays). Thus, the procedural-learning hypothesis would provide a *unitary* account in that it explains both the general and the specific gains in terms of tuning and optimization of the oculomotor scanning strategy to the regularities prevailing in the whole set of displays that observers encounter over the task. Please note that we evoke a view of optimization according to which visual search is adjusted toward a specific goal, namely, finding and responding to a target letter T in a cluttered array of distractor letters L. Given that this target differs from the distractors only in the combination of two shape features and the T vs. L junction, the search as such is likely inefficient, in terms of producing relatively steep slopes of the function relating RTs to the number of elements in the search display (e.g., ref. [14].). Nevertheless, through procedural learning, performance is optimized to achieve the goal reliably with a minimum of effort.

In sum, in hard search tasks requiring serial eye movements to find the target, repeatedly scanning identically composed item arrays leads to a decrease of RTs and fixation numbers compared to novel displays. However, the eye-movement savings (likely the main driver of the RT savings) occur only relatively late during the trial; and even after a reasonable amount of display repetitions, a considerable number of (some 4–6) fixations is still needed for the eye to reach the target (cf[6].). This suggests that contextual learning may foremostly aid, or optimize, the selection of fixation locations along (at least parts of) the oculomotor scanpath, thereby increasing the likelihood of hitting the target location relatively early during the search. In other words, contextual learning may drive adaptations of participants' general scanning strategies that broadly structure their search in a display-generic manner that is adapted to repeatedly encountered displays collectively, rather than individually (and that is little influenced by nonrepeated displays, which – by virtue of being encountered only once – cannot consistently contribute to shaping this strategy). Of course, such display-generic learning may operate alongside display-specific learning of the spatial target-distractor relations in individual repeated displays. Arguably, however, relatively direct, display-specific guidance of attention and the eye to the target location may only play limited role at least in hard search tasks requiring serial scanning.

To test this alternative, proceduralization account, we set out to, first of all, establish (and thus replicate) contextual facilitation in terms of the standard summary RT and eye-movement measures that have informed theorizing in the extant contextual-cueing literature. Then, we went on to examine oculomotor-scanpath-similarity measures – in particular, Dynamic Time Warping, Discrete Fréchet Distance, and Area Between Curves – that are diagnostic of similarity in the spatio-temporal sequence of fixations across individual (repeated and nonrepeated) displays, as well as the sequences produced by individual participants. These analyses were designed to reveal detailed information about the proceduralization of search performance, which is lost in the standard averaging of dependent measures both across individual repeated and, respectively, nonrepeated displays and across individual observers.

According to the procedural-optimization hypothesis, (1) scanpaths should become more homogeneous for individual displays across participants over trial blocks, with scanpaths for repeated displays becoming more similar compared to those for nonrepeated displays. Given that (any acquired) display-specific contextual-memory templates take time to become (fully) activated to provide direct guidance (e.g.[6,5],), higher scanpath homogeneity would particularly reflect display-generic eye-fixation sequences during the earlier, unguided parts of the search. (2) There should be an increased similarity when scanpaths for different displays are compared within individual participants: similarity measures should be higher for pairs of (differently composed) repeated displays compared to pairs of (different) nonrepeated displays. In contrast, the display-specific hypothesis of contextual-cueing would predict that scanpaths become more dissimilar for pairs of repeated (relative to nonrepeated) displays.

Methods

Participants. The sample size was determined based on Vadillo et al.'s[21] meta-analytical study of contextual cueing (i.e., which reported a rather large effect-size score of Cohen's d = 1.00). A power analysis based on this meta-analysis indicated that to find a main effect of contextual cueing on RT performance with 85% power, a minimum sample size of N = 11 participants would be needed. Based on this estimate, when analyzing contextual-facilitation effects at the level of each of our 4 individual repeated displays, this would require at least a 4 times larger sample size. In fact, a sample size of $N > 40$ participants is comparable with other, relevant studies of contextual cueing that have examined contextual facilitation at the level of individual learning blocks/epochs (e.g.[22],) or of single displays (e.g.[23,24],). Based on these considerations, we recruited $N = 46$ participants for the present experiment (38 identifying themselves as female, the remaining 8 as male; 3 left-handed; mean age = 23.28 [SD = 5.62, range = 19–43] years; no data on ethnic identity was collected). Written consent was obtained from each participant; with an ethics approval by the German Research Council (DFG; under GE 1889/4-2). Note that for nonsignificant effects, we additionally report Bayes statistics, where we used the Bayesian Information Criterion as approximation to the Bayes factor (BF_{10}; see ref. [25,26]).

General approach. Our goal was to bring together established RT and oculomotor measures of the contextual-cueing effect, which focus on group mean values, with oculomotor-scanpath measures that quantitatively describe search behavior in a more fine-grained manner, in particular, at the levels of individual displays or individual participants (see Fig. 1A and B). Additionally, by replicating established measures from the literature (ref. [5,6,27,28]), we aimed to ensure the representativeness of our own data for contextual-cueing studies at large, thus increasing the confidence in the generality of our analyses and findings. Please note that our study was not preregistered as we had a particular – exploratory – focus that seeks to find primary evidence for an alternative, procedural, account of statistical learning in search tasks, by also demonstrating the applicability and potential of scanpath comparison techniques to visual search in repeated versus non-repeated target-distractor arrays and thus generating ideas that justify further research.

For our analysis approach to be feasible, we adjusted the experimental design in two respects: First, and motivated by a previous study of contextual cueing[29], we reduced the number of learnable, repeated target locations, as well as the number of target locations in nonlearnable, nonrepeated displays to four each, with one target location per display quadrant; this was meant to ensure that the memory signals for the respective target location and the corresponding (possibly display-specific) scanpath would have as little interference from other repeated displays as possible and that allocation of attention over space and time would be maximally different. Second, we presented the same repeated and non-repeated display arrangements to all participants, in the same trial order. Using the same set of displays allowed us to control the perceptual content of the display set throughout the experiment; in particular, using the same arrangements for non-repeated displays ensured a "fair" comparison between scanpaths, eliminating confounds originating from, across participants, variably composed distractor-target configurations in non-repeated displays. Methodologically, these adjustments made it possible to compare pairs of scanpaths at different levels and relating to (1) the similarity of fixation sequences through an individual display when viewed by pairs of different participants and (2) the similarity of scanpaths for an individual participant viewing (pairs of) different displays. These design measures enabled us to perform a thorough test of the contrasting predictions made by the specific and the generic procedural-optimization accounts.

We acknowledge that the number of (4) consistently arranged target-distractor displays employed here is relatively low compared to the 8–12 repeated arrays typically used in the relevant studies (see, e.g., ref. [21],). Assuming that having to deal with fewer repeated displays fosters the acquisition of contextual regularities, the facilitation effect generated under the present conditions may turn out more robust than the meta-analytical effect reported by Vadillo et al.[21], with a Cohen's d effect-size score of 1.00. However, this is not supported by the present RT data (see below), which revealed a Cohen's d = 0.90 (95% CI: 0.30 − 1.51) – rendering it unlikely that contextual facilitation is a simple function of the number of different repeated displays encountered in an experiment. Nevertheless, our participants were presented with identical sets of repeated (and nonrepeated) displays. While this was a necessity for our scanpath analysis to work (in particular, for permitting scanpaths to be compared between different participants searching the same displays), it remains a possibility that the results are bound to these displays. To address this, we used linear mixed models, in which we explicitly took into account the *random* variability coming from individual nonrepeated displays (as well as individual participants) when estimating the effects of our fixed factors of context

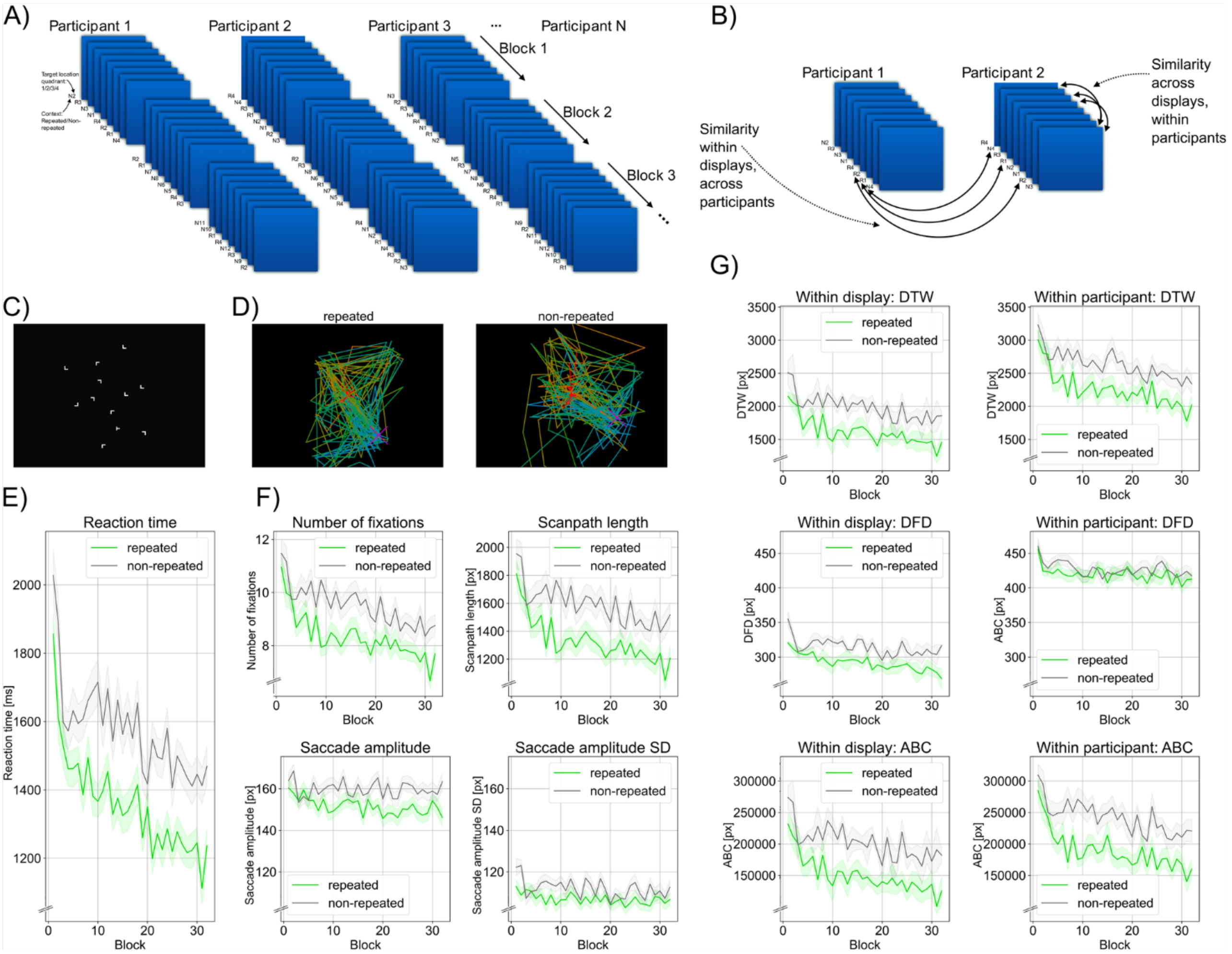

Fig. 1 Design, analytic approach and results from the current study A and **B** illustrate the trial schedule: Each of our $N = 46$ participants viewed the same set of (repeated and nonrepeated) displays, allowing us to compare the scanpath similarities within displays and within participants. **C** shows an example display used in the search task, with a right-oriented T target letter located in the lower right quadrant; and (**D**) illustrates participants' scanpaths when searching for Ts (in displays with the target position in the lower-right quadrant marked by a violet cross) in repeated and nonrepeated displays. **E** shows the reaction-time benefit for repeated over nonrepeated displays (in ms) as a function of block; and (**F**) outlines more effective processing in repeated vs. nonrepeated displays when oculomotor parameters are examined: number of fixations, scanpath length, and means and standard deviations of saccadic amplitudes. The data (except for fixation number) represent pixel values: 52 pixels correspond to 1 degree of visual angle. **G** shows blockwise differences between repeated and nonrepeated displays in scanpath similarity measures when computed using the metrics of Dynamic Time Warping, Fréchet Distance, and Area Between Curves, both within displays and within participants. The data are given in pixel values (or, respectively, squared pixel values for ABC). The shaded areas in (**E**–**G**) depict the standard error of the mean.

and block (the specific structure of our models is presented in the Supplementary Methods).

Apparatus and stimuli. The experimental routine was programmed in Matlab with Psychtoolbox extensions (ref., [30,31]) and run on an Intel PC under the Windows-7 operating system. Participants were seated in a dimly lit laboratory booth in front of a 19-inch CRT monitor (AOC, Amsterdam; display resolution 1024 ×768 pixels; refresh rate: 85 Hz) at a viewing distance of 60 cm (controlled by a chin rest). The search displays consisted of 12 gray items (luminance: 1.0 cd/m^2; 1 target and 11 distractors) presented against a black background (0.11 cd/m^2). All stimuli extended 0.35° of visual angle in both width and height. As depicted in Fig. 1C, the items were arranged on three (invisible) concentric circles around the display center (with radii of 91, 182, and, 273 pixels for circles 1, 2, and 3, respectively). In *repeated* displays, the locations and orientations of the distractors were held constant across trials; in *non-repeated* displays, all distractors (i.e., their locations and orientations) were generated anew on each trial. Note that in all presented displays, the location of the target was repeated but the (left/right) orientation of the target was determined randomly and was, thus, unpredictable. As a result, a repeated context could only be associated with a specific (repeated) target location, but not with a specific target identity. Following Chun and Jiang[1], this approach is used in most CC studies to ensure that contextual facilitation of RTs is owing to the repeated context guiding attention/the eyes, rather than facilitating the selection of the manual response (invariably) associated with a given repeated display. Importantly, both the set of $N = 4$ repeated displays and the set of $N = 128$ randomly generated nonrepeated displays were kept constant across all 46 participants, so that each participant encountered identical repeated and non-repeated configurations. Note, however, that trial order was randomly chosen within each block of $N = 4$ repeated plus $N = 4$ nonrepeated trials for individual participants. This enabled us to keep low-level, individual display properties constant across participants and thus compute dependent – scanpath – measures for each individual display (with variations between participants providing the error term).

There were overall 8 possible target locations, 4 of which were used for *repeated* displays (with constant distractor layouts) and the other 4 for *non-repeated* displays (with random distractor arrangements). Keeping the target locations constant in non-repeated as well as repeated displays is a standard procedure in studies of contextual cueing, to permit the learning of invariant target-distractor contexts to be dissociated from absolute target-position learning (i.e., target-location probability cueing; e.g., ref. [32,33],): absolute target-location learning is effectively controlled for by maintaining constant target positions in both repeated and nonrepeated displays. All targets, in both types of display, were located on the second ring, controlling for the distance of the target from to the display center in all conditions. Furthermore, the targets were placed in all four quadrants with equal probability. Specifically, the (12) display items were randomly assigned to 12 out of a total of 40 possible locations (4, 8, 12, and 16 locations on ring 1, 2, 3, and 4, respectively), the only constraint being that each quadrant contained an equal number of items (either 3 distractors or 1 target and 2 distractors). This means that in principle, there were more than $2.07*10^9$ variants for generating individually unique search displays. Selecting our repeated (and non-repeated) displays randomly from this large space of possible configurations rendered it quite unlikely that they were structurally highly similar.

Importantly, participants were not informed about the fact that some of the search arrays were presented repeatedly. The "T" target was rotated randomly by 90° to either the left or the right. The 11 remaining items were L-shaped distractors rotated randomly at orthogonal orientations (0°, 90°, 180°, or 270°). Figure 1C presents example display layouts (see also Supplementary Fig. 1) To record eye movements, a video-based eye-tracker was used (EyeLink 1000; SR Research Ltd., Mississauga, Ontario, Canada; version 4.594), monitoring participants' right eye. A standard 9-point (grid) calibration of the eye tracker was performed initially and validated at the beginning of each fourth block of (32) trials. Calibration was considered accurate when fixation positions fell within ~1.0° (=diameter) of visual angle for all calibration points. The average calibration error was 0.52°, the average of the maximal errors was 0.86. No data had to be removed due to poor calibration. Calibration accuracy was further checked by the experimenter on each trial. The default psychophysical sample configuration of the eye-tracking system (i.e., saccade velocity threshold set at 35°/s, saccade acceleration threshold set at $9500°/s^2$) was adopted for identifying saccadic eye movements.

Trial sequence. A trial started with the presentation of a central fixation cross (0.10° x 0.10°, luminance: 1.0 cd/m2) for 500 ms. Next, the fixation cross was removed from the screen, and, following a blank interval of 200 ms, the search display was presented. Observers were instructed to find the target "T" and respond as quickly and accurately as possible to its (left vs. right) orientation, while being allowed to move their eyes freely. Each search display stayed on the screen until a manual response was elicited. If the "T" was rotated to the right (left), observers responded by pressing the right (left) arrow button on a computer keyboard with their right (left) index finger. Following a response error, the word "Wrong" appeared in the screen center for 1000 ms. Each trial was followed by a blank inter-trial interval of 1000 ms. The experiment consisted of 256 trials (32 blocks x 8 trials each, 50% repeated displays in each block). Participants were free to proceed to the next block at their own pace. The search task took some 30 min to complete.

Recognition test. At the end of the experiment, observers performed a yes/no (repeated/nonrepeated display) recognition test, permitting us to assess whether they had acquired any explicit memory of repeated configurations presented in the preceding search task (a standard procedure in contextual-cueing experiments; see e.g., ref., [1]). To this end, observers were presented with 4 repeated displays and 4 newly composed displays. The task was to indicate whether or not a given display had been shown previously, by pressing the left or the right mouse button, respectively. The 4 repeated and the 4 newly generated displays were presented in random order. Observers' responses in the recognition task were nonspeeded and no error feedback was provided.

Statistical analysis. Comparisons of scanpaths were carried out in Python[34]. Statistical analysis was performed using Python[34], as well as R (version 3.4.3, ref. [35,36]). We analyzed our dependent measures of reaction times, error rates, and oculomotor variables using the lme4 package in R for linear mixed effect modeling, including target quadrant and participant as random factors in addition to the fixed factors of block and context. All tests reported in this study were conducted as two-sided parametric tests and the aptness of these tests was checked by visualization and formal methods.

Previous studies of contextual cueing reported substantial target-quadrant- and participant-dependent variations in baseline RTs (e.g., ref. [37,38],). Accordingly, we used linear random-intercept models (for the numerical dependent variables of RTs, error rates, oculomotor measures) to account for unwanted variability deriving from individual displays (with different target locations/quadrants) and individual participants. Also, by considering target quadrant and participant as random factors, we ensured that our results would be as generalizable as possible to other studies of contextual cueing. Note that the analyses of the scanpath metrics (see below) required specific variability coming from comparisons of individual displays and, respectively, of individual participants. For this reason, target quadrant and participant were not included as random factors in these analyses. Our model fits (i.e., effect sizes) were quantified in terms of Nakagawa's R^2,[38] using the package "performance" in R.

Dependent measures. To establish comparability between the present investigation and previous contextual-cueing studies, as well as to validate our dataset as being representative for visual-search paradigms, we begin our analysis with an examination and replication of established measures of the contextual-facilitation effect: RT, fixation number, and saccade amplitude (see, ref. [6], or[5]). We then proceed to the presentation of scanpath measures of contextual facilitation, including the total length of the scanpaths (ref. [27,39]) and the standard deviation of the lengths of the saccades constituting each scanpath. The latter essentially provides a new measure of the variability of saccade lengths across individual observers and displays, where a decrease in variability can be considered a measure of automaticity[40]. This is followed by overlay-plot visualizations of individual participants' spatio-temporal sequence of oculomotor behavior (which are also meant to demonstrate the usefulness of our scanpath approach to eye-tracking investigations of visual search in general). From these visualizations, a quantitative measure of contextual cueing, namely: scanpath similarity or consistency, is derived.

Analysis of scanpath similarity. The similarity of scanpaths was computed using established measures in the field (cf[41,42].), in particular, Dynamic Time Warping, Discrete Fréchet Distance, and Area Between Curves.

Dynamic Time Warping is a measure of similarity between two fixational series of different lengths. Two individual scanpaths may be highly similar with regard to the placing (i.e., the spatial coordinates) of individual fixations, but the temporal alignment of these sequences may be less consistent across individual trials. The strength of Dynamic Time Warping is that it can quantify the similarity of the shapes of scanpaths with distinct time series. Specifically, this metric compares two fixational series by aligning them in the time domain, thus minimizing the Euclidean distance between the aligned series. Concerning the Discrete Fréchet Distance: this metric can also deal with fixational time series of different length (and tempo). The Fréchet Distance considers both the location and ordering of the individual fixations along two scanpath curves and can be defined with regard to an analogy: a person that is walking a dog on a leash, with the person walking on one (scan-) path/curve and the dog on another path/curve. The discrete version of the Fréchet Distance only compares distances between fixations and not points in between. Hence, the Discrete Fréchet Distance corresponds to the length of the shortest leash possible for traversing both curves. Finally, we computed scanpath similarity based on Area Between Curves, which, like Dynamic Time Warping and the Discrete Fréchet Distance, permits comparisons of scanpaths with different lengths, although the particular scanpath measure is based on the area that falls between two scanpath curves. As Area Between Curves is well-suited to quantify hysteresis[43], this measure should be particularly sensitive to capture scanpath similarity when trajectories have the same start and end points (initial fixation point and target location).

We chose to explore three scanpath metrics, rather than just one, in order to provide a maximally precise and unbiased measurement of the effects of search task training on oculomotor behavior (as there is not yet a consensus which one of the scanpath measures is to be preferred over the others[41]. Our specific trial schedule (see Fig. 1A and B) allowed us to examine the similarity of scanpaths in multiple ways (see Supplementary Table). First, we compared each possible pair of gaze patterns arising from identical displays over different participants. This approach enabled us to compute scanpath similarity for each experimental block (each consisting of 4 repeated and 4 nonrepeated arrays), thereby addressing the important question of how the consistency of viewing patterns changes as a function of practice on the task. Second, we computed similarity of oculomotor trajectories between each pair of different displays viewed by the same participant. This analysis was intended as a strong test of the display-specific vs. general-procedural accounts of contextual repetitions on search-task training.

To more formally examine whether these observations represent meaningful effects, we computed scanpath similarity for each experimental block (1–32). To recap our hypothesis: if participants are learning a generic search procedure that is increasingly effective, then similarity of scanpaths should increase over time for both repeated and nonrepeated arrays, though this effect should be more pronounced for the former displays which, due to being repeated, accrue a greater weight in shaping the generic search procedure. But the prediction would be fundamentally different for the display-specific hypothesis of contextual cueing, according to which experience with individual repeated displays leads to the build-up of display-specific memories and associated scanning behavior. Accordingly, scanpath similarity obtained for pairs of individual repeated displays with different spatial composition should decrease with increased search-task training (and similarity measures should effectively be lower than those for nonrepeated displays). Two analyses were conducted (see Fig. 1A and B and Supplementary Table). In the first, *within-display* analysis, similarity of eye-movement sequences was calculated from each pair of different participants when viewing a given, individual (repeated or nonrepeated) display. Second, in the within-participant analysis, similarity measures were generated from eye-movement sequences in pairs of different displays when searched by an individual participant. Both analyses were conducted for three similarity measures: Dynamic Time Warping, Discrete Fréchet Distance, and Area Between Curves. Statistical inference was based on linear mixed models with the fixed factors of Context and Block and the random factors of Target Quadrant (in the analysis of within-display similarity) and Participant (within-participant scanpath analysis).

Validity check of scanpath measures. Taking the distance of each fixation from the target position into account, search can be divided into an initial inefficient and a subsequent efficient phase (ref. [6]; see also ref. [8,39]): only in the latter does the distance of a given fixation from the target decrease monotonically with each successive fixation. That is, in hard search tasks requiring oculomotor scanning, high-resolution (foveal) vision ultimately ends in the target region. Moreover, these studies found that the RT advantages for repeated over nonrepeated displays were accompanied by fewer fixations in the ineffective, but not the effective, search phase. Based on these observations, we expected higher scanpath similarity – indicative of display-generic scanning procedures – to manifest particularly in the initial 50% of the scanpath trajectories – as compared to the final 50%, as the eye eventually approach the unique target regions in individual repeated and nonrepeated displays (which would lower the within-participant display similarity). A 50/50 comparison of the scanpath-similarity scores confirmed this prediction: when measuring scanpath similarity across all repeated and nonrepeated displays encountered by individual participants, we found similarity (as measured by all three scanpath metrics) to be overall higher in the initial vs. the final scanpath parts: Dynamic Time Warping, DTW, $t(45) = 45.39$, $p < 0.001$, η^2 (partial) = 0.98, 95% CI: 0.97, 0.99; Discrete Fréchet Distance, DFD, $t(45) = 37.98$, $p < 0.001$, η^2 (partial) = 0.97, 95% CI: 0.95, 0.98; Area Between Curves, ABC, $t(45) = 33.81$, $p < 0.001$, η^2 (partial) = 0.96, 95% CI: 0.94, 0.97;. At the same time, the average distance of fixations from the individual-unique target location in repeated and nonrepeated arrays were reliably shorter for the final vs. the initial fixations, $t(45) = 43.37$, $p < 0.001$, η^2 (partial) = 0.98, 95% CI: 0.96, 0.98. This indicates that scanpath similarity decreases as the eye homes in on physically different target locations in the various displays. Most importantly, our theoretical scanpath and empirical distance measures showed a correlation. To examine this, we calculated a scanpath-similarity difference measure by subtracting the similarity scores in the final from those in the initial scanpath parts; accordingly, higher (i.e., positive) difference values indicate higher similarity in the initial part. Likewise, we computed a difference measure for the average physical distance of initial and final fixations in the scanpath part from the target position (i.e., distance initial fixations minus distance final fixations); accordingly, higher (i.e., positive) values indicate larger target-fixation distances in the first part. These scanpath-difference and fixation-distance measures were computed for each individual participant and then correlated in the complete sample. We found significant positive correlations between the two measures for each scanpath metric, ranging from $r = 0.29$ (ABC, 95% CI: 0.002, 0.54, $p = 0.049$) over $r = 0.37$ (DFD; 95% CI: 0.092, 0.60, $p = 0.0011$) to $r = 0.56$ (DTW, 95% CI: 0.32, 0.73, $p < 0.001$). These findings indicate that our three scanpath-similarity measures are consistently precise in capturing basic

Table 1 The table presents summary and inference statistics for the four computed oculomotor measures of fixation number, scanpath length, average saccade amplitude, and standard deviation of saccade amplitude (the three latter in pixels).

Oculomotor measures	
Number of fixations	**Scanpath length**
Main effect of Context	Main effect of Context
Mean repeated = 8.44	Mean repeated = 1349.26
Mean nonrepeated = 9.62	Mean nonrepeated = 1611.78
F(1, 8701.8) = 136.62	F(1, 8702.3) = 158.04
$p < 0.001$	$p < 0.001$
η^2 (partial) = 0.02 (95% CI: 0.01, 0.02)	η^2 (partial) = 0.02 (95% CI: 0.01, 0.02)
Main effect of Block	Main effect of Block
$\text{Mean}_{\text{first block}} = 11.29$	$\text{Mean}_{\text{first block}} = 1903.37$
$\text{Mean}_{\text{last block}} = 8.36$	$\text{Mean}_{\text{last block}} = 1387.34$
F(31, 8701.7) = 6.73	F(31, 8702.1) = 5.50
$p < 0.001$	$p < 0.001$
η^2 (partial) = 0.02 (95% CI: 0.01, 0.03)	η^2 (partial) = 0.02 (95% CI: 0.01, 0.02)
Interaction Block x Context	Interaction Block x Context
F(31, 8702.3) = 1.04	F(31, 8703.0) = 1.00
$p = 0.41$	$p = 0.46$
η^2 (partial) = $3.69*10^{-3}$ (95% CI: 0.00, 0.00)	η^2 (partial) = $3.55*10^{-3}$ (95% CI: 0.00, 0.00)
BF = $7.74*10^{-55}$	BF = $4.43*10^{-55}$
Variance explained	Variance explained
$R^2_{\text{conditional}} = 0.13$ (95% CI: 0.11, 0.17)	$R^2_{\text{conditional}} = 0.10$ (95% CI: $6.8*10^{-2}$, 0.13)
$R^2_{\text{marginal}} = 3.7*10^{-2}$ (95% CI: $3.7*10^{-2}$, $5.1*10^{-2}$)	$R^2_{\text{marginal}} = 3.7*10^{-2}$ (95% CI: $3.6*10^{-2}$, $5.2*10^{-2}$)
Average saccade amplitude	**Standard deviation of saccade amplitude**
Main effect of Context	Main effect of Context
Mean repeated = 151.72	Mean repeated = 107.02
Mean nonrepeated = 160.01	Mean nonrepeated = 112.24
F(1, 8701.8) = 84.73	F(1, 8702.4) = 42.59
$p < 0.001$	$p < 0.001$
η^2 (partial) = $9.64*10^{-3}$ (95% CI: 0.01, 0.01)	η^2 (partial) = $4.87*10^{-3}$ (95% CI: 0.00, 0.00)
Main effect of Block	Main effect of Block
$\text{Mean}_{\text{first block}} = 162.55$	$\text{Mean}_{\text{first block}} = 118.17$
$\text{Mean}_{\text{last block}} = 155.63$	$\text{Mean}_{\text{last block}} = 109.95$
F(31, 8701.7) = 1.48	F(31, 8702.2) = 1.82
$p = 0.041$	$p = 0.0036$
η^2 (partial) = $5.24*10^{-3}$ (95% CI: 0.00, 0.00)	η^2 (partial) = $6.44*10^{-3}$ (95% CI: 0.00, 1.00)
Interaction Block x Context	Interaction Block x Context
F(31, 8702.3) = 0.97	F(31, 8703.2) = 0.88
$p = 0.52$	$p = 0.66$
η^2 (partial) = $3.44*10^{-3}$ (95% CI: 0.00, 0.00)	η^2 (partial) = $3.12*10^{-3}$ (95% CI: 0.00, 0.00)
BF = $2.52*10^{-55}$	BF = $6.36*10^{-56}$
Variance explained	Variance explained
$R^2_{\text{conditional}} = 0.12$ (95% CI: $9.9*10^{-2}$, 0.17)	$R^2_{\text{conditional}} = 8.2*10^{-2}$ (95% CI: $6.1*10^{-2}$, 0.12)
$R^2_{\text{marginal}} = 1.6*10^{-2}$ (95% CI: $1.6*10^{-2}$, $2.8*10^{-2}$)	$R^2_{\text{marginal}} = 1.3*10^{-2}$ (95% CI: $1.4*10^{-2}$, $2.5*10^{-2}$)

properties of serial visual search: they decrease in the later, efficient parts of the scanpath, as the eyes move nearer the different target regions – providing a validity check (vis-à-vis established effects) for our analysis approach.

Reporting summary. Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

To preview our results: (1) we replicate previous findings of more efficient visual search in terms of expedited RTs and reduced fixation number (and other established oculomotor measures) for repeated vs. nonrepeated search arrays. (2) Individual participants exhibit scanpath patterns that are increasingly similar across blocks of repeated and nonrepeated displays – both within displays and within participants. Repeated displays nevertheless exhibit a higher consistency within displays and within participants, reflected in a higher degree of scanpath similarity across repeated displays.

Reaction times. For the RT analyses, error trials and extreme RTs three standard deviations below and above the mean were excluded from the data. This outlier criterion led to the removal of ~3% of all trials. Overall, observers had an average error rate of ~1.5%, without any indication of a speed-accuracy trade-off. Regarding error rates, no main effect (context: F(1, 2810.2) = 0.13, $p = 0.72$, η^2 (partial) = $4.63e*10^{-5}$, 95% CI: 0.00, 0.00, BF = $1.97*10^{-81}$; block: F(31, 2811.0) = 1.10, $p = 0.32$, η^2 (partial) = $3.91e*10^{-4}$, 95% CI: 0.00, 0.00, BF = $9{,}97*10^{-164}$) nor interaction effect (F(31,2810.2) = 0.97, $p = 0.51$, η^2 (partial) = 0.01, 95% CI: 0.00, 0.01, BF = $5.55*10^{-78}$) reached significance.

The analysis of the mean RTs revealed a main effect of Context: participants responded faster to repeated relative to nonrepeated displays (1359 vs. 1574 ms, F(1, 8701.6) = 268.0207, $p < 0.001$, η^2 (partial) = 0.03, 95% CI: 0.02, 0.04). The main effect of Block was also significant, reflecting a decrease in RTs with increasing block number (block 1 = 1960 ms; block 32 = 1345 ms, F(31, 8701.5) = 13.00, $p < 0.001$, η^2 (partial) = 0.04 (95% CI: 0.03, 0.05). The Context × Block interaction was nonsignificant (F(31, 8701.9) = 0.88, $p = 0.67$, η^2 (partial) = $3.13*10^{-5}$ (95% CI: 0.00, 0.00), BF = $5.86*10^{-56}$), indicative of a stable contextual-facilitation effect across blocks (cf. Figure 1E and Supplementary Fig. 2), corresponding to an overall explained variance of $R^2_{\text{conditional}} = 0.19$ (95% CI: 0.16, 0.23) and, removing the random effects, an $R^2_{\text{marginal}} = 6.4*10^{-2}$ (95% CI: $5.9*10^{-2}$, $7.5*10^{-2}$), respectively.

Recognition performance. Participants' comparison of the hit rates against the 50% baseline – chance – performance yielded a nonsignificant result, t(45) = 1.09, $p = 0.28$, η^2 (partial) = 0.03 (95% CI: 0.00, 0.17), BF = 0.28. Thus, there was little indication of explicit, above-chance recognition of displays that had been encountered repeatedly over the course of the search task.

Oculomotor performance. Fig. 1F (see also Supplementary Fig. 3) presents a series of oculomotor measures, derived from fixations and saccades, as a function of block number aggregated over trials, separately for repeated and nonrepeated displays; Table 1 summarizes the respective descriptive and inference statistics when submitting the data to a linear mixed model with Block and Context as fixed factors and Participant and Target Quadrant as random factors. The upper left subplot of Fig. 1F illustrates the second classical finding: a decline in the number of fixations across blocks, with overall fewer fixations made in repeated vs. nonrepeated displays. The upper right plot of Fig. 1F depicts scanpath length. Since target positions were placed at equal distance from the screen center, the scanpath length also coincides with the so-called scan-pattern ratio[27]: the total distance (in pixels) that the eyes traversed across a given search display until arriving at the target location, divided by the shortest distance possible between the initial fixation and the target location – essentially quantifying how directly the eyes approached the target (see, e.g.[7],). This measure turned out significantly smaller for repeated compared to nonrepeated displays. Moreover, as can be seen in the lower left plot of Fig. 1F, repeated configurations showed a smaller mean saccade amplitude. Of note, the average standard deviation of the saccade amplitudes

Table 2 Summary inference statistics for the two levels of analysis – within display and, respectively, within participant – for the three scanpath similarity measures of Dynamic Time Warping (DTW), Discrete Fréchet Distance (DFD), and Area Between Curves (ABC).

Similarity within display

DTW
Main effect of Context
Mean $_{\text{repeated}} = 1613.82$
Mean $_{\text{nonrepeated}} = 1987.74$
$F(1,2760.0) = 154.62$
$p < 0.001$
η^2 (partial) $= 0.05$ (95% CI: 0.04, 0.07)
Main effect of Block
Mean $_{\text{first block}} = 2329.61$
Mean $_{\text{last block}} = 1657.38$
$F(1,2760.1) = 6.48$
$p < 0.001$
η^2 (partial) $= 2.34{*}10^{-3}$ (95% CI: 0.00, 0.01)
Interaction Block x Context
$F(1,2760.1) = 1.10$
$p = 0.32$
η^2 (partial) $= 3.98{*}10^{-4}$ (95% CI: 0.00, 0.00)
$\text{BF} = 4.12{*}10^{-33}$
Variance explained
$R^2_{\text{conditional}} = 0.50$ (95% CI: 0.46, 0.66)
R^2marginal $= 0.40$ (95% CI: 0.35, 0.54)
DFD
Main effect of Context
Mean $_{\text{repeated}}$: 291.7
Mean $_{\text{nonrepeated}}$: 314.64
$F(12760.0) = 14.90$
$p < 0.001$
η^2 (partial) $= 5.37{*}10^{-3}$ (95% CI: 0.00, 0.01)
Main effect of Block
Mean $_{\text{first block}}$: 337.8
Mean $_{\text{last block}}$: 292.78
$F(1,2760.1) = 2.11$
$p = 0.001$
η^2 (partial) $= 7.64{*}10^{-4}$ (95% CI: 0.00, 0.01)
Interaction Block x Context
$F(1,2760.0) = 0.71$
$p = 0.88$
η^2 (partial) $= 2.57{*}10^{-4}$ (95% CI: 0.00, 0.00)
$\text{BF} = 1.44{*}10^{-32}$
Variance explained
$R^2_{\text{conditional}} = 0.47$ (95% CI: 0.46, 0.66)
$R^2_{\text{marginal}} = 0.38$ (95% CI: 0.34, 0.57)
ABC
Main effect of Context
Mean $_{\text{repeated}} = 149860.73$
Mean $_{\text{non-repeated}} = 202522.02$
$F(1,2760.0) = 193.54$
$p < 0.001$
η^2 (partial) $= 0.07$ (95% CI: 0.05, 0.08)
Main effect of Block
Mean $_{\text{first block}} = 252733.5$
Mean $_{\text{last block}} = 153740.48$
$F(1,2760.1) = 5.80$
$p < 0.001$
η^2 (partial) $= 2.10{*}10^{-3}$ (95% CI: 0.00, 0.01)
Interaction Block x Context
$F(1,2760.0) = 1.06$
$p = 0.37$
η^2 (partial) $= 3.84{*}10^{-4}$ (95% CI: 0.00, 0.00)
$\text{BF} = 7.64{*}10^{-33}$
Variance explained
$R^2_{\text{conditional}} = 0.52$ (95% CI: 0.49, 0.69)
$R^2_{\text{marginal}} = 0.41$ (95% CI: 0.33, 0.57)

was also significantly reduced for repeated vs. nonrepeated displays – see lower right plot of Fig. 1f. For all measures, the block x context interaction was not significant (see Table 1).

Scanpath analysis. In a first, qualitative analysis, we visualized scanpaths across blocks and target positions. In more detail, we plotted the scanpath representations for each of the four target positions in repeated and, respectively, nonrepeated displays (i.e., the four repeated and nonrepeated trials within a block). As can be seen from Fig. 1A (see also Supplementary Fig. 4), in the last block of the search task – after ample opportunity for contextual adaptation – there was a higher degree of similarity between scanpaths for repeated relative to nonrepeated contexts. Specifically, in the repeated condition for each display, saccades tended to be more often and more closely executed in parallel direction across observers, indicative of systematic biases in observers' oculomotor behavior. Also, in the repeated compared to the nonrepeated condition, the color of the lines connecting successive fixation locations (with green denoting the initial and blue the final saccade on a given trial, and intermediate colors denoting saccades in between) tended to be more clustered spatially, indicating that saccades were executed in a more systematic order as well.

Similarity analysis. The main finding of this study is that scanpath similarity is increasing throughout the experiments, with repeated displays gaining a significant advantage early on which remains throughout the experiment. These findings are indicative of the convergence towards an optimal search strategy on the level of the set of displays (see discussion). For both analyses, within displays (Table 2) and within participants (Table 3) respectively (also, see Fig. 1G and Supplementary Fig. 5), we found significant main effects of Context and Block, while the Context × Block interactions were nonsignificant. The pattern reflects a steady increase in scanpath similarity (corresponding to smaller numerical values in Dynamic Time Warping, Discrete Fréchet Distance, and Area Between Curves) over the course of the experiment (main effect of Block), which was however higher in repeated vs. nonrepeated displays (main effect of Context) with a stable context effect emerging early on (no significant interaction). Of note, the three similarity measures yielded qualitatively similar results, despite being sensitive to slightly different aspects of the scanpaths, attesting to a high reliability of our findings.

Thus, our scanpath similarity measures provide strong support for a procedural-optimization hypothesis, according to which participants, over time on task, learn a generic search procedure that is increasingly effective for all – repeated and nonrepeated – displays, though repeated displays weigh more highly in the optimization as a result of being searched more often.

Discussion

To gain an in-depth understanding of contextual facilitation, we analyzed established measures of the contextual-facilitation effect that focus on aggregate oculomotor indices, as well as new measures based on spatio-temporal scanpath sequences. Concerning individual eye-movement patterns: replicating prior reports, we found that detecting a target involves fewer fixations in repeated compared to nonrepeated target-distractor arrangements (e.g. ref. [5,6,43],), as well as a shorter scanpath length and, accordingly, a smaller scan-pattern ratio (ref. [7,27]). We also found the saccade amplitudes to be significantly shorter for repeated displays (in contrast to[6]). In addition, we established a aggregate oculomotor index of contextual facilitation (that hitherto had not been reported in the literature): a reduced standard deviation of the saccade amplitudes for repeated vs. nonrepeated displays.

Contextual-cueing of visual search: general procedural guidance of the eyes. Having established comparability of the present findings with those reported in prior studies of contextual cueing, we went on to examine the oculomotor scanpaths in order to differentiate between a template-based, display-specific and a procedural, display-generic scanning account of acquired contextual facilitation that may drive the gains in the aggregate eye-

Table 3 Summary inference statistics for the two levels of analysis – within display and, respectively, within participant – for the three scanpath similarity measures of Dynamic Time Warping (DTW), Discrete Fréchet Distance (DFD), and Area Between Curves (ABC).

Similarity within participant

DTW
Main effect of Context
$\text{Mean}_{\text{repeated}} = 2247.55$
$\text{Mean}_{\text{nonrepeated}} = 2628.75$
$F(1,189) = 99.32$
$p < 0.001$
η^2 (partial) = 0.34 (95% CI: 0.24, 0.44)
Main effect of Block
$\text{Mean}_{\text{first block}} = 3119.54$
$\text{Mean}_{\text{last block}} = 2180.57$
$F(31,189) = 2.80$
$p < 0.001$
η^2 (partial) = 0.31 (95% CI: 0.11, 0.34)
Interaction Block x Context
$F(31,189) = 0.58$
$p = 0.96$
η^2 (partial) = 0.09 (95% CI: 0.00, 0.01)
$\text{BF} = 8.73{*}10^{-47}$
Variance explained
$R^2_{\text{conditional}} = 0.34$ (95% CI: 0.29, 0.43)
$R^2_{\text{marginal}} = 8.9{*}10^{-2}$ (95% CI: $8.5{*}10^{-2}$, 0.12)
DFD
Main effect of Context
$\text{Mean}_{\text{repeated}}$: 419.14
$\text{Mean}_{\text{nonrepeated}}$: 427.24
$F(1,189) = 91.22$
$p < 0.001$
η^2 (partial) = 0.33 (95% CI: 0.22, 0.42)
Main effect of Block
$\text{Mean}_{\text{first block}}$: 458.01
$\text{Mean}_{\text{last block}}$: 415.03
$F(31,189) = 2.84$
$p < 0.001$
η^2 (partial) = 0.32 (95% CI: 0.12, 0.35)
Interaction Block x Context
$F(31,189) = 0.64$
$p = 0.93$
η^2 (partial) = 0.10 (95% CI: 0.00, 0.03)
$\text{BF} = 1.81{*}10^{-49}$
Variance explained
$R^2_{\text{conditional}} = 0.28$ (95% CI: 0.21, 0.36)
$R^2_{\text{marginal}} = 2.6{*}10^{-2}$ (95% CI: $3.2{*}10^{-2}$, $5.2{*}10^{-2}$)
ABC
Main effect of Context
$\text{Mean}_{\text{repeated}}$: 188,744.43
$\text{Mean}_{\text{nonrepeated}}$: 242078.25
$F(1,189) = 107.64$
$p < 0.001$
η^2 (partial) = 0.36 (95% CI: 0.26, 0.46)
Main effect of Block
$\text{Mean}_{\text{first block}}$: 297143.1
$\text{Mean}_{\text{last block}}$: 190412.32
$F(31,189) = 2.87$
$p < 0.001$
η^2 (partial) = 0.01 (95% CI: 0.00, 0.07)
Interaction Block x Context
$F(31,189) = 0.61$
$p = 0.95$
η^2 (partial) = $3.22{*}10^{-3}$ (95% CI: 0.00, 0.04)
$\text{BF} = 4.72{*}10^{-47}$
Variance explained
$R^2_{\text{conditional}} = 0.30$ (95% CI: 0.26, 0.37)
$R^2_{\text{marginal}} = 0.10$ (95% CI: $0.2{*}10^{-2}$, 0.13)

movement indices (such as the reduced total fixation number) for repeated displays. Using two distinct approaches – of comparing eye movement sequences between pairs of identical displays when viewed by different participants, and, respectively, individual participants viewing different displays – and three metrics of scanpath similarity (Dynamic Time Warping, Discrete Fréchet Distance, and Area Between Curves), we found that, while the consistency of the scanpaths increased overall with increasing time on task for both repeated and nonrepeated displays, these practice-dependent gains were more strongly driven by displays sampled repeatedly (vs. displays sampled only once). This effect pattern is indicative of common regularities shared between the (statistical) sample of search displays, which influence scanpaths in a way that is independent of the particular arrangement encountered on a trial or even the individual participant – akin to generic, one-for-all procedural learning. Had there been display-specific learning of individual repeated displays, the similarity measures obtained from any two such displays should have been reduced, relative to nonrepeated (baseline) displays, as each repeated array should have come to elicit its unique scanning pattern. However, at variance with this prediction from display-specific learning accounts of contextual cueing, we found the similarity measures to be actually different.

Of note, this pattern of scanpath-similarity effects does not rule out that the scanpaths become tuned to specific displays at some point along their progression, for instance, when the eye finally homes in on the target location;[6] nor do we take this to rule out the possibility of display-specific learning under all circumstances (considered further below). Rather, we take our findings to demonstrate that, over the course of a hard search task, the notion of display-general learning provides an apt account of contextual facilitation. In line with this[6]; (see also ref. [8,39]) have proposed that the (efficient) phase, in which the eyes come closer to the target with each successive fixation, does not differ significantly between repeated and nonrepeated displays; but the two display types differ with respect to the number of fixations in the preceding inefficient phase (with reduced fixation numbers in repeated displays). However, in contrast to Tseng and Li[6], our scanpath-similarity analyses, which take into account the entire spatio-temporal series of fixation events, show that eye movements in the so-called ineffective phase are, in fact, not (in terms of Tseng & Li[6] p. 1371) "wasted". Instead, our findings of increasing scanpath homogeneity with extended time on task suggest that what may look an ineffective phase actually constitutes an important period during which procedural learning of a general scanning scheme becomes functional.

In terms of skill acquisition[18], when participants perform a new task of searching displays with novel, as yet to-be-discovered statistical properties, one would expect that they learn to adapt and optimize their oculomotor scanning behavior with respect to the display statistics in rather generic terms – as opposed to acquiring search-guiding memory representations tailored to specific, individual display arrangements, as assumed by standard accounts of contextual cueing. These scanning strategies are optimal in the sense that they save cognitive effort: they do not require memorizing arbitrary distractor-target configurations and expensively checking a given display arrangement against a set of representations in contextual memory. As an unavoidable side effect, developing a strategy that is optimally adjusted to the statistical search environment at large (with repeated displays having a greater weight in shaping this strategy due to their increased frequency of occurrence) would also optimize the scanning of nonrepeated displays (which, due to their random variation, have little weight in determining the strategy). Finally, based on evolutionary considerations, learning the overall statistics of a scene environment would not only be more efficient, but also be more robust to environmental changes (in old scenes as well as the addition of new repeated scenes, which could be more easily incorporated in an environment-generic strategy) compared to learning highly specific display-target configurations.

Consistent with these functional considerations, there is also evidence from an fMRI study (including eye movements), by Manelis and Reder[44], in line with a procedural-learning account of contextual cueing: When comparing the first with the last epoch of the experiment, Manelis and Reder[44] found a significant decrease in functional connectivity between hippocampus and sensory-procedural areas (which they did, however, not expressly attribute to mechanisms related to procedural oculomotor learning). Of note, hippocampal activity is not only involved in motor-sequence learning (e.g., in finger-tapping tasks[45]), but also in other statistical learning tasks, such as serial RT tasks (e.g., ref. [46,47],), particularly in the initial stages, with a decrease in activity later on[48]. This is consistent with a critical, but (over the course of practice) diminishing role of the hippocampus in procedural motor learning and would go some way to explain why patients with damage to the hippocampus do not display a contextual-cueing benefit[49]. Thus, a procedural-learning account involving the hippocampus could explain the phenomenon of contextual cueing without assuming the acquisition of display-specific contextual representations.

Such a display-generic, procedural account of oculomotor contextual cueing would also explain why participants usually do not explicitly recognize repeatedly encountered display arrangements:[1] hippocampal involvement would revolve around procedural aspects of search performance, rather than explicit (episodic) memory. This view is not necessarily incompatible with existing functional accounts of long-term memory, according to which hippocampus contributes to the formation of inter-element associations: for the present investigation, these associations would involve the binding of individual eye fixation – thus making procedural memory a specific instance of a more general, associative hippocampal memory system (e.g., ref. [49,50],).

Limitations. While our results suggest that display-unspecific, procedural learning of saccadic trajectories plays an important role in oculomotor contextual facilitation, it is important to note that if oculomotor scanning is allowed or encouraged, these developing strategies might look different from conditions in which observers are instructed to search the display without eye movements (e.g., ref. [51],). In line with this, electrophysiological (EEG) studies report evidence of display-specific contextual cueing – more precisely: target-side-specific lateralizations of event-related potentials indicative of attentional resource allocation – as early as around 200 ms post display onset[52] (see also ref. [53]). Arguably, disallowing eye movements would impede the evolution of generic scanning procedures, in particular, when brief exposure times prevent extended search. In contrast, more natural scenarios that require/allow oculomotor scanning foster the acquisition of display-unspecific routines adapted to the statistical regularities in the set of search displays encountered, which is dominated by repeated display arrangements. A somewhat related idea is that participants may have acquired specific memory representations of the display arrangement, but do not use them when they can scan the display freely[54]. The ERP task design, by contrast, likely forces display-specific learning, perhaps due to the need of holding individual display arrangements in working memory in order to solve the task[55]. Thus, there might be dual mechanisms underlying contextual facilitation, with the relative dominance of the display-generic over the display-specific mechanism being determined by the task demands, that is, the extent to which eye movements are possible/required or discouraged/dis-allowed.

Conclusions

Contextual cueing is an important predictive-coding mechanism characterized by facilitation of search performance in repeated search arrays. As such, this facilitatory effect can be accommodated equally by accounts assuming associative learning of individual target locations in individual distractor arrangements (and the reproduction of individual scanpaths for these arrays) or, respectively, the acquisition of generic oculomotor patterns that optimize the scanning of the (for the participants initially overall new) set of search displays. The current study was designed to test the latter (display-generic) against the former (display-specific) account, by systematically investigating participants' eye-movement trajectories in repeated displays and comparing them against nonrepeated displays. Our findings are in line with a display-general scanning account of contextual facilitation: over time on task, scanpath sequences became increasingly similar across participants and displays, with total scanpath similarity being higher for repeated displays. We propose that at least under natural search conditions, contextual facilitation largely or exclusively derives from the acquisition of procedural oculomotor scanning programs, which become operational quite early during a given search trial. Conceptualizing contextual cueing as procedurally optimized oculomotor trajectories also offers new ways (1) for understanding the difficulty to update established contextual memories following consistent target-position changes within otherwise unchanged distractor arrangements (ref. [1], Experiment 3, ref. [7,56,57]); as well as (2) for understanding the apparent high capacity of contextual memory (see, e.g., ref. [58]; after being presented with 12 different target-distractor arrangements per day over a 5-day period, Jiang et al.[58]'s participants showed contextual facilitation for the total number of 60 arrangements when tested at the end); and (3), it provides a possible explanation as to why contextual cueing leads to fMRI activations in sensory brain areas contributing to procedural learning (e.g., ref. [44]; see also ref. [59], for confirmatory evidence using MEG). Moreover, a display-generic account of contextual learning would provide a coherent and, in terms of Occam's razor, the simplest explanation: it explains the search advantage for repeated versus nonrepeated displays, as well as the practice-dependent improvement of search in novel displays[1] in terms of procedural learning or skill acquisition[18]. Finally, given our evidence that oculomotor search is optimized independently of a particular configuration as a skill of performing a visual search task in general, we propose to use the more neutral term contextual facilitation (instead of display-specific contextual cueing) to describe the effects of procedural spatial learning in visual search. Repeated displays merely have a stronger influence in the optimization process, bringing about the facilitation effect.

Data availability

The raw data that support the analysis and results are publicly accessible at https://osf.io/snjpk/.

Code availability

The analysis code is available at https://osf.io/snjpk/.

Received: 7 February 2023; Accepted: 21 August 2023;

References

1. Chun, M. M. & Jiang, Y. Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognit. Psychol.* **36**, 28–71 (1998).
2. Moran, R., Zehetleitner, M., Liesefeld, H. R., Müller, H. J. & Usher, M. Serial vs. parallel models of attention in visual search: accounting for benchmark RT-distributions. *Psychonom. Bull. Rev.* **23**, 1300–1315 (2016).
3. Geyer, T., Zehetleitner, M. & Müller, H. J. Contextual cueing of pop-out visual search: When context guides the deployment of attention. *J. Vision* **10**, 20 (2010).

4. Iiang, Y. & Chun, M. M. Contextual cueing Reciprocal influences between. *Attention and Implicit Learning*. **48**, 277 (2003).
5. Peterson, M. S. & Kramer, A. F. Attentional guidance of the eyes by contextual information and abrupt onsets. *Percept. Psychophys.* **63**, 1239–1249 (2001).
6. Tseng, Y.-C. & Li, C.-S. R. Oculomotor correlates of context-guided learning in visual search. *Percept. Psychophys.* **66**, 1363–1378 (2004).
7. Manginelli, A. A. & Pollmann, S. Misleading contextual cues: How do they affect visual search? *Psychol. Res.* **73**, 212–221 (2009).
8. Kroell, L. M., Schlagbauer, B., Zinchenko, A., Müller, H. J. & Geyer, T. Behavioural evidence for a single memory system in contextual cueing. *Visual Cognit.* **27**, 551–562 (2019).
9. Harris, A. M. & Remington, R. W. Contextual cueing improves attentional guidance, even when guidance is supposedly optimal. *J. Exp. Psychol. Human Percept. Perform.* **43**, 926 (2017).
10. Brady, T. F. & Chun, M. M. Spatial constraints on learning in visual search: modeling contextual cuing. *J. Exp. Psychol. Human Percept. Perform.* **33**, 798 (2007).
11. Beesley, T., Vadillo, M. A., Pearson, D. & Shanks, D. R. Configural learning in contextual cuing of visual search. *J. Exp. Psychol. Human Percept. Perform.* **42**, 1173 (2016).
12. Wolfe, J. M., Cave, K. R. & Franzel, S. L. Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Human Percept. Perform.* **15**, 419 (1989).
13. Wolfe, J. M. Guided search 2.0 a revised model of visual search. *Psychonom. Bull. Rev.* **1**, 202–238 (1994).
14. Wolfe, J. M. Guided Search 6.0: An updated model of visual search. *Psychonom. Bull. Rev.* 1–33 (2021).
15. Goujon, A., Didierjean, A. & Thorpe, S. Investigating implicit statistical learning mechanisms through contextual cueing. *Trend Cognit. Sci.* **19**, 524–533 (2015).
16. Sisk, C. A., Remington, R. W. & Jiang, Y. V. Mechanisms of contextual cueing: A tutorial review. *Attent. Percept. Psychophys.* **81**, 2571–2589 (2019).
17. Fitts, P. M. Perceptual-motor skill learning. In Categories of human learning (pp. 243-285). Academic Press (1964).
18. Fitts, P. M. & Posner, M. I. Human performance. Brooks. Cole, Belmont, CA, 5, 7-16 (1967).
19. Shiffrin, R. M. & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychol. Rev.* **84**, 127 (1977).
20. Anderson, J. R. Acquisition of cognitive skill. *Psychol. Rev.* **89**, 369 (1982).
21. Vadillo, M. A., Konstantinidis, E. & Shanks, D. R. Underpowered samples, false negatives, and unconscious learning. *Psychonom. Bull. Rev.* **23**, 87–102 (2016).
22. Peterson, M., Mead, P., Kelly, S., Esser-Adomako, E. & Blumberg, E. Spatial context target relearning following a target relocation event: Not mission impossible. *Attent. Percept. Psychophys.* **84**, 474–489 (2022).
23. Smyth, A. C. & Shanks, D. R. Awareness in contextual cuing with extended and concurrent explicit tests. *Memory Cognit.* **36**, 403–415 (2008).
24. Geyer, T., Rostami, P., Sogerer, L., Schlagbauer, B. & Müller, H. J. Task-based memory systems in contextual-cueing of visual search and explicit recognition. *Sci. Rep.* **10**, 1–13 (2020).
25. Jeffreys, H. *Theory of probability* (3rd ed.). Oxford: Oxford University Press (1961).
26. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
27. Brockmole, J. R. & Henderson, J. M. Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *Quart. J. Exp. Psychol.* **59**, 1177–1187 (2006).
28. Myers, C. W. & Gray, W. D. Visual scan adaptation during repeated visual search. *J. Vision* **10**, 4 (2010).
29. Sewell, D. K., Colagiuri, B. & Livesey, E. J. Response time modeling reveals multiple contextual cuing mechanisms. *Psychonom. Bull. Rev.* **25**, 1644–1665 (2018).
30. Brainard, D. H. & Vision, S. The psychophysics toolbox. *Spat. Vision* **10**, 433–436 (1997).
31. Pelli, D. G. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat. Vision* **10**, 437–442 (1997).
32. Shaw, M. L. & Shaw, P. Optimal allocation of cognitive resources to spatial locations. *J. Exp. Psychol. Human Percept. Perform.* **3**, 201 (1977).
33. Geng, J. J. & Behrmann, M. Spatial probability as an attentional cue in visual search. *Attent. Percept. Psychophys.* **67**, 1252–1268 (2005).
34. Van Rossum, G. & Drake, F. L. Python reference manual. Vol. 111, 1-52 (Amsterdam: Centrum voor Wiskunde en Informatica, 1995).
35. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/ (2018).
36. Geringswald, F., Herbik, A., Hoffmann, M. B. & Pollmann, S. Contextual cueing impairment in patients with age-related macular degeneration. *J. Vision* **13**, 28 (2013).
37. Olson, I. R. & Chun, M. M. Perceptual constraints on implicit learning of spatial context. *Vis. Cognit.* **9**, 273–302 (2002).
38. Nakagawa, S., Johnson, P. C. D. & Schielzeth, H. The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J. Royal Soc. Interf.* **14**, 20170213 (2017).
39. Pollmann, S. & Manginelli, A. A. Anterior prefrontal involvement in implicit contextual change detection. *Front. Human Neurosci.* **3**, 28 (2009).
40. Logan, G. D. Toward an instance theory of automatization. *Psychol. Rev.* **95**, 492 (1988).
41. Fahimi, R. & Bruce, N. D. On metrics for measuring scanpath similarity. *Behav. Res. Methods* **53**, 609–628 (2021).
42. Jekel, C. F., Venter, G., Venter, M. P., Stander, N. & Haftka, R. T. Similarity measures for identifying material parameters from hysteresis loops using inverse analysis. *Int. J. Mater. Form.* **12**, 355–378 (2019).
43. Zang, X., Jia, L., Müller, H. J. & Shi, Z. Invariant spatial context is learned but not retrieved in gaze-contingent tunnel-view search. *J. Exp. Psychol. Learn. Memory Cognit.* **41**, 807 (2015).
44. Manelis, A. & Reder, L. M. Procedural learning and associative memory mechanisms contribute to contextual cueing: Evidence from fMRI and eye-tracking. *Learn. Memory* **19**, 527–534 (2012).
45. Albouy, G., King, B. R., Maquet, P. & Doyon, J. Hippocampus and striatum: Dynamics and interaction during acquisition and sleep-related motor sequence memory consolidation. *Hippocampus* **23**, 985–1004 (2013).
46. Hazeltine, E., Grafton, S. T. & Ivry, R. Attention and stimulus characteristics determine the locus of motor-sequence encoding. A PET study. *Brain: A J. Neurol.* **120**, 123–140 (1997).
47. Müller, R. A., Kleinhans, N., Pierce, K., Kemmotsu, N. & Courchesne, E. Functional MRI of motor sequence acquisition: effects of learning stage and performance. *Cognit. Brain Res.* **14**, 277–293 (2002).
48. Albouy, G. et al. Both the hippocampus and striatum are involved in consolidation of motor sequence memory. *Neuron* **58**, 261–272 (2008).
49. Chun, M. M. & Phelps, E. A. Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nat. Neurosci.* **2**, 844–847 (1999).
50. Henke, K. A model for memory systems based on processing modes rather than consciousness. *Nat. Rev. Neurosci.* **11**, 523–532 (2010).
51. Lleras, A. & Von Muhlenen, A. Spatial context and top-down strategies in visual search. *Spatial Vision* **17**, 465–482 (2004).
52. Zinchenko, A., Conci, M., Töllner, T., Müller, H. J. & Geyer, T. Automatic guidance (and misguidance) of visuospatial attention by acquired scene memory: evidence from an N1pc polarity reversal. *Psychological Sci.* **31**, 1531–1543 (2020).
53. Schankin, A. & Schubö, A. Contextual cueing effects despite spatially cued target locations. *Psychophysiology* **47**, 717–727 (2010).
54. Võ, M. L. H. & Wolfe, J. M. When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *J. Exp. Psychol. Human Percept. Perform.* **38**, 23–41 (2012).
55. Ballard, D. H., Hayhoe, M. M. & Pelz, J. B. Memory representations in natural tasks. *J. Cognit. Neurosci.* **7**, 66–80 (1995).
56. Geyer, T., Seitz, W., Zinchenko, A., Müller, H. J. & Conci, M. Why are acquired search-guiding context memories resistant to updating? *Front. Psychol.* **12**, 564 (2021).
57. Zellin, M., von Mühlenen, A., Müller, H. J. & Conci, M. Statistical learning in the past modulates contextual cueing in the future. *J.Vision* **13**, 19–19 (2013).
58. Jiang, Y., Song, J. H. & Rigas, A. High-capacity spatial contextual memory. *Psychonom. Bull. Rev.* **12**, 524–529 (2005).
59. Chaumon, M., Schwartz, D. & Tallon-Baudry, C. Unconscious learning versus visual perception: dissociable roles for gamma oscillations revealed in MEG. *J. Cognit. Neurosci.* **21**, 2287–2299 (2009).

Acknowledgements

This research was supported by grants of the German Research Council (DFG) to Thomas Geyer (GE 1889/4-2 & RTG 2175). The funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

WS and AZ conceived and designed the experiments; AZ performed the experiments: WS analyzed the data; HJM contributed to conceptual analysis and discussion; WS, HJM, TG wrote the paper.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44271-023-00019-8.

Correspondence and requests for materials should be addressed to Werner Seitz.

Peer review information *Communications Psychology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Antonia Eisenkoeck. A peer review file is available

Study 2 – Seitz et al. (2024)

Seitz, W., Zinchenko, A., Müller, H. J., & Geyer, T. (2024). Distraction from long-term memory in visual search: Acquired procedural and template-based memory interfere with contextual cueing after target re-location. *Visual Cognition*, 1-22.

Author Contributions

WS and AZ conceived and designed the experiments; AZ performed the experiments; **WS** analyzed the data; HJM contributed to conceptual analysis and discussion; **WS**, HJM, TG wrote the paper.

Visual Cognition

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pvis20

Distraction from long-term memory in visual search: Acquired procedural and template-based memory interfere with contextual cueing after target re-location

Werner Seitz, Artyom Zinchenko, Hermann J. Müller & Thomas Geyer

To cite this article: Werner Seitz, Artyom Zinchenko, Hermann J. Müller & Thomas Geyer (23 Aug 2024): Distraction from long-term memory in visual search: Acquired procedural and template-based memory interfere with contextual cueing after target re-location, Visual Cognition, DOI: 10.1080/13506285.2024.2315786

To link to this article: https://doi.org/10.1080/13506285.2024.2315786

VISUAL COGNITION
https://doi.org/10.1080/13506285.2024.2315786

Routledge
Taylor & Francis Group

Distraction from long-term memory in visual search: Acquired procedural and template-based memory interfere with contextual cueing after target re-location

Werner Seitz [a], Artyom Zinchenko[a], Hermann J. Müller[a,b] and Thomas Geyer[a,b,c]

[a]Department Psychologie, Ludwig-Maximilians-Universität München, München, Germany; [b]Munich Center for Neurosciences – Brain & Mind, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany; [c]NICUM – Neuroimaging Core Unit Munich, Ludwig-Maximilians-Universität München, München, Germany

ABSTRACT
Participants can learn to faster detect targets embedded in repeatedly encountered spatial arrangements of distractors – termed the "contextual cueing" (CC) effect. However, cueing is severely compromised following target relocation in an otherwise unchanged distractor arrangement. Previous research demonstrated that this re-location cost is due to persistent misguidance of search towards the original location. Since CC reflects top-down guidance of contextual memory, this misguidance is an instance of a "distraction" effect that operates from acquired memory, rather than being driven by salient but irrelevant stimuli in the display. While traditional accounts of CC emphasize the acquisition of search-guiding memory "templates" specific to particular displays, contextual learning also tunes attentional (oculomotor) scanning routines to the overall statistics of the display arrangements, yielding a context-unspecific LT "proceduralization" of search. Using reaction-time and oculomotor-scanning measures, we confirmed both mechanisms to contribute to initial contextual learning as well as the "distraction" effect produced by relocating the targets of repeated displays. We suspect that guidance and misguidance of search by repeated contexts involve two complementary LT mechanisms: procedural optimization of broad, i.e., display-generic, scanning routines, and learning of where to expect the target in specific repeated-context displays.

ARTICLE HISTORY
Received 5 May 2023
Accepted 13 December 2023

KEYWORDS
Visual search; statistical learning; contextual cueing; attentional capture; oculomotor search

Highlights

- Visual search is facilitated by statistical learning of the spatial target-distractor relations ("contexts") in repeatedly encountered display layouts, due to more effective eye guidance towards repeated target locations
- This contextual-learning effect is associated with a cost following consistent relocation of the target to the opposite hemifield in an otherwise unchanged (repeated) distractor layout
- This relocation cost results from an acquired attentional bias to preferentially scan displays in left-right direction, in addition to ongoing, top-down attentional prioritization of the originally learnt target locations
- This suggests that search experience with repeated target-distractor contexts trains *both* display-general procedures about how to optimally scan the search arrays *and* display-specific representations ("templates") of repeated target-context relations

The present study is concerned with an unusual type of "distraction" of visual search: mis-direction of search processes by different types of long-term (LT) memory acquired as a result of repeatedly searching for a target predictably positioned within a constant spatial array of non-target, or distractor, items. The "distraction", or interference, happens when, after learning these spatial target-distractor relations, the target is consistently repositioned to a different location within the same distractor array (or distractor "context"). On the surface, this interference effect appears similar to attentional capture by a salient but task-irrelevant distractor in the search display (e.g., Chelazzi et al., 2019) – just that the displays do not contain any such item that could divert search

CONTACT Werner Seitz seitz-werner@hotmail.de Department Psychologie, Lehrstuhl für Allgemeine und Experimentelle Psychologie, Ludwig-Maximilians-Universität München, Leopoldstraße 13, 80802 München, Germany

away from the new target location.[1] Instead, the source of the interference must lie in the activation of search-guiding LT-memory mechanisms, whose predictions about the location of the target or, respectively, procedures of how to find it most efficiently are no longer applicable. The present study was designed to investigate this phenomenon of LT-memory generated distraction in more detail, the aim being to elucidate the LT-memory mechanisms that bring about the interference. – To render the background of our study, in the next two sections, we provide a brief review of the phenomenon of "contextual cueing" of visual search and the dominant and emerging conceptual models proposed to account for it. This informs the rationale of our study, which we develop in the third section.

Contextual cueing

Visual search is facilitated, that is, search reaction times (RTs) are expedited, when the searched-for target item is consistently located within a stable configuration, or *context*, of non-target, or *distractor*, elements. There are two accounts of the RT benefits deriving from experience with repeated displays. One assumes that display repetitions facilitate response-selection and/ or motor-execution stages of processing, when participants make a decision about which motor (hand) effector is required for a correct response (e.g., Hout & Goldinger, 2010; Kunar et al., 2007; Schankin & Schubö, 2010). The other, attentional account assumes that contextual cueing arises because the acquired target-distractor spatial associations (stored in long-term memory, LTM) come to guide search, predicting or "cueing attention to" the target location (e.g., Chun & Jiang, 1998; Geyer et al., 2010; for reviews, see, e.g., Goujon et al., 2015; or Sisk et al., 2019). Note that in visual search, statistical contingencies can be based also on other environmental properties, such as learning to associate the repeated target location with constant background colours, textures, or scene backgrounds (e.g., Kunar et al., 2014; Rosenbaum & Jiang, 2013). Further, if the target location is allowed to vary across trials in a nevertheless constant distractor array, participants can learn to ignore the repeated distractor locations, which leads to RT facilitation even with unpredictable targets (though the effect is smaller than the "standard" CC effect; e.g., Kunar & Wolfe, 2011; Vadillo et al., 2021). – In the present study, we examine statistical learning under "standard" conditions when both the target location and the spatial layout formed by the non-target, distractor, elements are repeated across trials (as in Chun & Jiang's, 1998, original study). Specific emphasis is given to participants' overt scanning strategies and how experience with the repeated displays modifies these strategies. We distinguish between template-based and procedural LT memories, that is, learning about where and how to focus visual search.

In more detail, we examine the attentional account of contextual cueing (CC) using eye-fixation measures that index the focus of attention, with emphasis on eye movements following an unexpected change of the target location in an otherwise unchanged distractor arrangement. As is known from the literature: while the initial acquisition of context-target mappings requires just a few (~3) repetitions of each individual search display, changes of the target location from a learnt target position in one half of the display to a different position in the other half, in an otherwise unchanged distractor layout, greatly diminishes the CC effect, and it takes ~80 repetitions of each repeated display for contextual-cueing to be fully re-established (Zellin et al., 2014; see also Yang et al., 2021). Thus, while contextual regularities are acquired fast and efficiently, established context memories are rather inflexible: they resist updating, and prolonged practice is required for a new target location to be integrated within an existing representation (Conci & Zellin, 2022; Makovski & Jiang, 2010; Manginelli & Pollmann, 2009; Zellin et al., 2013; see also Geyer et al., 2021). The loss of CC after target location changes is unlikely due to restrictions in memory capacity (Jiang et al., 2005) or general limitations in acquiring invariant contextual information (Brockmole & Henderson, 2006; Brockmole & Le-Hoa Võ, 2010; Jiang & Wagner, 2004; Zang et al., 2017). However, the previous target re-location studies were mainly concerned with the adaptability of acquired LTM representations per se (using RT summary measures of CC). Accordingly, they tell us little about what the information is that is extracted from experience with repeated target-distractor arrangements and represented in contextual LTM, which needs to be updated following target relocation within otherwise unchanged contexts.

The attentional account of contextual learning assumes that display repetitions lead to the build-up of representations, or *templates*, in LTM, likely located in the medial temporal lobes (MTL; Chun & Phelps, 1999; Manns & Squire, 2001), of *specific* display arrays. When re-encountering such a display on later occasions, the corresponding template is activated and top-down guides search towards the target location (e.g., Chun & Jiang, 1998). For instance, activated contextual templates may raise the priority of the target location on the search-guiding "attentional-priority map" (e.g., Fecteau & Munoz, 2006) – complementing any bottom-up-/stimulus- or top-down-/knowledge-driven guidance signals. Accordingly, the loss of contextual cueing after the re-location of the target would be attributable to *LTM-induced distraction*: attention continues to be misguided to the old target location, until this persistent bias can be overcome by incorporating the new target position in the originally acquired contextual template (e.g., Zinchenko et al., 2020a).

However, at odds with an exclusive role of MTL mechanisms, evidence from imaging and patient studies suggests that statistical context learning induces training-related changes also in visuo-perceptual regions directly involved in encoding the search arrays (e.g., Chaumon et al., 2009; Manelis & Reder, 2012). From a theoretical perspective, these findings lend support to process-based accounts of the acquisition and automatization of search skills (e.g., Fitts, 1964; Fitts & Posner, 1967; Shiffrin & Schneider, 1977), according to which statistical learning directly optimizes the processes by which (visual-search) tasks are actually accomplished, with repeatedly encountered search displays eventually accruing a greater weight in shaping these processes than those encountered only once (e.g., Seitz et al., 2023).

The aim of the present study was to investigate the interplay of these two mechanisms of contextual facilitation – that is, acquired context-based guidance and procedural optimization of search – in repeated search arrays, with a particular focus on how these processes relate to *attentional misguidance* – or *memory-based distraction* – following target re-location within an otherwise unchanged distractor context. This is an important question as the available studies agree on the idea that CC is a form of top-down scene guidance of attention (e.g., Wolfe, 2020).

Contextual cueing as skill acquisition

Searching a repeated display several times triggers learning about where to find the search target, leading to the acquisition of contextual search templates in LTM (e.g., Ort & Olivers, 2020) that point to the target position within the spatial context of distractor items. Activation of such a contextual template by a given (repeated) search display renders target localization (and, thus, responses to the target) more efficient compared non-repeated (never before encountered) displays – constituting the contextual-cueing effect (Chun & Jiang, 1998). Typically, in studies of contextual cueing, the target is a (left- or right-oriented) T-shaped letter embedded in an array of (orthogonally oriented) L-shaped distractor letters – a hard task requiring serial scanning. Expedited localization of such a target in repeated search displays is evidenced by the number of eye movements required to detect the target (and select an appropriate response) decreasing over the course of learning (e.g., Peterson & Kramer, 2001; Tseng & Li, 2004; Zinchenko et al., 2020b), consistent with the idea of more efficient guidance of the (overt) attention towards the target position – for instance, by a template-based (top-down) raising of the attentional priority assigned to this position within the distractor context (Wolfe & Horowitz, 2017). Interestingly, unexpected (but consistent) changes of the target location in an otherwise unchanged distractor context lead to an increase in the number of eye fixations needed to find the relocated target (though the number is still reduced compared to non-repeated displays; e.g., Brockmole & Henderson, 2006; Manginelli & Pollmann, 2009). This increase in fixation number has been attributed to the persistence of misleading attentional biases induced by the originally learnt (but after the change no longer applicable) target position.

Of note, though, even after a reasonable amount of practice with repeated displays (with a stable target location), the number of fixations needed to reach the target remains relatively high (amounting to some 4–6 fixations), rather than the target being among the very first items to attract an eye movement. This may suggest that, in hard search, tasks requiring serial oculomotor scanning, *display-specific* learning of spatial target-distractor relations actually plays only a limited role in guiding search toward the display region that may contain the critical item on future occasions. Instead, contextual learning may aid, or optimize, the selection of fixation locations along (at least parts of) the oculomotor scanpath, thereby increasing the likelihood of "hitting" the target location relatively early during the search (Kroell et al., 2019; Tseng & Li, 2004). In other words, in addition to acquiring relatively direct, *display-specific* guidance to the target location, contextual learning may drive adaptations of participants' general scanning strategies that broadly structure their search (e.g., Wolfe, 2021) in a *display-generic* manner, that is: a manner that is adapted to repeatedly encountered displays collectively, rather than individually (and that is little influenced by non-repeated displays, which – by virtue of being encountered only once – cannot consistently contribute to shaping this strategy).

Now, in relation to the target relocation effect (i.e., the increased number of fixations required after an unexpected change of the target position within an otherwise unchanged distractor context): the assumption that experience with repeated displays promotes automaticity of broad, display-generic scanning behaviour would predict an increase in the number of fixations needed to find the re-positioned target if the new target location, albeit display-dependent, is not fully compatible with individuals' proceduralized scanning strategy, such as a preference to move the eyes initially to the left half of a display. This is in contrast with the display-specific template account of CC, which predicts an increase in fixations following target relocation for each individual (initially learnt) target-distractor arrangement, which is dependent on the specific position of the target in the left or, respectively, right half of this display.

In the present study, we adopted a contextual learning/re-learning design and used the difficult T vs. L's letter search for studying statistical learning and re-learning in this task. During the initial learning phase, we expected attention, and the eyes, to land increasingly faster at the target location in repeated vs. non-repeated arrays; in the subsequent re-learning phase, upon the re-location of the target to another position within repeated displays, we expected this CC effect to turn into a cost. While this would replicate prior studies using the contextual learning/re-learning design, our main aim was to examine how the target relocation cost actually comes about, that is: Is it attributable to a persistent display-specific misguidance of attention to the originally learnt target location, which would be in line with the standard explanation of contextual cueing (as a relatively direct memory-based elevation of the selection priority assigned to the originally learnt target location)? Or is it due to the relocated target being positioned (relatively) away from the fixation locations making up some proceduralized, display-generic scanpath. Of course, the relocation cost may also arise from a combination of both mechanisms acquired as a result of contextual learning. The present study was designed to address this question.

Rationale of the present study

To this end, in the current study, we recorded eye movements to examine visual search through repeated vs. non-repeated (T vs. L's) letter arrays in an initial learning phase and a subsequent target-relocation phase (with constant distractor contexts in repeated displays). Our particular focus was on (quantitatively) investigating participants' broad scanning behaviours and whether and how these change with repeated encounters of (repeated) search arrays (i.e., learning experience). This would reveal information relevant to answering the question at issue, namely, whether, or to what extent, the "distraction" caused by target re-location in repeated displays (evidenced by an increase in search RTs and fixation number following target re-location) is a consequence of a proceduralized scanning bias (established during the initial learning phase) or the direct misguidance of (overt) attention to the initial target location. For instance, initial contextual learning may reinforce participants' tendency to scan search displays in a particular, idiosyncratic way, such as in left–right direction (e.g., Zelinsky, 1996; for reviews, see, e.g., Clarke et al., 2019; Husain & Rorden, 2003). Such a, for example, left-to-right strategy would predict that targets consistently presented in a (top, bottom) left display quadrant in the initial learning phase would require more fixations when they are re-positioned to a right quadrant in the re-location phase. To our knowledge, such display- (i.e., target-location) specific re-location effects have never been examined before. Importantly, such effects would go beyond those predicted by current LT-template-based accounts of contextual cueing, which attribute search guidance (in the initial learning phase), and consequently misguidance (in the re-location phase), to template activation by a repeated display top-down biasing attention towards the initial target location (by raising its priority on the attentional-priority map). According to these accounts, contextual repetitions should increase the "attend-to-me" signal for each repeated target position in each display quadrant (or hemifield) to a similar degree, leading to comparable search-guidance and misguidance effects for the specific target positions in individual repeated display arrangements.

To test these competing accounts, we set out, first of all, to replicate contextual facilitation in terms of the standard summary RT and eye-movement measures (e.g., the average number of saccades/fixations) that have informed theorizing in the extant contextual-cueing literature. Then, in order to investigate (i.e., to either corroborate or refute) the alternative account, we went on to examine oculomotor-scanpath measures that are diagnostic of fixations toward the original and re-located target positions. These analyses were designed to reveal detailed information about (i) the proceduralization of search performance with experience from repeated search arrays, which is lost in the "standard" averaging of dependent measures (such as RT and fixation number); and (ii) to pinpoint the source/s of contextual cueing as arising from the acquisition of display-generic scanning regimes or, respectively, adjustments of display-specific target priority coding; as both these contextual learning effects are expected to persist into the re-location phase (e.g., turning template-based guidance into misguidance), we can examine whether and to what extent they would explain the RT/fixation costs.

For our approach to be feasible, we adjusted our experimental design in two respects: First, and motivated by a previous study of contextual cueing (Sewell et al., 2018), we reduced the number of learnable, repeated target locations, as well as the number of target locations in non-learnable, non-repeated displays to four each, with one target location per display quadrant; this was meant to ensure that the memory signals for the respective target location would have as little interference from other repeated displays as possible, thus give the display- (i.e., target-location-) specific template accounts the fairest chance to show their explanatory value. Second, we presented the same repeated and non-repeated display arrangements to all participants, however, with a randomized order within blocks. Using the same display arrangements allowed us to control the perceptual content of the display set throughout the experiment; in particular, by using the same arrangements for repeated displays, we were able to detect display-generic (i.e., display-independent), broad oculomotor scanning strategies by minimizing confounds deriving from display composition (which typically varies across individual participants). And using identical non-repeated displays ensured a fair measure of contextual cueing, defined as the difference in performance between repeated and non-repeated displays, further eliminating confounds originating from, across participants, variably composed distractor-target configurations in non-repeated displays.

To preview our results: (i) We replicate previous findings of the speeding of visual search in terms of improved RTs and reduced fixation numbers in repeated displays. (ii) We find slowing of search following target relocation events, although savings were still evident compared to non-repeated displays. (iii) Disruption of search was not constant across all previous target positions. Instead, targets presented in a left display quadrant initially and a right quadrant after relocation led to pronounced slowing of search. (iv) But even when controlling for quadrant-specific variation of RTs (and fixation numbers), there was still contextual guidance toward original, i.e., learned, target positions, which led to misguidance after repositioning of the target to another location within the learned display layout. We take these findings to argue that expedited search in repeated displays is due to both the acquisition of search-guiding context memories (template-based LT memory) for specific displays (target locations) and the development of display-generic strategies of how to optimally scan the displays encountered (procedural LT memory). However, both these display-specific and general procedural mechanisms persist into a relocation phase, contributing to the disruption of search following the repositioning of targets within otherwise unchanged distractor contexts.

Methods

Participants

46 participants took part in the study (38 identifying themselves as female, the remaining 8 as male; 3 left-handed; mean age = 23.28 [SD = 5.62, range = 19–43]

years). The sample size was determined on the basis of previous comparable (large-sample) studies of CC that used a learning-phase/re-location-phase design (~50 participants per experiment; e.g., Peterson et al., 2022; Geyer et al., 2023). We considered high statistical power to be necessary because the abolishment, or even reversal, of contextual facilitation is most pronounced in the block of trials immediately following target re-location (e.g., Manginelli & Pollmann, 2009; Yang et al., 2021). Worth telling is that this sample size was also supported by a priori power analysis, based on a mean Cohen's d = .75 for a context × phase interaction effect and coming from 24 published studies that investigated the adaptation of contextual cueing across separate experimental (training and relocation) phases in altogether 31 different cases (for details see Zinchenko et al., 2023). Accordingly, a sample size of N = 46 participants was appropriate to detect this effect size with 81% power (groups = 2, number of measurements = 32) given an alpha level of .05 and a nonsphericity correction of 1 using the R library "WebPower."

Note that although we carried out only a single experiment, the results are not unusual or surprising: in critical respects, they replicate those from several previous target-relocation studies, which all failed to show successful relearning following the change in the target position within otherwise unchanged distractor contexts (e.g., Annac et al., 2017; Conci & Zellin, 2022; Geyer et al., 2023; Makovski & Jiang, 2010; Yang et al., 2021). The replication of key findings from prior studies thus validates our (in relation to these studies somewhat changed) display-design features and the large sample of participants. Importantly, however, the present study goes beyond the pattern of results established in previous studies in that it allowed us to test critical hypotheses that were motivated by the existing findings: beyond replicating these patterns in generic, summary-statistical measures of search performance, our innovative analyses of individual trial blocks and individual (repeated) displays – which were made possible by our changed design features and the large sample size – yield new insights into the mechanisms underlying CC.

Apparatus and stimuli

The experimental routine was programmed in Matlab with Psychtoolbox extensions (Brainard, 1997; Pelli & Vision, 1997) and was run on an Intel PC under the Windows 7 operating system. Participants were seated in a dimly lit booth in front of a 19-inch CRT monitor (AOC, Amsterdam; display resolution 1024 × 768 pixels; refresh rate: 85 Hz) at a viewing distance of 60 cm (controlled by a chin rest). The search displays consisted of 12 grey items (luminance: 1.0 cd/m^2; 1 target and 11 distractors) presented against a black background (0.11 cd/m^2). All stimuli extended 0.35° of visual angle in both width and height. As depicted in Figure 1, the items were arranged on three (invisible) concentric circles around the display centre (with a radius of 1.74°, 3.48°, and 5.22° for circles 1, 2, and 3, respectively). The position of each item was static, i.e., there was no trial-wise jitter in repeated displays, etc., within the circular array. The "T" target was rotated randomly by 90° to either the left or the right. The 11 remaining items were L-shaped distractors rotated randomly at cardinal orientations (0°, 90°, 180°, or 270°).

In repeated displays, the locations and orientations of the distractors were held constant across trials; in non-repeated displays, all distractors (i.e., their locations and orientations) were generated anew on each trial. Note that in all presented displays, the location of the target was repeated but the (left/

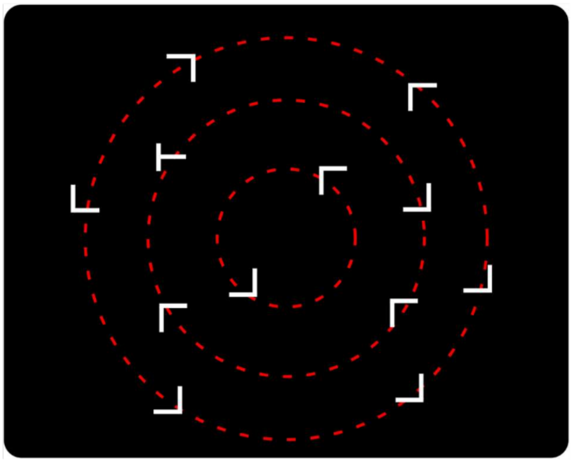

Figure 1. Example search display with a T target letter rotated by 270° relative to the vertical and 11 L distractor letters rotated 0°, 90°, 180°, or 270°. The red dashed circles, depicting the three concentric rings on which the search items were arranged, were not shown in the actual search displays. Targets in repeated and non-repeated displays were always presented at the second (middle) ring at a total of $N = 12$ different positions distributed equally across the four display quadrants.

right) orientation of the target was determined randomly, rendering it unpredictable. As a result, a repeated context could only be associated with a specific (repeated) target location, but not with a specific target identity. Following Chun and Jiang (1998), this approach is used in most contextual-cueing studies to ensure that contextual facilitation of RTs is owing to the repeated context guiding attention/the eyes, rather than facilitating the selection of the manual response (invariably) associated with a given repeated display. Importantly, both the set of (N = 4) repeated displays and the set of (N = 256) randomly generated non-repeated displays were kept constant across all 46 participants (though trial order was random in each block, consisting of $N = 4$ repeated plus $N = 4$ non-repeated trials), so that each participant encountered identical repeated and non-repeated configurations. This enabled us to keep low-level, individual display properties constant across participants and thus compute dependent – scanpath – measures for each individual display (with variations between participants providing the error term).

There were overall 12 possible target locations: 2 × 4 of these were used for repeated displays: 4 positions in the learning phase and 4 other positions in the re-location phase; the remaining 4 locations were used for non-repeated displays. That is: non-repeated displays had random non-target layouts in each trial together with a set of 4 unique target locations in the learning phase and the same set of 4 identical target locations in the subsequent re-location phase. All targets, in both repeated and non-repeated displays, were located on the second ring, controlling for the distance of the target from the display centre in all conditions. Furthermore, the targets were placed in all four quadrants with equal probability in all context (repeated, non-repeated) and phase (learning, re-location) conditions. At the transition from the initial learning to the re-location phase, the target in a repeated display swapped its position with that of a distractor in the opposite hemifield (e.g., a target on the left replaced a distractor on the right, with the distractor "moving" to the original target location). Of note, participants were not informed about the fact that some of the search arrays were presented repeatedly.

To monitor and record eye movements, a video-based eye-tracker was used (EyeLink 1000; SR Research Ltd., Mississauga, Ontario, Canada; version 4.594). Eye-movement recordings were calibrated at the start of the experiment and after every four blocks (of 64 trials). Calibration was considered accurate when fixation positions fell within ~1° for all calibration points. The default psychophysical sample configuration of the eye-tracking system (i.e., saccade velocity threshold set at 35°/s, saccade acceleration threshold set at $9500°/s^2$) was adopted for the eye-data samples.

Trial sequence

A trial started with the presentation of a central fixation cross (0.10° × 0.10°, luminance: 1.0 cd/m2) for 500 ms. Next, the fixation cross was extinguished and a blank interval was presented for 200 ms, after which the search display was presented. Participants were instructed to respond as quickly and accurately as possible to the orientation of the target "T" (left vs. right) and were permitted to move their eyes freely. Each search display stayed on the screen until a manual response was elicited. If the "T" was rotated to the right (left), observers responded by pressing the right (left) arrow button on a computer keyboard with their right (left) index finger. Following a response error, the word "Wrong" appeared on the screen for 1000 ms. Each trial was followed by a blank inter-trial interval of 1000 ms. The experiment consisted of 512 trials (2 phases × 32 blocks × 8 trials each, 50% repeated displays in each block). Participants were free to continue with the next block at their own pace. The experiment took overall some 50 min to complete.

Recognition test

At the end of the experiment, participants were given a recognition test in which they had to identify whether they had previously seen a given configuration or not (a "standard" procedure in contextual-cueing experiments; cf. Chun & Jiang, 1998). Participants were shown 4 displays that had been previously presented (with the original target positions) and 4 new ones and were asked to press a button indicating whether a given display was new or old. Participants judged the repeated and non-repeated displays twice, yielding a total of 16 recognition trials, to increase the accuracy and statistical power of the test (cf. Vadillo et al., 2016). Participants were not timed and were not given feedback on their responses.

Results

Individual mean error rates and reaction times (RTs) were calculated for each factorial (Block × Context × Phase) combination. For the RT analysis, error trials and RTs above and below 3 standard deviations from the mean were excluded from analysis, leading to the removal of < 2% of all trials. To take care of the likely dependence of our results on the, across participants, constant sets of repeated and non-repeated search displays, we employed linear mixed-effects models that explicitly take into account the random variability deriving from individual non-/ repeated displays when estimating the effects of the fixed factors of context, block, and phase. In more detail, we formalized the relationship between the dependent variables and the fixed/ random factors in our general mixed-effects model as follows (using the lme4 package in R, version 3.4.3; R Core Team, 2018):

$$
\begin{aligned}
&generalmixed - effectsmodel \\
&= lmer(dependentvariable \sim context{*}phase{*}block \\
&+ (1|participant) + (1|targetquadrant), data = \ldots)
\end{aligned}
$$

where the *dependent variable* is either reaction time, number of fixations, or fixation direction; *context* refers to repeated vs. non-repeated target-distractor arrangements; *phase* is the initial learning vs. the relocation period; and *block* reflects time on task (L1–L32 in the learning and R1–R32 in the relocation phases). The random factor *participant*, which refers to individual participants, allows our general mixed-effect model to incorporate per-participant adjustments in the intercept of the functions relating our dependent variables to our fixed factors. The random factor *target quadrant* refers to position-specific intercepts, that is, the intercepts for the display quadrant in which a target is located at learning and relocation, respectively, allowing for per-quadrant adjustments of our averaged dependent measures.

Additionally, since we were also interested in how our dependent measures would vary as a function of *individual* repeated displays, we built more specific mixed-effects models in which we explicitly considered target quadrant as fixed effect and participant as random effect. These models are presented in greater detail in the respective result sections.

Result 1: contextual cueing expedites search, but diminishes after target relocation

Figure 2 depicts mean RTs and fixation number for repeated and non-repeated displays in individual blocks of the learning and re-location phases. The mean RT across all experimental conditions was 1,423 ms. The (block × context × phase) mixed-effect model revealed a main effect of block and a main effect of context (*Fs* $>=$ 10.31, *ps* $<$ 0.001). Importantly, there was an interaction between context and phase, $F(1, 17446) = 63.81$, $p < 0.001$, reflecting a substantial contextual-facilitation effect (i.e., faster search RTs for repeated vs. non-repeated displays) in the learning phase (207 ms); this compares with a (by 66%) reduced, but still significant facilitation effect (of 72 ms) in the relocation phase ($F(1, 8697.4) = 42.23$, $p < 0.001$).[2]

The mean number of fixations required for the eyes to land at the target location was 8.68. Analysing this number by another (block × context × phase) mixed-effect model revealed main effects of block and context (*Fs* $>=$ 5.70, *ps* $<$ 0.001). The theoretically important context × phase interaction was significant, too, $F(1,17446) = 32.72$, $p < 0.001$. Fewer fixations were required to detect the target in repeated than in non-repeated displays (7.29 vs. 8.13 fixations), but contextual facilitation in terms of eye movements was greater in the learning phase (CC effect: 1.2 fixations) than in the relocation phase (CC effect: 0.48 fixations; reduction: 60%), though CC remained reliable in the latter phase ($F(1, 8697.5) = 17.54$, $p < 0.001$).

Result 2: the reduction of CC during re-location is due to misguidance of the eyes toward previous target positions

Given our search task was a hard one (involving serial scanning), we expected fixations to approach each display item with some probability (independently of their status as a target or non-target, i.e., distractor,

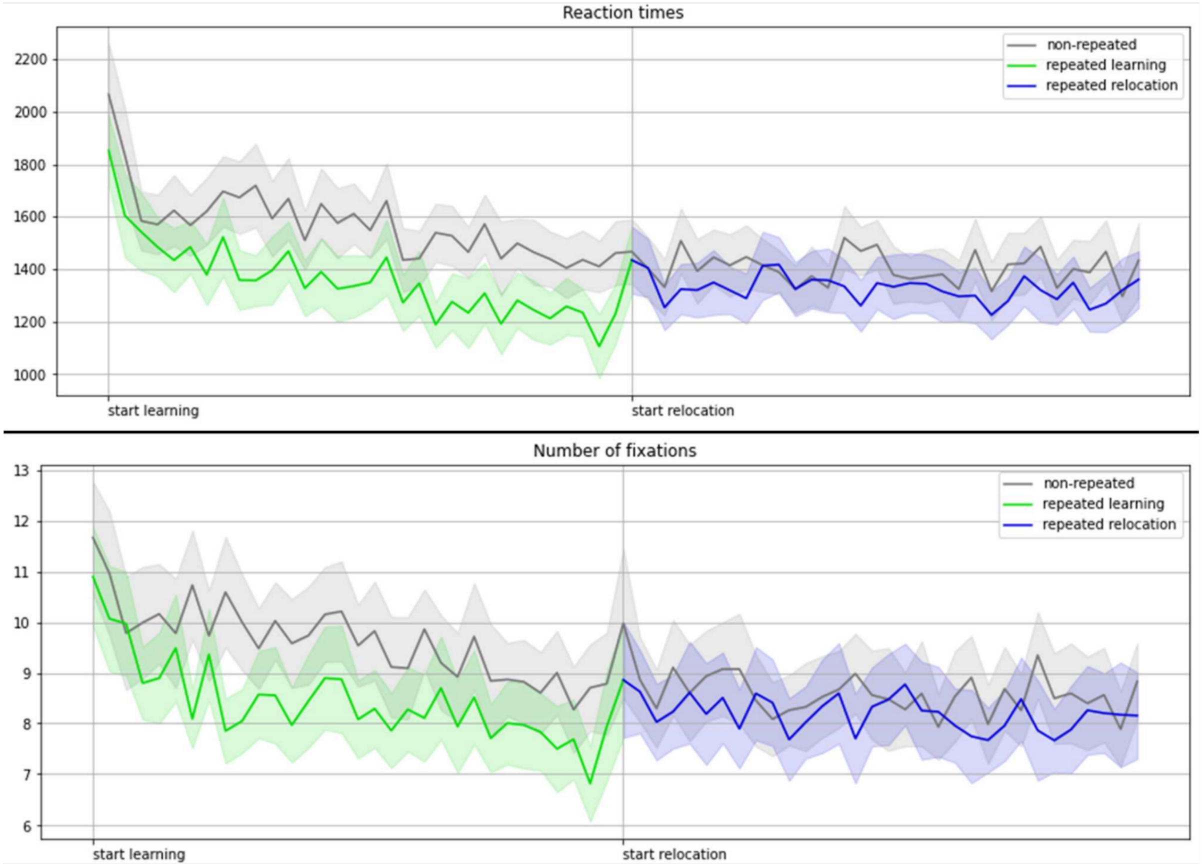

Figure 2. Mean reaction times in milliseconds and fixation number (upper and lower panels, respectively), with associated standard error bars, for repeated and non-repeated contexts (dashed and solid lines, respectively) in the learning and relocation phases. Each phase consisted of 32 blocks.

element). However, if a given repeated display activates an acquired bias toward a particular – target – item, fixations should approach its locations with a higher probability. Importantly, this bias should persist during the target relocation phase. One measure to quantify misguidance then is to determine the distance that each fixation on the entire scanpath has (directly) relative to the original – that is, learnt – target location in the relocation phase and compare this measure to the "baseline" distance that each fixation has from the future, relocated target in the initial learning phase (in which the relevant location contained a distractor). Specifically, we computed the distribution of minimal distances of any fixation to the "other" target for (i) the last learning block, (ii) the first relocation block, and (iii) the last relocation block, the latter to assess how a scanning bias toward the original target location persisted during the relocation phase. Average trends in fixational distances in these blocks are illustrated in Figure 3.

As can be seen, the distance plots are bimodal, with one peak centred around 270 px – which corresponds to the distance between the display centre (containing the fixation marker, where the search begins) and the locations of the original (L32) and the relocated (R1, R32) targets. Here, distances with a peak of 270 px (pixels) therefore correspond to those scanpaths for which the closest distance to the "other" target was likely to be the initial fixation, i.e., those scanpaths which monotonically diverged from the "other" target. The second peak centres around an average distance of 100 px (relative to the future, relocated target) in the last block of learning (L32), while it centres around 0 px (relative to the original target) in both the first and the last block in the relocation phase. The peak corresponding to a distance of 100 px to future targets can be thought

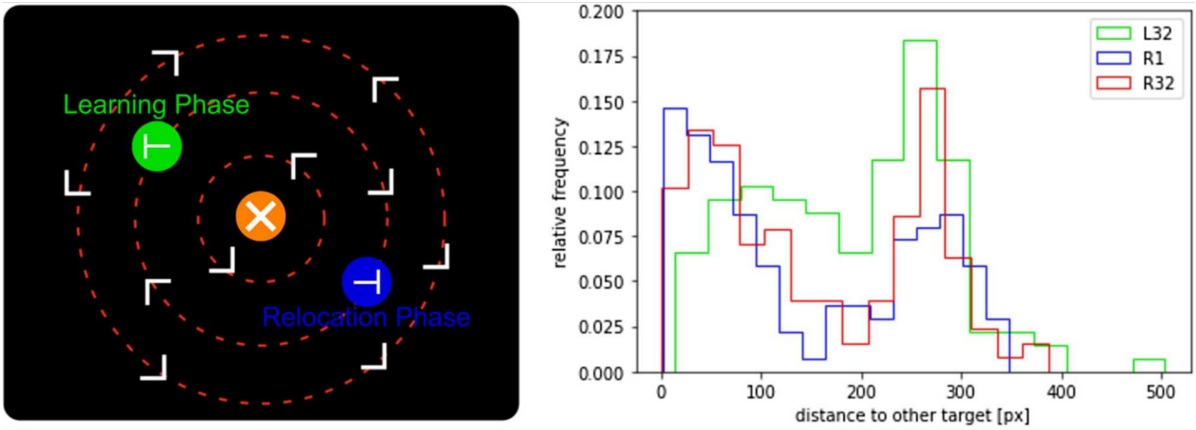

Figure 3. The left panel presents an adaptation of the example search display from Fig. 1, now highlighting an original and a relocated target (in the left and right display halves, respectively) in an otherwise unchanged distractor array. The right panel presents the proportion of distances (in pixels, px) of any fixation within a scanpath from the future (relocated) target location during the last block of learning (block L32) and from the original, i.e., learnt, target location in the first and last relocation blocks (R1 and R32, respectively). Each distance plot – L32, R1, and R32 – shows two distinct peaks centreing around the display centre (marked with an "X") and around the positions of learnt ("Learning Phase") and relocated ("Relocation Phase") targets.

of as a baseline when compared to the closest distance (with a peak of roughly 0 px) to previous target positions as both would incorporate differences arising as a consequence of relocation. Direct *t* tests of the fixational distances in the three (L32, R1, R32) blocks revealed the distances to the original targets to be significantly shorter than those to the relocated targets (all *t*'s > 3.53, *p*'s < .01).

This finding may reflect a visual exploration strategy that enables participants to better work through "difficult" T- (target) vs. L- (distractor) type search arrays, involving occasional visits of future target positions – a strategy that is operating even after 32 blocks of repetition and continuing in the relocation phase, though it leads to misguide of search in the form of eye fixations towards original, i.e., learned target positions. Alternatively, visual exploration may have been influenced by factors specific to our display design. Recall that all targets, in repeated and non-repeated arrays, were presented on the second ring around the central fixation cross (to equate target eccentricity effects between the two types of displays). It is thus possible that participants detected this regularity and so preferentially fixated near these (target) locations. Nevertheless, the learning of this regularity offers no ready (or at a best a partial) account of the present findings. Learning of, and attendant search guidance from, the additional ring-2 regularity would predict that fixations should land at future (during the learning phase) and previous (during the relocation phase) target locations to a similar extent. Critically, however, there was an average fixational distance of 100 px relative to a distractor item when this distractor "only" became a future, relocated, target (L32), compared to a fixational distance of 0 px relative to a distractor item in the relocation phase, when this distractor appeared at the original – that is, contextually learnt – target position (R1, R32). Thus, the pattern of distance effects shown in Figure 3 could at most reflect a combination of target position learning (happening with both future and original target locations) and contextual learning, the latter making it particularly likely that the eyes move in the direction of the original target positions (in a target relocation phase). Accordingly, a comparison of fixation locations between future vs. original target positions (100 vs. 0 px distances in green vs. blue distributions in Figure 3) would provide a conservative approach to estimating the pure effect of contextual learning, controlling for potential effects arising from the learning of (ring-2) target positions. Worth noting is that the fixational bias toward original target positions remains evident even after extensive practice with the relocated displays (in relocation block R32): statistically, CC was comparable between the first and the last relocation block, in terms of both RTs, $t(45) = 0.26$, $p = 0.80$, and fixation number, $t(45) = 0.64$, $p = 0.53$.

Another way to look at the data is to use gaze plots of our individual participants' closest fixations to future target locations in the last block of the learning phase (L32) and their closest fixations to the original target locations in the first block of the relocation phase (R1), for each individual repeated display. As can be seen from Figure 4, there is a strong clustering of fixations toward the original target positions. Here, distance is colour-coded (from warm to cold colours), showing a marked cluster of orange dots for relocated displays, but a more or less constant gradient of colours during the learning phase. Interestingly, rather than being spatially even, the clustering is more pronounced for original target locations in (previous) left-side display quadrants. This observation was supported by a DBSCAN cluster analysis (Birant & Kut, 2007), which estimates clusters by "just" looking at the local density of fixations without the requirement to set the number of clusters or their size in advance. As a measure of clustering, we computed the Davies Boudin Score (DBS), which yields low values for well-defined clusters and high values otherwise. Note that all DBS values for the learning phase were above 1, while those of the relocation phase were substantially below 1. For visualization, Figure 4 also includes a circle with a radius of 135 px: neither cluster in the last block of learning is represented exclusively within this boundary, while one and only one cluster is always represented within this region in the first relocation block (reminiscent of the "valley" at 145 px between the two peaks in Figure 3). In more detail, clustering scores were overall lower (indicative of higher spatial clustering/regularity) for fixations towards the original target locations compared to fixations toward future locations (mean DBS values of 6.99 and 0.33, respectively). While this suggests that search was misguided toward previous target positions, this negative bias was not constant but rather varied as a function of the quadrant of the original target: clustering scores were particularly low for fixations toward (original) targets in the left display half compared to targets in the right half (DBS of 0.24 and 0.43, respectively), indicative of previous targets in a left-side display quadrant decelerating search after their re-location to a right-side quadrant. To quantify this observation, we built a specific mixed-effect model on the RT costs (i.e., the RT difference between learning and relocation), with minimal fixation distance at relocation and target quadrant as fixed effects and participant as random effect. This model yielded a main effect of distance ($F(1,89) = 32.85$, $p < 0.001$) and a main effect of target quadrant ($F = 3.31$, $p < 0.05$). Of note, the interaction between distance and target quadrant was also significant, $F(3, 89) = 2.78$, $p < 0.05$, indicating that the effect of fixation distance on RTs was greater for targets relocated to the right display half.

To further assess eye directions on a given trial, we computed the *ordinal number* of the fixation that was closest to the future target location for the last learning block and the closest fixation to the original target location in the first re-location block, as well as the *relative position* of fixations toward future/original targets along the complete series of fixations on a given trial. The relative position was quantified by dividing the *ordinal number* of the fixation by the *total number* of fixations on a given scanpath, yielding scores between 0 and 1. Note that we assigned the first fixation in a scanpath an index of zero. Figure 5 plots average trends in ordinal fixation number when fixations were closest to the "other" target (i.e., future target in learning and original target in re-location). As can be seen, not all fixations within a sequence of fixations had equal propensity to land in close proximity of the "other" target. Specifically, there was a stronger tendency for early fixations to approach the "other" target location. Notably, this effect was seen for both future and original targets. To test this, we submitted the ordinal number and relative position of the closest fixation (within a scanpath) to the "other" target as dependent variable to separate mixed-effect models (one for the ordinal number and one for the relative position), each with block (last block in learning, first block in relocation) and individual repeated display (1-4) as independent variable and participants (1-46) as random factor. The two-way, block × display interaction turned out significant ($Fs >= 5.10$, $ps < .002$): during the learning phase, close fixations to future target locations (which, in the learning phase, were occupied by a distractor) occurred earlier for repeated displays that had a future target in a left-side quadrant (and thus the original target in a right-side quadrant); by contrast, in the relocation phase, original target locations in a left-side quadrant (which, in the relocation phase, were occupied by a distractor) led to earlier close fixations. That is, the pattern of close fixations was

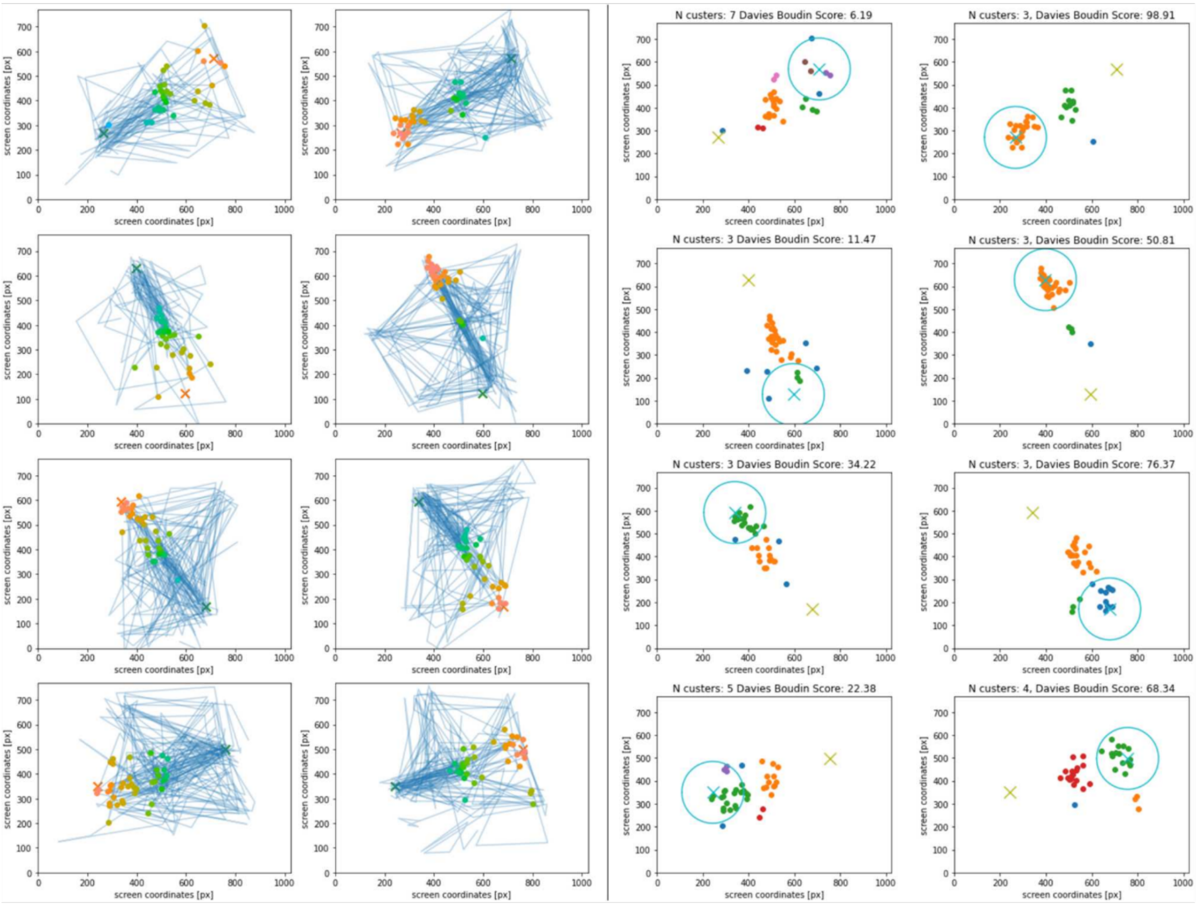

Figure 4. Visual-exploration scanpaths through each of the four repeated display configurations during the last block of learning (L32) and the first block of relocation (R1); the many blue lines in the left panel represent our N = 46 individual participants. The actual target is marked by a blue x, and the respectively "other" target (i.e., future target during learning and the previous target during relocation) by an orange x, while the rows show each (context matched) repeated display. The fixation on each scanpath that is closest to the "other" target is highlighted by a coloured dot; in the two columns on the left, the colour changes according to distance from warm to cold colours, while blue lines outline saccades between the other fixations (here the corners between the lines). During the learning phase, the colours show a gradient, while in the relocation phase, mainly two extremes can be seen, which indicates that these fixations are clustered. The right two columns show the outcome of a DBSCAN cluster analysis (Birant & Kut, 2007). Here, colour represents members belonging to a cluster. A circle of a radius of 135 px highlights an area which covers one (and in each case only one) well-defined cluster for the relocation phase (while this is never the case in the learning phase).

predictable from the placement of the original and relocated targets in individual display quadrants, rather than individual "contextual-cueing" displays. This conclusion was supported by two additional mixed-effects models (on the ordinal fixation number and relative scanpath position), with block and target quadrant (top-right, top-left, bottom-left, and bottom-right) as fixed effects and participant as random effect. Here, only the main effects of target quadrant were significant (Fs > = 4.12, ps <.007).

In sum, we carried out an analysis of the temporal sequence of fixations by assuming that individual participants display a visual-exploration strategy that explains a large proportion of the variation in fixational placements towards the original and relocated target positions. An analysis of the ordinal position of each fixation that was closest to the other target showed that when search begins, fixations that depart from the display centre (towards the "true" target) will necessarily diverge from the other target – so, these early fixations (marked by an ordinal position of 0) will inevitably constitute those closest to the "other" target. As such, the finding of comparable peaks of nearest fixations for both

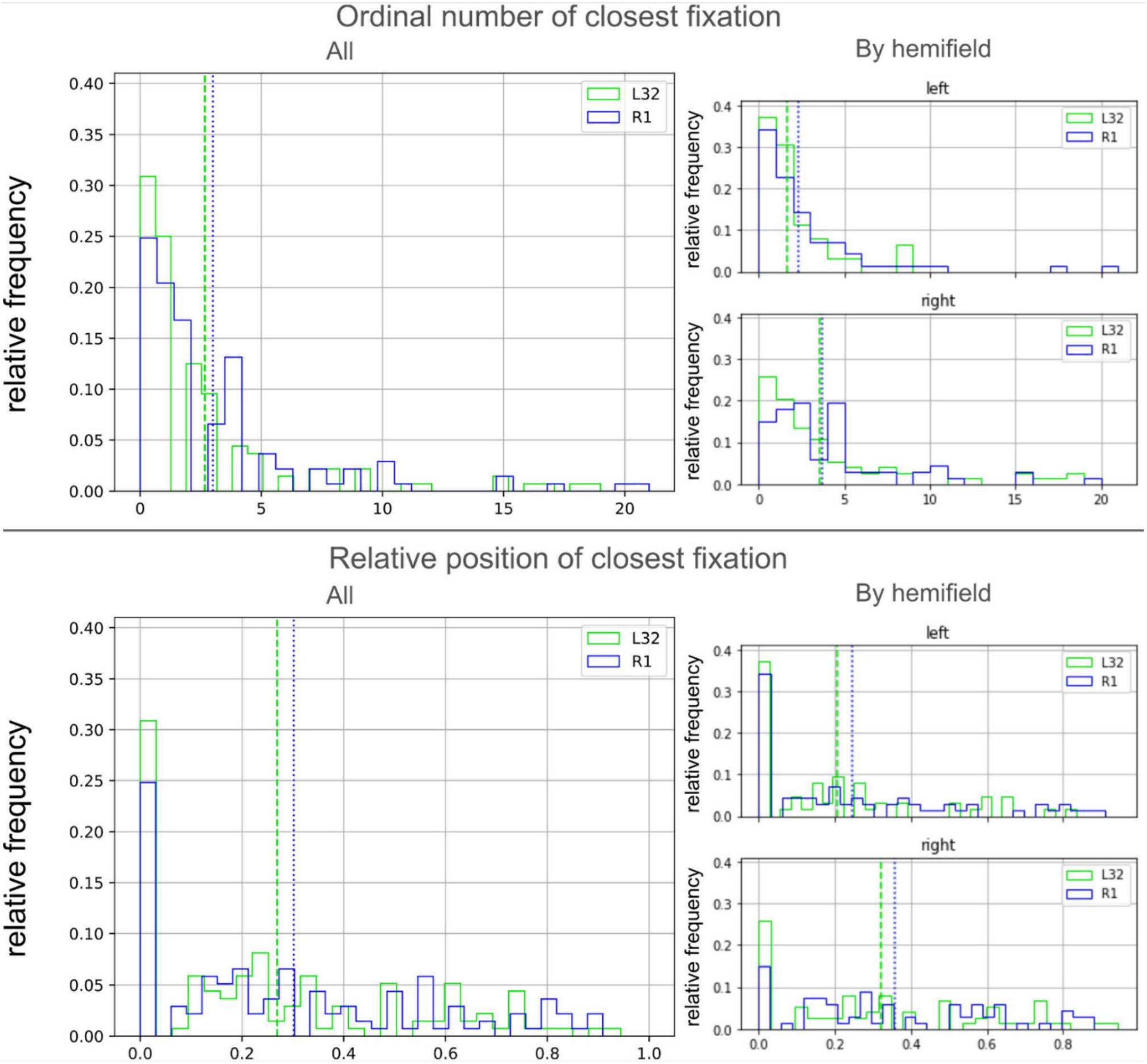

Figure 5. Ordinal number and relative position of fixations near the future and the original target location in the sequence of fixations making up the complete scanpath for the last learning block (L32) and the first relocation block (R1). The top panels on the right side show original and, respectively, relocated targets appearing in the left display half (quadrants 2 and 3), and the bottom panels on the right side show original and, respectively, relocated targets appearing in the right display half (quadrants 1 and 4). Dashed/dotted lines indicate the mean for illustration purposes. Note: ordinal numbers are counted from 0 onwards in order to maintain consistency with the relative position, for which the first fixation should be 0/n instead of 1/n, where n is the number of fixations on each scanpath.

future and relocated targets constitutes important piece of evidence for the operation of a display-general visual-exploration strategy, without necessitating the assumption of oculomotor guidance from display-specific target-context templates in LTM coming into play. Critically, the temporal pattern of near fixations was nevertheless predictable from the placement of the original and relocated targets in individual – that is, left-side – display quadrants. We take this to suggest that participants work through "difficult" T vs. L's search displays preferentially employing a left-to-right scanning strategy.

Next, adopting a procedure used in previous studies of contextual cueing (Brockmole & Henderson, 2006; Manginelli & Pollmann, 2009), we examined the average eye direction for the first five fixations from the beginning of search and across all 32 blocks of the learning phase and all 32 blocks of the relocation phase. These studies found that following target relocation, fixations are reliably (mis-)

guided towards the original target location, an effect that was seen during the initial five fixations in the trial. Accordingly, the effects of experience with repeated arrays should become measurable with these initial fixations.

Figure 6 shows the x-coordinates of fixations 1–5 relative to the centre of the display. As can be seen, there is a "leftness" of fixations that increases almost linearly with practice on the task – an effect that is seen for each individual fixation 1–5 and that persists even in the relocation phase. Each fixation's "leftness" was submitted to our general mixed-effects model, with context, block, and phase as fixed effects and target quadrant and participant (intercepts) as random effects in a crossed design – the latter permitting us to estimate the effects of our fixed factors in the development of a leftward scanning preference, independently of our specific individual displays (and participants).

This model revealed significant main effects of phase for all fixations ($Fs > 16.35$, $ps < 0.001$). For fixations #2 and #3, the main effect of block was also significant, $Fs > 2.07$, $ps < 0.001$. Critically, there was a significant main effect of context for fixation #2 ($F(1,16132) = 5.8247$, $p = 0.016$): scanning of repeated versus non-repeated contexts was associated with an increased leftness of fixation #2. A similar numerical (albeit statistically non-significant) difference was also evident for fixation #3 ($F(1,14100) = 2.17$, $p = 0.14$) and fixation #4 ($F(1,11993) = 3.65$, $p = 0.056$), while the context effect was again significant for fixation #5 ($F(1,10197) = 4.43$, $p = 0.035$). Thus, initial oculomotor scanning becomes increasingly left-directed with practice on the task (an effect seen with both repeated and non-repeated arrays, indicated by main effects of phase/block), though repeated displays carry greater weight in shaping this "leftness" strategy due to their increased frequency of occurrence (indicated by a main effect of context).

Result 3: misguidance of search towards previous target positions emerges from acquired left-right scanning and ongoing prioritization of original – that is, learnt – target positions in individual CC arrays

Finally, we computed RTs and the number of fixations in the last block of learning (L32) and the first block of re-location (R1) for each target position/display quadrant in each of our four repeated displays. See the "context-matched" and "target-matched" conditions in Figure 7, which shows RTs/fixations to original and re-located targets in different quadrants of identical repeated-context arrays (context-matched condition) and RTs/fixations to original and re-located targets presented in identical quadrants, but in different repeated-context arrays (target-matched condition). The last learning block is shown in green and the first re-location block in blue. Individual lines represent individual participants and larger lines indicate the mean of each block and quadrant, while the large grey lines indicate the overall mean for that quadrant (across the two blocks). Worth noting is that guidance and misguidance of search almost entirely depended on the target's display quadrant: while original targets located in the second and third quadrants (corresponding to the left display half) have an advantage in terms of both RT and fixation

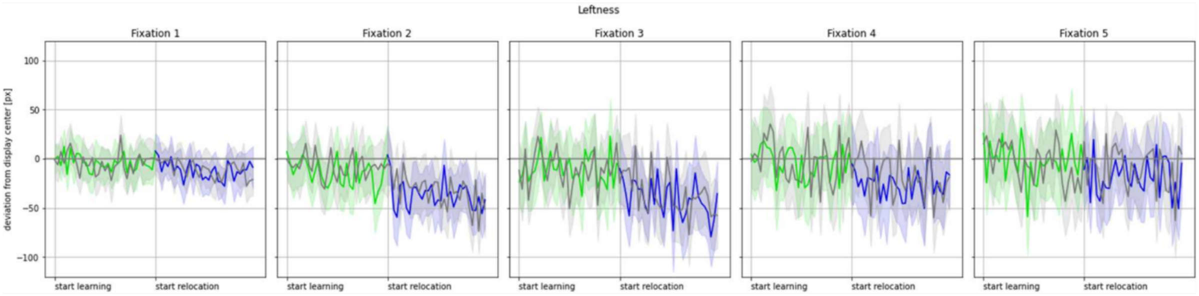

Figure 6. Average eye direction in repeated arrays for the first five fixations from the beginning of search in the all 32 blocks of the initial learning phase and all 32 blocks of subsequent re-location phase (green and blue lines, respectively). Negative values indicate that a fixation is to the left from the display centre (marked by the horizontal line). Grey lines indicate fixation direction in non-repeated search layouts.

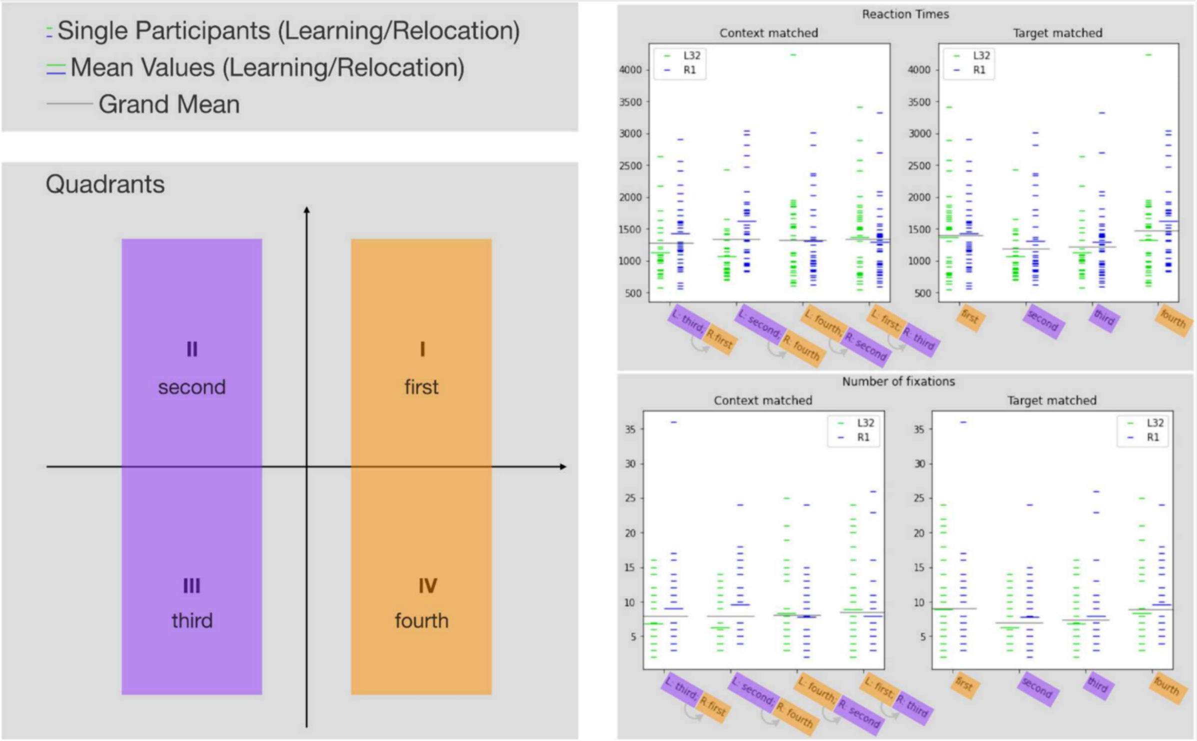

Figure 7. Reaction times and number of fixations of our N = 46 participants when searching for targets in repeated displays in the last block of the learning phase and the first block of the re-location phase (green and blue colours), respectively, separately for the four display quadrants (L = learning, R = relocation; first, second, third, and fourth corresponds to the top-right, top-left, bottom-left, and bottom-right quadrants, respectively). The left-hand panels show performance for the original and re-located targets when analysed for identical repeated-context displays (but when the respective targets occur in different display quadrants; = context-matched analysis) or identical display quadrants (with the respective targets being presented in different repeated-context displays arrays; = target matched).

number in the initial learning phase, search is particularly disrupted when these targets changed their positions to the first and fourth quadrant (on the right display half) during the re-location phase. Specifically, re-locating a target from the second and third, i.e., left-side, quadrants to the first and fourth, right-side, quadrants in an otherwise identical contextual array resulted in a rather large average cost (over 200 ms and over 4 fixations), while changes in the reverse order led to hardly any slowing of search. Separate mixed-models for RTs and fixation number with the fixed factors block (last block in the learning phase, first block in the relocation phase) and target quadrant (1-4) and the random factor participant (1-46) revealed the theoretically important block × quadrant interactions to be significant (*Fs* > 3.29, *ps* < 0.021).

Of note, quadrant-specific differences in search also come to the fore when matching the target quadrants between learning and re-location (in which case the “target-matched” displays are necessarily different in terms of spatial composition/individual distractor contexts). This can be seen from the right-hand panels of Figure 7, which illustrate how search unfolds with original and re-located targets presented in the very same quadrants (though within different repeated-context displays). Here, targets presented in the second and third (i.e., left-side) quadrants “outperform” those appearing in the first and fourth (right-side) quadrants. In other words, targets located in the second and third (left) quadrants are associated with the fastest RTs and lowest fixation numbers – importantly, independently of block.

However, as can also be seen, RTs were faster and fewer fixations were required to find original relative to re-located targets – critically, even when display quadrants were equated between the two (original, re-located) target types. In other words, a target in a given display quadrant is processed differently depending on whether (or not) there was *display-specific training* for that particular (original) target location.

Nevertheless, display-specific learning was not constant across quadrants, as is indicated by more pronounced differences when targets were relocated from the left two quadrants to right quadrants (see Figure 7, upper and lower left plots). Corresponding to each of these plots, we computed mixed-effects models, with the fixed factors block (last block of learning and first block of relocation), as well as target quadrant ("target matched"; upper and lower right plot for RT and NF, respectively) and display (i.e., "context matched"; upper and lower left plots for RT and NF) and the random factor participant. These models revealed a significant (RT: $F(1, 230.52) = 9.70$, $p = 0.0021$) and marginally significant (NF: $F(1,231.40) = 2.84$, $p = 0.094$) main effect of block, in line with the hypothesis of a general cost. Importantly, in the models with display as a fixed effect (corresponding to the left two plots in Figure 7), there was a significant block × display interaction (Fs $> =$ 5.21, $p < 0.022$), while the main effect of display was not significant (Fs $< =$ 0.26, ps $> =$ 0.86). This means that before and after relocation, the cost differed significantly in magnitude between contexts. Conversely, in models with target quadrant as fixed effect (corresponding to the two plots on the right side of Figure 7), there was a main effect of target quadrant (Fs $> =$ 3.17, ps $<.025$), whereas the interaction between target quadrant and block was non-significant (Fs $< =$ 0.67, ps $> = .57$). This suggests that the overall number of fixations and reaction times were determined by identical quadrants, with different contexts (rather than by context per se), and that the magnitude of the relocation cost was not influenced by the target quadrant.

Accordingly, the analysis of reaction times and the number of fixations illustrates that the relocation cost is a result of *both*, broad scanning strategies, with a bias to inspect the left half of the display first, *and* display-specific mechanisms, worsening overall performance by continuing prioritization of the previous target location (even) after target relocation. In other words, the pattern of guidance and misguidance of search acquired through experience with repeated-context arrays results from the joint operation of two – display-specific and display-unspecific – learning mechanisms.

Result 4: participants lack explicit knowledge about display repetitions.

Participants' mean hit rate (repeated displays correctly identified as repeated) was 52.5%, which compares with a mean false-alarm rate (non-repeated display incorrectly judged as repeated) of 47.5%, $t(45) = 0.88$, $p = 0.38$. Thus, there was little indication of explicit, above-chance recognition of displays that had been encountered repeatedly over the course of the search task.

Discussion

The goal of this study was to evaluate whether the guidance and misguidance of search, the latter following target re-location in an otherwise unchanged distractor layout, arise as a consequence of up-modulations of the attentional priority assigned to the (original) target location within a specific distractor context (contextual-template-based LT memory) or from the development of oculomotor scanning routines that are optimally adjusted to the set of displays encountered, with repeated displays shaping the routine more than non-repeated displays (oculomotor procedural LT memory). Looking at two distinct – reaction-time and oculomotor scanning (e.g., number of fixations) – measures, we found consistent evidence that display repetitions shape participants' broad, (relatively) display-independent, scanning procedures. Overall, practice with repeated search arrays led to a bias to start scanning the displays on the left. The bias of the eyes to return to the previous target location in the left half (rather than going towards the right half, where the new target appears) could explain the slowing of RTs to re-located targets after the change and can be seen in the large costs for displays in which the target was originally located on the left and then relocated to the right, compared to practically no cost for contexts with the original target position on the right (although here, a learnt context with the original target on the right yielded comparable RTs and numbers of fixations to a display in which the relocated target is positioned on the left, i.e., appearing at a position that is preferentially scanned earlier). Moreover, processing differences remained even when equating the quadrants of original and re-located targets (i.e., when comparing different repeated-context displays with the original and, respectively, re-located target in the same, left quadrants). Since these differences cannot be explained by a proceduralized general scanning routine (which would be the same for the two types of display), they indicate instead that practice with repeated displays (also) leads to context-specific, template-driven changes in the coding of attentional priority for individual target locations. Further, we find that participants have a strong bias to visit the previous target location after relocation, indexed by the distance to previous target locations being significantly smaller than would be expected (i.e., when compared to future target locations) from search in general. Additionally, the distance of the closest fixation to the previous target location in relocation predicts the relocation cost. Thus, these findings favour an account that proposes two qualitatively different components of contextual cueing: learning to broadly structure the search in (initially) leftward direction and learning of individual target locations within individual repeated-context displays. Critically, these mechanisms continue to operate after target relocation events, constituting two sources underlying the misguidance of search towards previous target locations.

Contextual cueing involves the development of oculomotor scanning routines

One intriguing and new finding of the present study is how the savings in RTs and fixation number actually come about in repeated search displays. The currently dominant account attributes these savings to the search being "cued", or "guided", (more or less) directly to the target location by acquired (long-term) memory representations, or "templates," of specific distractor-target arrangements. Templates are activated by re-encountering, on a given trial, the corresponding arrangement. This then top-down raises the attentional priority of the target location (e.g., Beesley et al., 2015; Brady & Chun, 2007; Chun & Jiang, 1998; Geyer et al., 2010), enhancing its potential to summon attention. According to this proposal, the number of attention shifts required to detect a target in a particular item arrangement will decrease with increasing re-encounters of this arrangement, due to the build-up of a search-guiding memory template for this particular display. Support for this comes from studies that examined fixation number (e.g., Peterson & Kramer, 2001) and/or (across repeated displays) aggregated oculomotor-scanpath indices (e.g., Manginelli & Pollmann, 2009; Tseng & Li, 2004) as a dependent measure, which showed that finding the target in repeated (vs. non-repeated) displays required overall fewer eye movements. However, while this "specific-template" account is attractive (not least because it ties in seamlessly with the functional architecture assumed by general theories of search guidance, such as Guided Search; Wolfe, 2021), the results from our non-aggregated measures support an additional, and alternative, "procedural" account of the savings that does not rely on the notion that observers acquire contextual memories specific to particular distractor-target arrangements. In fact, that there is procedural learning in contextual-cueing paradigms is a universal finding in virtually all pertinent studies (for reviews, see, e.g., Goujon et al., 2015; Sisk et al., 2019): search RTs decrease, typically quite substantially, over the course of practice on the task (i.e., across consecutive trial blocks). Importantly, this improvement is evident in non-repeated (as well as repeated) displays – which is why it is attributed to general procedural learning, which optimizes task performance through the development of an automatic (search) routine (e.g., Fitts, 1964; Fitts & Posner, 1967; Shiffrin & Schneider, 1977). The fact that this improvement is more marked for repeated (vs. non-repeated) distractor-target arrangements is simply attributable to repeatedly searched displays accruing greater weight in shaping this scanning routine compared to non-repeated arrangements which, by definition, are encountered only once (Seitz et al., 2023). Accordingly, there is no need to assume qualitative difference in search performance between repeated and non-repeated displays to account for (at least part of) the contextual-cueing effect. – With the search displays implemented in the present study, generic optimization of oculomotor scanning is evidenced by participants developing a leftward bias in initial saccade direction.

Broad display scanning and high-fidelity memory templates

While our results indicate that procedurally learnt, display-general oculomotor scanning routines play a significant role in contextual cueing, they also highlight the importance of display-specific learning. In particular, we found indications (in the re-location phase) of persistent search biases towards original target positions even when we matched the (original, re-located) target quadrants. This is consistent with the idea that contextual cueing involves associative learning of specific spatial distractor-target relations (e.g., Geyer et al., 2021; Goujon et al., 2015). Perhaps the strongest support for this idea comes from electrophysiological studies that examined lateralized event-related potential (ERP) components, such as the N2pc (e.g., Luck, 2014), to track the (lateralized) allocation of covert attention (e.g., Liesefeld et al., 2017). These studies provided evidence that focal attention is deployed rapidly and directly to the target location in repeated displays (e.g., Johnson et al., 2007; Schankin & Schubo, 2009; Stokes et al., 2012; Zinchenko et al., 2020a), with the lateralized ERP markers correlating well with the RT-cueing effects exhibited by the individual participants (e.g., Chen et al., 2022; see also Schankin & Schubo, 2009).

An important aspect of the present findings, which we have not emphasized up to now, is that participants frequently oriented their eyes to different display regions before eventually reaching the target item. Even with targets in the top-left display quadrant, which yielded the fastest reaction times, participants on average required more than five fixations to land in the target region (see Figure 7). This contrasts with the above-mentioned ERP studies, which used only limited presentation times (e.g., 700 ms in Zinchenko et al., 2020a; see also Schankin & Schubö, 2010) and, importantly, disallowed eye movements (which would produce artefacts in EEG studies).[3] Arguably, disallowing eye movements would impede the evolution of generic scanning procedures, in particular, when the short exposure time does not permit extended search. Applied to the present idea of dual mechanisms underlying contextual cueing, the task demands in these electrophysiological studies are quite different to those under natural viewing, and eye-movement, conditions. In particular, the ERP task design likely forces display-specific learning (perhaps due to the need of holding individual display arrangements in working memory in order to solve the task; e.g., Ballard et al., 1995). In contrast, more natural scenarios, by allowing oculomotor scanning under conditions of unlimited display presentation, foster the acquisition of display-unspecific routines adapted to the statistical regularities in the set of search displays encountered (which is dominated by repeated display arrangements) – possibly because scanning a display is simply more efficient in terms of energy than memorizing all of them (for a similar argument about not utilizing a learned context, see Wolfe & Horowitz, 2017). Accordingly, the amount and ease of procedural learning is likely to be one of the factors that mediate if and to what extent the learnt context is constrained to the spatial vicinity of the target item, versus being rather global in scope (see, e.g., Brady & Chun, 2007, for a local hypothesis; Olson & Chun, 2002; Peterson & Kramer, 2001; Shi et al., 2013, for a global hypothesis; for a review; Tseng & Li, 2004, see Goujon et al., 2015).

LTM-based attentional capture

Our results also have implications regarding the mechanisms underlying attentional capture: they support accounts according to which capture is a function of participants' statistical learning experiences (for review, see, e.g., Chelazzi et al., 2019). The central observation is that capture is minimal, or even entirely prevented, when bottom-up salient distractors appear in display regions, or at specific locations, where they occur frequently as compared to only rarely, which has been attributed to the acquisition of pro-active distractor suppression mechanisms that curtail the attentional priority achievable by items at frequent distractor locations (e.g., Ferrante et al., 2018; Goschy et al., 2014; Sauter et al., 2018; Wang & Theeuwes, 2018). However, while these mechanisms reduce the potential of distractors at frequent locations to interfere with search, they may also yield a disadvantage in that targets presented at de-prioritized (i.e., frequent distractor) locations would require extended time to be found and further processed (e.g., Sauter et al., 2021; Wang & Theeuwes, 2018). An analogous, though in terms of effect direction, reversed pattern is also evident in contextual-cueing studies using a learning/re-location design: an initial search advantage for targets presented at original locations in repeated distractor arrangements turns into a disadvantage in a subsequent target re-location phase, as search continues to be biased toward original target locations. That is, there is a distraction – or "attentional capture" – effect that does not arise from the item at the original target location being bottom-up (physically) salient (the T vs. L's search task affords little bottom-up guidance; see, e.g., Moran et al., 2013), but rather by this item being in some way top-down preferred by acquired LT-memories. Here, we show that this distraction effect involves persistent contributions from two types of acquired LT-memory: spatial template-based memory attentionally prioritizing the (learnt) target location within specific distractor contexts, and procedural memory, i.e., oculomotor scanning routines adapted to statistical regularities in the set of visual scenes encountered.

In summary, our results suggest that oculomotor scanning strategies are adaptable to general statistical regularities prevailing in scenic environments and act as a kind of "hinge" between the abundance of visual information contained in the scene and attentional selection. That is, rather than just being the passive consequence of shifts of attention, eye movements can proactively support attentional orienting towards locations that most likely contain information to be processed further (for accomplishing the task at hand) – before the actual attentional selection takes place. Proceduralization of oculomotor scanning may be an efficient default strategy because it minimizes the (resource-costly) need to encode and retrieve scene-specific memories while also minimizing interference from other displays by reducing the amount of learned information, as well as being more adaptive to changes at locations of the scenes which are not target-relevant, thus reducing "distraction" by potentially misleading cues from scene-specific memories (in line with Olson & Chun, 2002; and Brady & Chun, 2007).

Somewhat paradoxically, the present study shows that generically optimized scanning strategies can facilitate attentional selection of *target* locations when a layout of non-target items is repeatedly encountered, manifesting in an effect that resembles "attentional capture" after a relocation of the target, originating from contextual LTM. Although we demonstrated this with respect to statistical target location learning in a hard (T vs. L's) search scenario, we conjecture that the same mechanism would be at play in similar scenarios with either frequent and infrequent target locations, or designs in which a moderately bottom-up salient "distractor" item is placed consistently within repeated arrangements of the other, non-target and target, items (where the target is positioned randomly within this arrangement, while being of a similar, low saliency as the non-target items).[4] In the first case, participants may learn generic scanpaths that seek out frequent target locations earlier or, in the second case, circumvent the locations of such frequently encountered distractor items. This would predict that when the target is then consistently "relocated" to the distractor position (effectively removing the distractor) in a given display, search performance should be impaired – because such locations are deprioritized in the generic scanpaths acquired in the preceding learning phase. A new study would be necessary to test this prediction in a paradigm that is closer to the "standard" additional-singleton paradigm employed in most extant studies of distractor handling.

Conclusion

Our results support the novel view according to which there is guidance and also misguidance of search from experience with repeated display layouts, which arises from both LT memory for specific distractor-target arrangements and LT memory for generic oculomotor procedures. We conjecture that both components aid the search-guiding, attentional-priory map (e.g., Wolfe, 2020). Our findings extend current notions of visual search by demonstrating that learnable display-generic scanning strategies actively aid search even before attentional selection takes place, by bringing locations with a higher task relevance into view earlier, as well as by minimizing costs associated with the encoding and recall of (a large number of) display-specific contextual memories while also minimizing potential interference. We conjecture that procedural scanpath learning also extends to the deprioritization of locations consistently occupied by relatively salient "distractor" items, which remains to be tested in future research.

Notes

1. Here, we just use an analogy for illustrating our new idea – of mis-guidance of visual search being an instance of a visual distraction effect that operates from acquired long-term memory (LTM). This is not to imply that necessarily the same mechanisms are at play in attentional capture by salient but task-irrelevant stimuli and attentional capture from contextual LTM.

2. Although there was a numerical difference between repeated and non-repeated displays from block 1 onwards, the CC effect was fairly unstable in blocks 1–4 and stabilized only during the subsequent blocks.
3. Of note, to make the task performable without eye movements within the brief display duration, the display items were increased in size compared to the stimulus sizes used in standard contextual-cueing experiments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by a German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) grant GE 1889/4-2 to T.G.

ORCID

Werner Seitz http://orcid.org/0000-0001-8483-5109

References

Annac, E., Conci, M., Müller, H. J., & Geyer, T. (2017). Local item density modulates adaptation of learned contextual cues. *Visual Cognition*, *25*(1–3), 262–277. https://doi.org/10.1080/13506285.2017.1339158

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*(1), 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Beesley, M., Vadillo, M. A., Pearson, D., & Shanks, D. R. (2015). Pre-exposure of repeated search configurations facilitates subsequent contextual cuing of visual search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 348–362. https://doi.org/10.1037/xlm0000033

Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & knowledge engineering*, *60*(1), 208–221. https://doi.org/10.1016/j.datak.2006.01.013

Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 798–815. https://doi.org/10.1037/0096-1523.33.4.798

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

Brockmole, J. R., & Henderson, J. M. (2006). Short article: Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, *59*(7), 1177–1187. https://doi.org/10.1080/17470210600665996

Brockmole, J. R., & Le-Hoa Võ, M. (2010). Semantic memory for contextual regularities within and across scene categories: Evidence from eye movements. *Attention, Perception & Psychophysics*, *72*(7), 1803–1813. https://doi.org/10.3758/APP.72.7.1803

Chaumon, M., Schwartz, D., & Tallon-Baudry, C. (2009). Unconscious learning versus visual perception: Dissociable roles for gamma oscillations revealed in MEG. *Journal of Cognitive Neuroscience*, *21*(12), 2287–2299. https://doi.org/10.1162/jocn.2008.21155

Chelazzi, L., Marini, F., Pascucci, D., & Turatto, M. (2019). Getting rid of visual distractors: The why, when, how, and where. *Current opinion in psychology*, *29*, 135–147. https://doi.org/10.1016/j.copsyc.2019.02.004

Chen, S., Shi, Z., Zinchenko, A., Müller, H. J., & Geyer, T. (2022). Cross-modal contextual memory guides selective attention in visual-search tasks. *Psychophysiology*, *59*(7), e14025. https://doi.org/10.1111/psyp.14025

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience*, *2*(9), 844–847. https://doi.org/10.1038/12222

Clarke, A. D., Nowakowska, A., & Hunt, A. R. (2019). Seeing beyond salience and guidance: The role of bias and decision in visual search. *Vision*, *3*(3), 46. https://doi.org/10.3390/vision3030046

Conci, M., & Zellin, M. (2022). Stimulus-driven updating of long-term context memories in visual search. *Psychological Research*, *86*(1), 252–267. https://doi.org/10.1007/s00426-021-01474-w

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390. https://doi.org/10.1016/j.tics.2006.06.011

Ferrante, O., Patacca, A., Di Caro, V., Della Libera, C., Santandrea, E., & Chelazzi, L. (2018). Altering spatial priority maps via statistical learning of target selection and distractor filtering. *Cortex*, *102*, 67–95. https://doi.org/10.1016/j.cortex.2017.09.027

Fitts, P. M. (1964). Perceptual-motor skill learning. In *Categories of human learning* (pp. 243–285). Academic Press.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. (vol. 5, pp. 7–16). American Psychological Association.

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*, 650245. https://doi.org/10.3389/fpsyg.2021.650245

Geyer, T., Zehetleitner, M., & Müller, H. J. (2010). Contextual cueing of pop-out visual search: When context guides the deployment of attention. *Journal of Vision*, *10*(5), 1–17. https://doi.org/10.1167/10.2.3

Geyer, T., Zinchenko, A., Seitz, W., Balik, M., Müller, H. J., & Conci, M. (2023). Mission impossible? Spatial context relearning following a target relocation event depends on cue predictiveness. *Psychonomic Bulletin & Review*, 1–8.

Goschy, H., Bakos, S., Müller, H. J., & Zehetleitner, M. (2014). Probability cueing of distractor locations: Both intertrial facilitation and statistical learning mediate interference reduction. *Frontiers in Psychology*, *5*, 1195. https://doi.org/10.3389/fpsyg.2014.01195

Goujon, A., Didierjean, A., & Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. *Trends in Cognitive Sciences*, *19*(9), 524–533. https://doi.org/10.1016/j.tics.2015.07.009

Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception & Psychophysics*, *72*(5), 1267–1282. https://doi.org/10.3758/APP.72.5.1267

Husain, M., & Rorden, C. (2003). Non-spatially lateralized mechanisms in hemispatial neglect. *Nature Reviews Neuroscience*, *4*(1), 26–36. https://doi.org/10.1038/nrn1005

Jiang, Y., Song, J. H., & Rigas, A. (2005). High-capacity spatial contextual memory. *Psychonomic Bulletin & Review*, *12*(3), 524–529. https://doi.org/10.3758/BF03193799

Jiang, Y., & Wagner, L. C. (2004). What is learned in spatial contextual cuing— configuration or individual locations? *Perception & Psychophysics*, *66*(3), 454–463. https://doi.org/10.3758/BF03194893

Johnson, J. S., Woodman, G. F., Braun, E., & Luck, S. (2007). Implicit memory influences the allocation of attention in visual cortex. *Psychonomic Bulletin & Review*, *14*(5), 834–839. https://doi.org/10.3758/BF03194108

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

Kunar, M. A., John, R., & Sweetman, H. (2014). A configural dominant account of contextual cueing: Configural cues are stronger than colour cues. *Quarterly Journal of Experimental Psychology*, *67*(7), 1366–1382. https://doi.org/10.1080/17470218.2013.863373

Kunar, M. A., & Wolfe, J. M. (2011). Target absent trials in configural contextual cuing. *Attention, Perception, & Psychophysics*, *73*(7), 2077–2091. https://doi.org/10.3758/s13414-011-0164-0

Liesefeld, H. R., Liesefeld, A. M., Töllner, T., & Müller, H. J. (2017). Attentional capture in visual search: Capture and post-capture dynamics revealed by EEG. *Neuroimage*, *156*, 166–173. https://doi.org/10.1016/j.neuroimage.2017.05.016

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT Press.

Makovski, T., & Jiang, Y. V. (2010). Contextual cost: When a visual-search target is not where it should be. *Quarterly Journal of Experimental Psychology*, *63*(2), 216–225. https://doi.org/10.1080/17470210903281590

Manelis, A., & Reder, L. M. (2012). Procedural learning and associative memory mechanisms contribute to contextual cueing: Evidence from fMRI and eye-tracking. *Learning & Memory*, *19*(11), 527–534. https://doi.org/10.1101/lm.025973.112

Manginelli, A. A., & Pollmann, S. (2009). Misleading contextual cues: How do they affect visual search? *Psychological Research*, *73*(2), 212–221. https://doi.org/10.1007/s00426-008-0211-1

Manns, J. R., & Squire, L. R. (2001). Perceptual learning, awareness, and the hippocampus. *Hippocampus*, *11*(6), 776–782. https://doi.org/10.1002/hipo.1093

Moran, R., Zehetleitner, M., Mueller, H. J., & Usher, M. (2013). Competitive guided search: Meeting the challenge of benchmark RT distributions. *Journal of Vision*, *13*(8), 24–24. https://doi.org/10.1167/13.8.24

Olson, I. R., & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, *9*(3), 273–302. https://doi.org/10.1080/13506280042000162

Ort, E., & Olivers, C. N. (2020). The capacity of multiple-target search. *Visual Cognition*, *28*(5-8), 330–355. https://doi.org/10.1080/13506285.2020.1772430

Pelli, D. G., & Vision, S. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Peterson, M., Mead, P., Kelly, S., Esser-Adomako, E., & Blumberg, E. (2022). Spatial context target relearning following a target relocation event: Not mission impossible. *Attention, Perception, & Psychophysics*, *84*(2), 474–489. https://doi.org/10.3758/s13414-022-02443-0

Peterson, M. S., & Kramer, A. F. (2001). Attentional guidance of the eyes by contextual information and abrupt onsets. *Perception & Psychophysics*, *63*(7), 1239–1249. https://doi.org/10.3758/BF03194537

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. 2018. *https://www.r-project.org*

Rosenbaum, G. M., & Jiang, Y. V. (2013). Interaction between scene-based and array-based contextual cueing. *Attention, Perception, & Psychophysics*, *75*(5), 888–899. https://doi.org/10.3758/s13414-013-0446-9

Sauter, M., Hanning, N. M., Liesefeld, H. R., & Müller, H. J. (2021). Post-capture processes contribute to statistical learning of distractor locations in visual search. *Cortex*, *135*, 108–126. https://doi.org/10.1016/j.cortex.2020.11.016

Sauter, M., Liesefeld, H. R., Zehetleitner, M., & Müller, H. J. (2018). Region-based shielding of visual search from salient distractors: Target detection is impaired with same- but not different-dimension distractors. *Attention, Perception, & Psychophysics*, *80*(3), 622–642. https://doi.org/10.3758/s13414-017-1477-4

Schankin, A., & Schubo, 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.

Seitz, W., Zinchenko, A., Müller, H. J., & Geyer, T. (2023). Contextual cueing of visual search reflects the acquisition of an optimal, one-for-all oculomotor scanning strategy. *Communications Psychology*, *1*(1), 20.

Sewell, D. K., Colagiuri, B., & Livesey, E. J. (2018). Response time modeling reveals multiple contextual cuing mechanisms. *Psychonomic Bulletin & Review*, *25*, 1644–1665.

Shi, Z., Zang, X., Jia, L., Geyer, T., & Müller, H. J. (2013). Transfer of contextual cueing in full-icon display remapping. *Journal of Vision*, *13*(3), 1–1. https://doi.org/10.1167/13.3.1

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

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

Stokes, M. G., Atherton, K., Patai, E. Z., & Nobre, A. C. (2012). Long-term memory prepares neural activity for perception. *Proceedings of the National Academy of Sciences*, *109*(6), E360–E367. https://doi.org/10.1073/pnas.1108555108

Tseng, Y. C., & Li, C. S. R. (2004). Oculomotor correlates of context-guided learning in visual search. *Perception & Psychophysics*, *66*(8), 1363–1378. https://doi.org/10.3758/BF03195004

Vadillo, M. A., Giménez-Fernández, T., Beesley, T., Shanks, D. R., & Luque, D. (2021). There is more to contextual cuing than meets the eye: Improving visual search without attentional guidance toward predictable target locations. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(1), 116. https://doi.org/10.1037/xhp0000780

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

Wang, B., & Theeuwes, J. (2018). Statistical regularities modulate attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 13. https://doi.org/10.1037/xhp0000472

Wolfe, J. M. (2020). Visual search: How do we find what we are looking for? *Annual Review of Vision Science*, *6*(1), 539–562. https://doi.org/10.1146/annurev-vision-091718-015048

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, *28*(4), 1060–1092.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 58. https://doi.org/10.1038/s41562-017-0058

Yang, Y., Coutinho, M. V., Greene, A. J., & Hannula, D. E. (2021). Contextual cueing is not flexible. *Consciousness and Cognition*, *93*, 103164. https://doi.org/10.1016/j.concog.2021.103164

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

Zelinsky, G. J. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research*, *36*(14), 2177–2187. https://doi.org/10.1016/0042-6989(95)00300-2

Zellin, M., Conci, M., von Mühlenen, A., & Müller, H. J. (2013). Here Today, Gone Tomorrow – Adaptation to Change in Memory-Guided Visual Search. *PLoS One*, *8*(3), e59466. https://doi.org/10.1371/journal.pone.0059466

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. (2020b). Distributed attention beats the down-side of statistical context learning in visual search. *Journal of Vision*, *20*(7), 4–14. https://doi.org/10.1167/jov.20.7.4

Zinchenko, A., Conci, M., Müller, H. J., & Geyer, T. (2023). Environmental regularities mitigate attentional misguidance in contextual cueing of visual search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication, https://doi.org/10.1037/xlm0001297

Zinchenko, A., Conci, M., Töllner, T., Müller, H. J., & Geyer, T. (2020a). 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

Study 3 – Seitz et al., (in preparation)

Seitz, W., Zinchenko, A., Müller, H. J., Geyer, T., (in preparation). Learning how: A ‘mindless’ procedure alone gives rise to contextual-cueing – a weakly supervised connectionist model of statistical context learning in visual search

Author Contributions

WS conceived and designed the experiments; **WS** analyzed the data; HJM contributed to conceptual analysis and discussion; **WS**, AZ, HJM, TG wrote the paper.

Learning how: A ‘mindless’ procedure alone gives rise to contextual-cueing – a weakly supervised connectionist model of statistical context learning in visual search

Werner Seitz[1, 2], Artyom Zinchenko[1,], Hermann J. Müller[1,3], & Thomas Geyer[1,3,4]

[1]Department Psychologie, Ludwig-Maximilians-Universität München, Munich, Germany

[2]Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität München, Germany

[3]Munich Center for Neurosciences – Brain & Mind, Ludwig-Maximilians-Universität München, Germany

[4]NICUM - NeuroImaging Core Unit Munich, Ludwig-Maximilians-Universität München

Correspondence to:

Werner Seitz

Department Psychologie

Lehrstuhl für Allgemeine und Experimentelle Psychologie

Ludwig-Maximilians-Universität München

Leopoldstraße 13

80802 München

Germany

E-Mail: seitz-werner@hotmail.de

Keywords

Visual search, selective attention, statistical learning, contextual cueing, Hebbian Learning, computational modeling

Abstract

Because our environment is not random, it is beneficial to assimilate the statistics of sensory impressions and improve performance, such as visual search for a target object in a cluttered array of non-target objects (contextual cueing – CC – effect). Computational models of CC have so far focused on predicting the target location from a particular configuration of non-target items. This contrasts with recent findings according to which display repetitions train human participants' general procedures for the search task. Here, we test the latter idea by employing a connectionist model of visual search that exclusively learns a search procedure without acquiring any individual display-layout information. We show that an instance of a "learning how" mechanism not only proposes a viable alternative account to existing "learning that" mechanisms, but also generates more plausible key behavioral metrics and exhibits a central bias as an emergent phenomenon of learning-induced plasticity. These findings have implications for models of visual search and artificial intelligence: Learning a procedure from leveraging a task's structure alone can mimic the effects of top-down modulation of attention, while also reducing the need for supervision in learning, thereby making computational models that leverage procedural learning behaviorally more plausible and easier to train.

Introduction

Detecting, and responding to, specific objects in a cluttered sensory world is one of the most prevalent tasks of the human visual system (Wolfe, 1998; Wolfe & Horowitz, 2017). Since environments are typically stable, it is beneficial to extract information that is predictive of relevant ‘target’ objects and use this information to facilitate search on later occasions. For example, looking for a certain item in a shop for the first time is driven by knowledge about the item’s featural (size, colour etc.) properties. In subsequent searches for the same item in the same shop, another source of information becomes available: knowledge about the location of the target item on a certain shelf in the presence of other, distractor (i.e., non-target) items that, while not directly relevant for the task at hand, nevertheless provide a helpful context for finding the searched-for item.

An experimental paradigm that mimics this situation in the laboratory is visual search for a target item placed at a stable position within a constant – i.e., repeated – spatial arrangement of distractor items – originally introduced by Chun and Jiang (1998). The basic task requires participants to search through artificial arrays of letter-type stimuli, finding and responding to the orientation of a T-shaped target (rotated by either 90° or 270°) among (orthogonally oriented) L-shaped distractor items. Critically, search performance for repeated distractor-target arrangements – in which the distractor layout, or ‘context’, is predictive of the target location – is compared against non-repeated (‘baseline’) arrangements – in which distractor locations are arranged anew on each trial. The target locations themselves are fixed in both repeated and non-repeated displays, controlling for absolute target-position learning (cf. Geng & Behrmann, 2005; Jiang, Swallow, & Rosenbaum, 2013). This laboratory task thus allows the isolated study of *spatial* context learning, unconfounded by semantic and syntactic properties of the searched scenes (for evidence that such properties matter, see Vo & Wolfe, 2013). Chun and Jiang (1998; 2003) found the search reaction times (RTs) to decrease with increasing practice on the task not only for repeated but also for non- repeated display arrangements – reflecting procedural learning of how to best perform the task, akin to the acquisition of a (search) skill (Fitts, 1964; Fitts & Posner, 1967). Critically, the practice-related improvement turned out larger for repeated vs. non-repeated displays – which has been attributed to the beneficial effect of repeated distractor contexts for finding the target. Chun and Jiang (1998) proposed that repeatedly encountered context-target spatial relations are stored in spatial long-term (LT) memory. Upon re-encountering a repeated display, the appropriate memory is retrieved and guides, or 'cues', search to the target location – giving rise to the 'contextual-cueing' (CC) effect.

Central to accounts of CC is the more efficient allocation of focal attention to the target item in repeated displays (Chun & Jiang, 1998; for reviews, see Goujon et al., 2015, Sisk et al., 2019). In support of this, oculomotor investigations of CC have shown that detecting the target in repeated displays requires fewer eye movements (i.e., a reduced number of fixations, NF, where, in a fixation, covert and overt attention are assumed to be aligned) compared to the baseline of non-repeated displays (Tseng & Li, 2004; Peterson & Kramer, 2001). Relatedly, oculomotor scanpaths – i.e., the total distance traversed by sequential eye movements – are also shorter for repeated displays (e.g., Manginelli & Pollmann, 2009; Zang et al., 2015; Brockmole & Henderson, 2006).

Dominant theoretical accounts of CC attribute this pattern of effects to the build-up of contextual memories that associating the location of the target with the distractor layout within specific repeated display arrangements (Chun & Jiang, 1998; Tseng & Li, 2004; Beesley et al., 2015). In terms of Ryle (1945), such theories of CC may be considered instances of “knowing that”, i.e., what is acquired is some kind of (search-guiding) proposition that, given a certain distractor layout, the target is to be found at location [x,y]. Complementary to this would be accounts of “knowing how”: the capability to optimally perform specific tasks or actions (Ryle, 1945). Arguably, the CC paradigm is particularly well suited to investigating the latter type of account, which considers the refinement of the requisite – i.e., overall optimal – search procedure to be driven more strongly by repeated than by non-repeated displays. In line with this notion, recent analyses of the oculomotor scanpaths observers traverse to find the target show that procedural learning of how to the search the set (or ‘world’) of displays encountered is an important determinant of the CC effect (Seitz et al., 2023; Seitz et al., 2024): while scanpath similarity increases generally – across all types of displays – with training, critically, it increases more for repeated than for non-repeated displays. This supports a “knowing-how” account of CC, according to which statistical learning optimizes the processes by which the search task is accomplished, with repeatedly encountered contexts simply accruing a greater weight in tuning these processes than non-repeated contexts which, by definition, are searched only once.

Given this evidence, the motivation of the present study was to test computationally whether procedural learning of how to optimally search of a certain ‘world’ of visual displays can give rise to a CC effect in the absence of any display-specific learning, i.e., without associating a specific distractor arrangement with a specific target location. Along with this, we set out to test whether an only weakly supervised model – leveraging the implicit structure of the task rather than relying on explicit teaching signal (provided by an external supervisor) – could still exhibit adaptive behavior. If so, this would have implications for a “second” type of learning in AI and hold the possibility of designing/collating new datasets with a smaller burden on explicit annotation.

Previous models of CC

Based on the findings of faster response times (RTs) and fewer fixations required for detecting targets in repeated contexts, there have been several attempts to model the ‘cueing’, or guidance, of attention to learnt target locations by acquired long-term contextual-memory representations. Brady and Chun (2007) proposed a two-layer feedforward network which learned to associate a specific context with a specific target position. The input layer consisted of an 8 by 6 (location) matrix, with ‘occupied’ matrix cells encoding the presence of 1 out of a total of 12 display items at the respective ‘search-display’ location. Inputs triggered by a specific display arrangement then activated (via a bottom-up activation term) spatially corresponding neurons in an output layer that was considered to represent a ‘priority map’ (cf. Fecteau & Munoz, 2006) for the deployment of attention and eye movements. The input and the output layer were connected by weights that were adjusted by the delta rule each time the target was detected: weight updating was largest for the target location and (exponentially) descending for distractors with increasing spatial distance from the target location. The network’s performance was measured in terms of the number of activated output-layer units that had to be checked in decreasing order prior to arriving at the unit representing the target location. This model was well able to generate a CC effect, implying that the effect arises from a reduction in the number of fixations required to find the target in repeated display layouts. An augmented version of this basic two-layer network was proposed by Beesley et al. (2015) to account for the observation of facilitated RTs even when the target location is allowed to vary relative to an invariant distractor layout (though the facilitation is smaller than the ‘standard’ CC effect with fixed target locations; e.g., Vadillo et al., 2021; Kunar & Wolfe, 2011). To accommodate this finding, Beesley et al.’s model learns an additional set of ‘auto-associative’ weights between distractors within the input layer, reinforcing the distractor configuration.

Thus, in these modelling attempts, associations are formed – i.e., connections are strengthened – between the spatial configuration of items in the input layer and the location of the searched-for target in the ‘attentional-priority’ output layer in a supervised manner (for an overview, see Dayan & Abbott, 2005); additionally, associations may be acquired among repeatedly placed distractor items within the input layer using unsupervised learning. From a biological perspective, Goujon et al. (2015) put forward the related idea that Spike-Timing-Dependent Plasticity (STDP) – the neurobiological mechanism of associative Hebbian learning – alone might account for the acquisition of CC.

Present work

While associative learning (Goujon et al., 2015) presents a plausible mechanism for generating contextual facilitation as a form of statistical learning, the levels at which statistical regularities are actually leveraged have as yet not been comprehensively investigated, at least in terms of computationally explicit models. Empirically, recent work (Seitz et al., 2023; Seitz et al., 2024) has demonstrated that contextual facilitation – rather than being simply attributable to the acquisition of specific, repeatedly encountered distractor-target relations (display-specific learning) – may arise from procedural, oculomotor-scanpath learning adapted to the set of encountered display layouts at large (display-generic learning), i.e., the honing of a general search ‘skill’ in terms of Fitts and Posner (1967). Of note, display-specific learning as such could account only for the acquisition of a search advantage for repeated over non-repeated displays. In contrast, acquiring a generic procedural strategy to optimally scan the entire set of encountered displays (including both repeated and non-repeated layouts) can provide a parsimonious account of both contextual facilitation (i.e., the search advantage for repeated vs. non-repeated displays) and the general improvement in performance across time-on-task, which is evident with both repeated and non-repeated displays: while scanning is optimized for all types of display (accounting for the general improvement in performance), repeated displays are encountered multiple times and thus shape the scanning strategy more than non-repeated displays that are encountered only once. As a result, the similarity of the oculomotor scanpaths becomes higher between pairs of (differently composed) repeated displays compared to pairs of (different) non-repeated displays (see Figure 1A). This proposal of display-generic (rather than display-specific) learning is in line with Lashley's (1951) account of sequential behavior, according to which motor actions are not planned de novo each time, but instead previously executed movement patterns are modified or updated to fit the demands for the task at hand (see also Rosenbaum et al., 2007).

Based on this idea and the evidence from measures of scanpath similarity (Seitz et al., 2023, Seitz et al., 2024), we implemented a computational model which, by its very architectural constraints, cannot form any contextual associations between – or 'representation' of – the distractor-layout (input-layer pattern) and a target location (output layer) at all; rather, it can *only* learn fixation sequences based on previously executed saccades. That is, the model updates prior oculomotor patterns based on the current input, as proposed by theories of procedural learning (Lashely, 1951). Beyond this theoretical motivation, choosing this simple procedural learning rule was also underpinned by a pragmatic consideration: executing stereotypical eye movements driven by locations of relevance in the current input, as well as indirectly by those of previously ‘scrutinized’ displays, would, in itself, provide a fairly economical way of optimizing where and in which sequence fixations should be placed, given that the fixations will be tuned to general characteristics of relevance in the set of displays and thus be likely to capture the ‘salient’, evolutionarily informative aspects of the input.

Consequently, the current study was designed to examine the explanatory reach of this simple procedural learning mechanism, vis-à-vis the mechanisms implemented in previous computational theories of CC (Brady & Chun, 2007; Beesley et al., 2015). Importantly, the assumptions underlying the present scheme differ fundamentally from those of the previous models, which learn target locations as a function of context: the models inspect likely candidate positions guided by initially erroneous ‘memory signals’, that, across repeated searches, become gradually biased towards the target location. In contrast, the model investigated here does not acquire any associations between an item configuration and the target location predicted by that configuration; instead, it optimizes its behavior as an emergent property of reinforcing previously executed oculomotor patterns.

Thus, the present modeling attempt present a fundamental departure from previous schemes: Employing a computational model which, by design, can *only* learn a scanning procedure and *no* spatial-layout information at all allows us to examine whether contextual facilitation can arise merely as byproduct of an acquired, essentially ‘mindless’ generic search procedure. If so, the model results would demonstrate that to account for contextual facilitation, it is *not necessary* to assume the build-up of higher-level contextual-memory ‘templates’ in spatial long-term memory that are activated by the presentation of a specific repeated display and then top-down bias search towards the target location pointed to by the distractor configuration. Of note, our model incorporates the notion of CC being a mere byproduct of associative learning (cf. Goujon et al., 2015), though with associative learning in the model occurring simply between successive fixations, rather than extracted regularities in the item arrangement.

In terms of behavioral indices of CC, following previous studies (Brady & Chun, 2007; Beesley et al., 2015), we focused on the key measure of the number of fixations required by the model to detect the target. Further, for a more comprehensive validation of our computational model against the evidence of CC being driven by procedural learning (Seitz et al., 2023), we additionally examined the model's search behavior in terms of scanpath-similarity measures, i.e., the homogeneity of the scanpaths within 'participants' and between displays. If procedural learning alone suffices to produce CC, we expected the model's behavior to replicate previous findings diagnostic of procedural learning, in particular: a reduced number of fixations as well as a higher similarity of the scanpaths in later vs. earlier blocks of learning, especially for repeated vs. non-repeated displays.

Methods

Model architecture and learning

The model was implemented in PyTorch (Paszke et al., 2019), benefitting from CUDA (NVIDIA, Vingelmann, P., & Fitzek, F. H. P., 2020) optimized code, and trained on an NVIDIA RTX A4000 GPU. The model has two layers, an input ('search-display') layer and an output ('priority-map') layer, connected by weights – similar to Brady and Chun's (2007) model. In contrast to their model, however, the current model does not associate the activations of a given input display with the target location in the output layer in a supervised learning scheme. Hence, it does *not* and *cannot* learn to associate a given configuration of display items with the target location – in fact, it is never even informed about which item corresponds to the target.

The architecture of the model is illustrated in Figure 1B. While there is full connectivity between each neuron in the input layer and each neuron in the output layer, there are two distinct sets of weights. The first set are ‘direct’ weights, connecting spatially corresponding positions in the input and output layers (i.e., the diagonal of the weight matrix): these weights take on a fixed (pre-set) value, which does not change as a result of learning. Essentially, they provide a “bottom up” signal that represents the item locations in the output layer (with equal “priority”), similar to the bottom-up weights in Brady-and-Chun (2007). All other, off-diagonal connections (the second set) are learnable, with their weights initially set to zero. Thus, initially, each display (distractor and target) item receives an equal amount of activation on the output layer. By disabling changes of the diagonal weights, the network is prevented from learning to increase the weight between a given target location in the input layer and the target location in the output layer, as well as between sets of distractor elements and the target. That is, the network cannot use target-location information to optimize oculomotor scanning.

The model computes an oculomotor scanpath, i.e., a sequence of fixation coordinates on the output layer, starting at the display center and then successively sampling one position after the other until the target is located (similar to the model of Koch & Ullman, 1984). On each iteration, the model either selects the output neuron (item location) exhibiting the highest activation as the next fixation coordinate, or, if there are several output neurons with equal activation (before learning all of them), it randomly selects one of those. Next, if the selected item is not the target, the selected item is removed from the input layer (as a result of which it is no longer represented on the output layer and, so, its location can no longer be selected for fixation) – implementing an ‘inhibition-of-return’ mechanism (e.g., Klein & MacInnes, 1999). Scanning is terminated once the model ‘selects’ the target location – though, as already stated, the network itself is not informed whether or not the selected element is the searched-for target item.

Initially, before learning becomes effective, the model must sample approximately half the number of items in the display to locate the target – as is typical for ‘unguided’ search tasks, such as (T-type target vs. L-type distractor) letter search (e.g., Wolfe, 2021). Critically, as the network is not informed that the target item was selected (search is then simply terminated), it has no means to learn that a given context implies a certain target location. Following search termination, the next display is presented as input to the network.

Learning takes place in the second set of weights, i.e., the off-diagonal connections of the weight matrix. The network increases the weights between the current fixation location’s region and the region centered around the location selected for the next fixation according to a Hebbian learning rule, i.e., by changing of the weights connecting the pre- (input) and postsynaptic (output) neurons proportional to the neurons’ activations (for visual illustration, see Figure 1C). Of note, for each fixation location, neighboring (output) neurons receive lateral activation, too (with the amount of activation scaled by distance; see below). While this is physiologically plausible, essentially implementing a kind of population encoding (e.g., Behan & Kime, 1996), it enables the network to generalize its behavior to similar, rather than only identical, input patterns: if a saccade from a current region to another region is possible, which attracted similar fixations to previously executed eye movements, the network will be biased to execute this saccade. Specifically, the activations of ‘lateral’ neurons around the current and selected-to-be-next fixation locations are scaled by a negative exponential decay function of the distance to the respective fixation location, and the change in weights is proportional to the activation between each input and output neuron, according to the following equation:

$$\mathit{\Delta} w_{i,j} = \left(e^{-\varepsilon\left(i,f(t)\right)/\tau_{in}}\right) * \left(e^{-\varepsilon\left(j,f(t)/\tau_{out}\right.}\right) * lr,$$

whereby

$$\varepsilon\left(k, f(l)\right) \ = \ \sqrt{\left(x_k - x_{f(l)}\right)^2 + \left(y_k - y_{f(l)}\right)^2},$$

where $\mathit{\Delta} w_{i,j}$ is the change of a particular off-diagonal weight between neuron i in the input layer and neuron j in the output layer and $\varepsilon\left(i, f(t)\right)$ is the Euclidean distance of neuron i to the current fixation location $f(t)$ in the input layer, while $\varepsilon\left(j, f(t+1)\right)$ is the distance of neuron j to the selected next fixation location $f(t+1)$ in the output layer. The constant τ_{in} denotes a decay constant which regulates the amount of lateral activation as a function of the distance ε of neuron i from the current fixation location $f(t)$; analogously, τ_{out} is a decay constant that regulates the amount of lateral activation of neuron j as a function of the distance ε to the selected next fixation location $f(t+1)$, while lr is the learning rate. Different values of τ correspond to different amounts of spatial generalization/selectivity. The amount of lateral activation in a layer as a function of distance and its dependence on τ is illustrated in Figure 1D. Critically, the network had no information about which element is a distractor and which a target item. It only learned to associate fixation locations based on consecutive fixation events when searching repeated and non-repeated arrays.

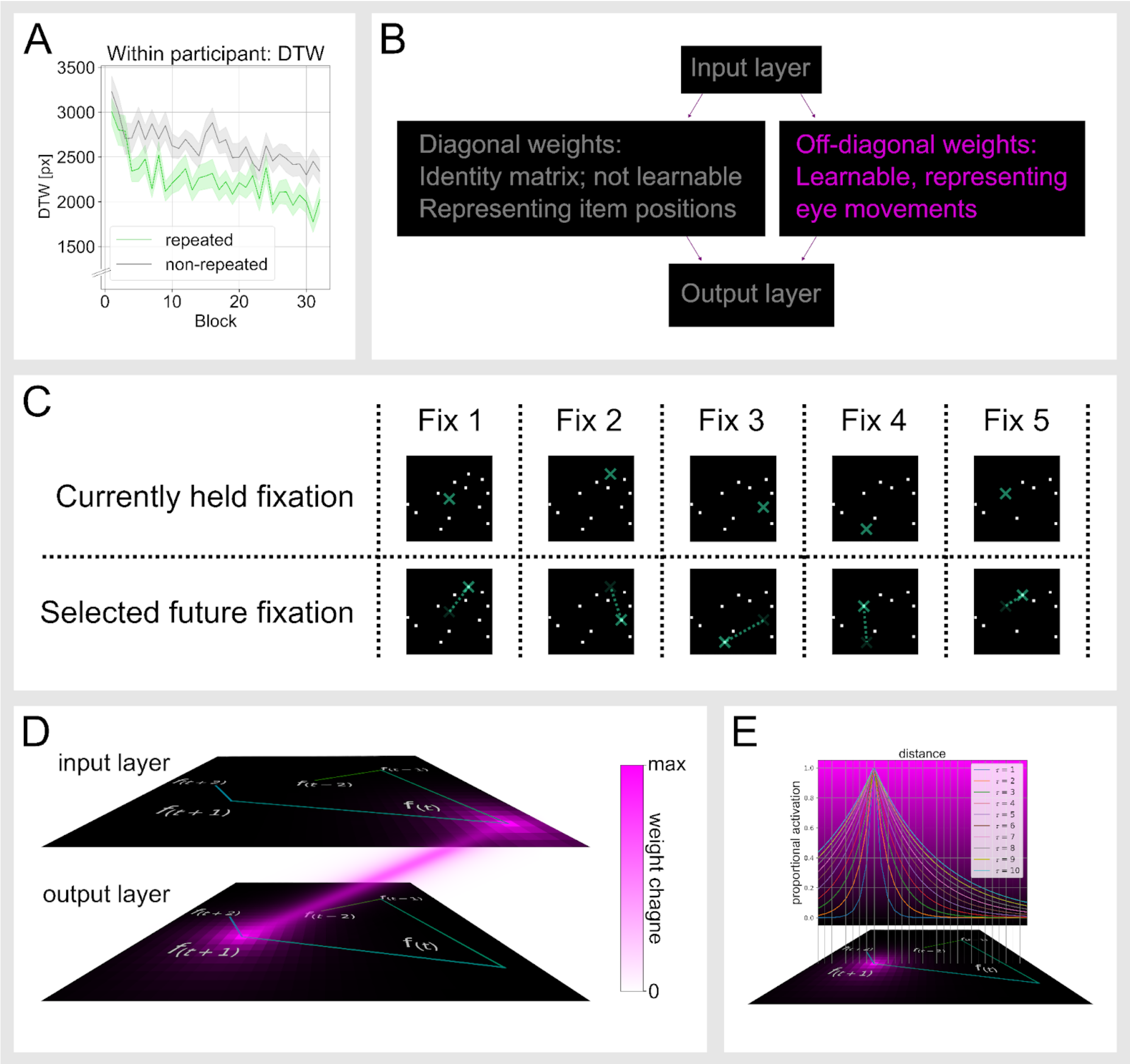

Figure 1. Figure 1A shows real human data (from Seitz et al., 2023) performing visual search in repeated and non-repeated target-distractor layouts and a practice-, i.e., block-, related increase in scanpath similarity (measured by Dynamic Time Warping – DTW; see, e.g., Fahimi & Bruce, 2020), which increased with increasing time-on-task (learning), with a greater increase for repeated displays. Figure 1B illustrates the architecture of the network examined in the present investigation, which consisted of two layers that are connected by two sets of (fixed and, respectively learnable) weights. Figure 1C outlines the overall dynamics of the model: The initial fixation is set at the center and a particular item is selected as the next fixation (location), which is then ‘removed’ from the input prior to determining a new fixation location; this iterative selection and rejection process is terminated upon a fixation falling on the target. Figure 1D visualizes the amount of weight change resulting from Hebbian learning between laterally activated neurons around the current fixation (in the input layer) and those of the selected next fixation (output layer). Figure 1F illustrates the decreasing amount of lateral activation of a given neuron as a function of its distance from the fixation location. Smaller values of τ yield a faster decrease in activation.

Training procedure

The network was initialized 25 times with a different (randomly generated) set of displays, corresponding to 25 “participants”. Each network underwent 32 training blocks, each consisting of 12 repeated/old-context and 12 non-repeated/new-context trials – mimicking Chun and Jiang’s (1998) original study of CC. The display matrix consisted of 31×31 possible locations, where, in each trial display, 11 locations were occupied by L-shaped distractor items and 1 location by a T-shaped target item. Target locations were distributed randomly, except that three (of the 12) repeated and three (of the 12) non-repeated displays had targets in each of the four display quadrants, and thus avoiding target-quadrant effects attributable to absolute target-location probability cueing (e.g., Jiang et al., 2013). Both the target and the distractor items activated the matrix element corresponding to the respective item location to the value of 1 in the input map. Note that we choose a relatively large 31×31 matrix to ensure that the network had sufficient degrees of freedom to produce variability in the spatiotemporal profile that its ‘eyes’ traverses within a given repeated and non-repeated display layout. This was a departure from Brady and Chun (2007), who only had 8×6 possible display locations to model selection of individual target and non-target items.

Results

We investigated whether learning idiosyncratic oculomotor scanning strategies suffice to elicit a CC effect both in terms of the number of fixations required to detect the target item and the consistency, or similarity, of the oculomotor scanpaths traversed, using Dynamic Time Warping (DTW) computed between each pair of individual repeated-display and, respectively, pairs of non-repeated-display scanpaths (each display containing a unique target position) in each block, using

Python's similarity-measure library (see Jekel et al., 2019). DTW quantifies the similarity of the shapes of the scanpaths with distinct time series by aligning them in the time domain, thus minimizing the Euclidean distance between the aligned series.

Analyses were performed using repeated-measures ANOVAs with the factors Context (repeated, non-repeated) and Block (1-32) – consistent with Seitz et al. (2023; 2024). For comparisons to existing data from the literature (see section: "Comparison to empirical data"), we collapsed 8 blocks of trials into 1 single Epoch, yielding a total of 4 epochs (as the studies considered also analyzed/ presented their data across 4 epochs). The results reported below were obtained from a network with a $\tau_{in} = 2.5$ and $\tau_{out} = 4.5$ and a learning rate of $5*10^{-7}$. The specific values of τ_{in} and τ_{out} were determined empirically (by selecting from simulations according to a grid-search scheme), so as to obtain behaviorally plausible values for the number of fixations typically observed in eye-movement studies of CC with stimuli generated analogously to the procedure implemented here (e.g., Peterson & Kramer, 2001; Manginelli & Pollmann, 2009; Zhao & Ren, 2020): collectively, these studies found that even after a reasonable amount of practice with repeated displays, the number of fixations required by human learners to home in on the target remained relatively high (typically some 4–6 fixations), rather than the target being among the very first items to attract an eye movement.

Analysis of simulated oculomotor parameters

The network showed a significant improvement in the number of fixations required to reach the target for both repeated and non-repeated displays (main effect of Block, F(31, 744) = 19.85, p < 0.001, ηp^2 = 0.45, 95% CI[0.39, 0.49]). Additionally, the network required significantly fewer fixations, on average, to reach the target in repeated vs. non-repeated displays (5.06 vs. 6.18 fixations; main effect of Context,

$F(1, 24) = 34.32$, $p < 0.001$, $\eta p^2 = 0.59$, 95% CI = [0.31, 0.75]). Further, there was a significant interaction between Context and Block ($F(31,744) = 5.35$, $p < 0.001$, $\eta p^2 = 0.18$, 95% CI = [0.11, 0.20]; see Fig. 2A, left panel), due to the Context, or contextual-facilitation, effect increasing with practice (as a result of a more marked improvement of performance across blocks for repeated vs. non-repeated displays).

Previous research (Seitz et al., 2023) has shown that repeated displays not only require fewer fixations, but also that the oculomotor scanpaths become more similar as the experiment progresses, and more so for repeated than for non-repeated displays. This effect pattern arises because the oculomotor search strategy itself is being adapted, with repeated displays weighing in stronger on the optimization of the search strategy than non-repeated displays (Seitz et al., 2023, Seitz et al., 2024). Since our model *only* learns a procedure – namely, to produce a stereotypical sequence of saccades based on previous fixational patterns –, we tested whether the scanpaths would become more similar in our network across blocks and, importantly, whether this effect would be more pronounced for the repeated displays. The results revealed (DTW) scanpath similarity to increase as the experiment progressed (main effect of Block: $F(31, 744) = 42.82$, $p < 0.001$, $\eta p^2 = 0.64$, 95% CI = [0.60, 0.67]), and the scanpaths for repeated displays to become more similar than those for non-repeated displays (main effect of Context, $F(1, 24) = 43.82$, $p < 0.001$, $\eta p^2 = 0.65$, 95% CI = [0.39, 0.78]). Additionally, there was a significant Context × Block interaction ($F(31, 744) = 4.39$, $p < 0.001$, $\eta p^2 = 0.15$, 95% CI = [0.09, 0.17]; Fig. 2A, right panel), evidencing a more marked increase in scanpath similarity across blocks for repeated vs. non-repeated displays. Of note, we obtained the same result pattern in a second run (see Fig. 2B) when generating a new set of non-/repeated displays, thus rendering it unlikely that the first set of results was driven by specific properties inherent in the displays in the first run. This pattern mirrors the human data.

Comparison to empirical data

We went on to examine how well our model fits to empirical data. For repeated displays, the median number of fixations was above 4, while it was above 5 for non-repeated displays. These numbers match previous reports fairly well, including Peterson and Kramer's (2001) 'benchmark' oculomotor study of CC. Figure 2B gives the median number of fixations of the simulation as well as that reported by Peterson and Kramer (2001). To assess the fit quantitatively, we computed the Mean Absolute Error between the model's results and Peterson and Kramer's data for both runs, which was 0.3 fixations (largest deviation = 0.5, relative error = 5%) for run 1 and 0.2 fixations (largest deviation = 0.6, relative error = 3%) for run 2.

Next, we fitted a curve to the contextual-facilitation effect produced by the model (see Brady & Chun, 2007, for a similar procedure), which was best described by a power function (see Chun & Jiang, 2003). The variance explained was 98%. Moreover, to assess how well the characteristics of the obtained curve translate into empirical reaction times, we compared the model to the data reported by Geyer et al. (2023), who tested a large sample of (N=45) participants, yielding a robust measure of CC. Assuming that the benefit of one fixation corresponds to a reaction-time speed-up of 25 ms (Wolfe, 1994; Brady & Chun, 2007), the saturation curve explains 58% of the variance in the RT data of Geyer et al. (2023), further validating our model. The curve is defined by:

$$y = 16.30 * \left(1 - e^{\ (-x/0.31)}\right) - 14.97,$$

where y corresponds to the magnitude of the CC effect per epoch, denoted as x. Assuming that contextual facilitation approaches an asymptotic value across epochs is also more plausible compared to a (monotonically increasing) logarithmic relationship (Brady & Jiang, 2007). Fixation numbers and reaction times have a natural lower bound (i.e., ≥ 0) and their characteristics over time are best described by power functions (e.g., Fitts & Posner, 1967; Anderson, 1982; Chun & Jiang, 2003). Consequently, rather than growing unboundedly, the difference between increasingly smaller values must approach some asymptotic value. Figure 2D summarizes our findings.

Central Bias as emergent phenomenon of procedural learning

Finally, we examined *where* our models 'looks' after training, by summing up the learnt weights for individual (output) locations across displays and participants. While one might have expected the network to display a strong bias towards individual target locations, surprisingly, it actually developed a bias to scan particularly central display locations (not necessarily coinciding with target locations) – an emergent behavior that resembles human visual exploration (e.g., Clarke & Tatler, 2014). This is illustrated in Figure 2D, which shows the summed weights of randomly selected "participants". As can be seen, the weights do not project in a 1:1 fashion to single target locations in repeated and non-repeated arrays; instead, they cluster around the central display region. We found such a bias also in other (control) simulations, in which we had the model start search on each trial at a peripheral (rather than the central) display location, e.g., the location in the top-left display quadrant (coordinates: $x = 3$,$y = 3$ in the 31×31 matrix). It is thus unlikely that the central bias is a particular consequence of our training procedure, with search, by default, starting at the display center (coordinates:$x = 15$,$y = 15$). Rather, the central bias is likely to reflect training-induced plasticity in the process of visual search akin to human behavior.

I support of this, fitting a bivariate Gaussian (with a least-squares approximation) to the average bias of all “participants”, along the lines of Clarke and Tatler (2014), we found that a truncated Gaussian could indeed describe the computational data well, with the center of the Gaussian located at the display-center coordinates $x = 15.4$, $y = 14.5$, with a standard deviation of 10.5 and 10.1 in the x and y directions, respectively. To assess the goodness of fit, we computed the mean average error between the bias values and the fitted Gaussian, divided by the mean bias value, which yielded a relative error of 1.7%.

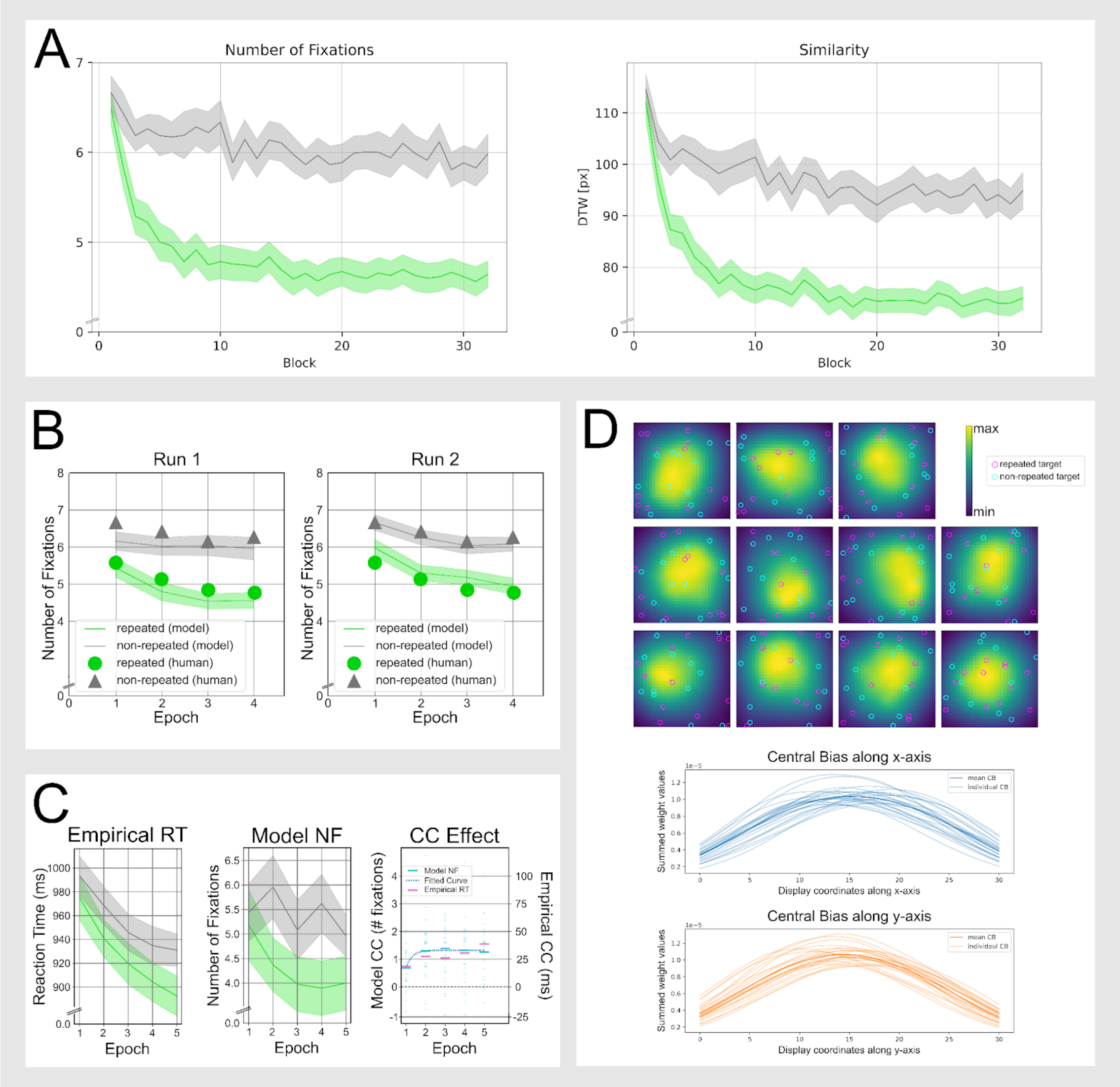

Figure 2. Figure 2A depicts the mean number of fixations (left panel) and the (DTW) similarity (right panel) of the scanpaths generated by the network, showing a Block x Context interaction for both measures. The lines in Figure 2B show the median number of fixations of the model (“Run 1”), as well as a second, independent version of the experiment (“Run 2”) conducted to validate our findings. The circular and triangular markers show the median number of fixations from Experiment 1 of Peterson and Kramer (2001). The left plot of Figure 2C depicts the empirical RT data from Geyer et al.’s (2023) large sample of participants; the middle panel of Figure 2C depicts simulated fixations of our model and the right panel shows the CC effect rendered by the model and human participants (from Geyer et al., 2023), respectively, as well as the curve fitted to the model data. Individual model “participants” are visualized in light blue. Figure 2D illustrates the bias of the weights towards the display center, as an emergent property of procedural learning. The upper panel shows the weights of 11 randomly selected “participants”, while the middle and lower panels show the bias in the x- and y-direction, respectively. Note: Projections of the fitted 2D Gaussian are not shown, because, due to the good fit, they would visually occlude the mean central-bias curves.

Discussion

The present study demonstrates that acquiring a display-general oculomotor scanning procedure in a visual search task can yield a CC effect even in the absence of any configural learning, i.e., associating a specific target location with a specific configuration of the distractor elements. Our model merely learns a simple procedure: if a saccade from a given fixation location to another location (among several alternatives) is possible and the model has previously executed a similar saccade, the model is biased to execute that saccade. While this limits performance, in that non-target elements will inevitably also be visited during search, the saccades ultimately terminate at the task-critical target locations. The network thus learns to generalize its search behavior in such a way that many saccades are biased towards target locations – in line with previous computational models of the CC effect (Brady & Chun, 2007; Beesley et al., 2015). Critically, though, our network focuses on the learning of entire series of scanpaths, rather than learning to prioritize individual display locations.

While our results demonstrate that procedurally learnt, display-general oculomotor scanning routines go a long way to account for CC, they do not rule out other mechanisms – of top-down search 'guidance' – proposed in the CC literature: these assume associative learning of specific, repeatedly encountered spatial distractor-target relations (LTM 'templates'), which, when activated by a given repeated display, bias the deployment of attention towards the target location (e.g., Wolfe, 2020), possibly augmented by facilitated response-selection and/or -execution processes (e.g., Kunar et al., 2007).

Of note, our network does not claim neurobiological plausibility (such as endeavoring to model the cellular operation of neurons in the superior colliculus; e.g., van Opstal & van Gisbergen, 1990). Rather, it was designed to provide a computationally explicit implementation of a hypothesis according to which the mere acquisition of a simple, “mindless” procedure suffices to generate a CC effect – thus, casting doubt on the necessity to assume (learnt) display-specific LTM representations to account for contextual facilitation. Of importance in this regard, our model produces a reasonable fit to empirical data obtained from human participants; in particular, at the end of learning, the number of fixations required to find the target remains above 4 or 5 for repeated and non-repeated displays, respectively – rather than the empirically implausible 1 or 2 fixations estimated by Brady and Chun’s (2007) model for repeated displays. Further, we observed a central bias as an emergent property of procedural learning. While it is thought that a central bias can arise from training with specific target locations (Parkhurst et al., 2002), our model suggests that just performing “mindless” scanning of the visual displays may suffice for the bias to emerge – in line with the observation of a central bias even in (“target-less”) free-viewing tasks (Carnosa et al., 2003).

Crucially, here, we provide a procedural account of learning which can render performance improvements in visual search that are not contingent on the learning of, and attendant ‘cueing’ of search towards, specific target locations. Procedural learning, in particular of oculomotor trajectories, thus appears to be an important factor that has so far not been explicitly considered in models of visual search such as Guided Search (e.g.; Wolfe, 2021). This also implies that eye movements, rather than being just the consequence of ‘targeted’ shifting of attention, actually promote the orienting of attention toward display regions that most likely contain relevant information for attentional processing – prior to the selection of specific locations/items based on bottom-up or top-down ‘cues’. Following from this, we propose that procedural learning plays an active role in visual search in the sense that convenient, general search procedures optimally adapted to the ‘world’ of displays encountered constitute another factor – besides the allocation of attention based on acquired target-distractor LT memories – that aid the allocation of attention in a manner unguided by other feature-based (i.e., bottom-up or more short-term top-down) factors.

Arguably, this type of learning is applicable to skill acquisition in general, which constitutes a universal characteristic of human behavior (Luchins, 1942; Ryle, 1946; Fitts, 1964; Fitts & Poser, 1967; Willingham et al., 1989). Although here we explore its potential with regard to accounting for a particular phenomenon – namely, CC –, we believe that the notion of skill acquisition has implications for a model-based understanding of oculomotor behavior in visual exploration and scene viewing in general.

Summary and Outlook

The key finding from our modeling work is that a network employing solely procedural learning can give rise to what looks like memory-based 'cueing' of attention to the target location in a hard visual-search task. Importantly, the model is never informed about the identity of individual items, nor does it learn to associate a target location with the distractor layout. Hence, it does not acquire any "knowing that" (Ryle, 1945). Rather, it just learns to adapt, and hone, its previously executed fixation patterns, with optimization occurring incidentally, simply as a byproduct of repeatedly searching a certain set of displays. The network thus acquires "knowing how" to best solve the task. Although here a motor procedure is learned, we suggest that the findings generalize to other forms of procedural learning.

While the model does not acquire specific context-target associations ('templates' in long-term memory) that top-down guide attentional selection ('contextual cueing'), it nevertheless facilitates search by learning a display-generic oculomotor procedure. This raises the question whether the findings from other statistical learning paradigms (e.g., Theeuwes et al., 2022; Bogaerts et al., 2022) may also be explained by the acquisition of (oculo-)motor routines (i.e., “learning how”), rather than pre-attentive guidance processes (i.e., “learning that”). We hypothesize that if a simple *procedure* can be readily set up to solve the task at hand (e.g., visual search), attentional guidance may actually not be the dominant mechanism driving the learning-dependent performance gains. Accordingly, future research would need to factor in participants’ capacity for procedural learning across a range of cognitive domains and tasks, which would require innovative experimental designs to separate procedural and attentional-guidance effects and/or establish to what extent procedural learning interacts with statistical learning of attentional guidance. Leveraging scenarios in which acquired procedures render performance errors whereas “cognitive” strategies would yield the correct outcome might be particularly insightful in investigating the interaction between “learning how” and “learning what” in an adaptive manner.

With regard to conceptual and computational models of (foveated) vision that have traditionally emphasized guidance by bottom-up and/or top-down information: we argue that, in light of our empirical and modeling results, these models might profitably consider whether procedural strategies may not equally well explain existing findings, or in what way the models may be augmented by incorporating procedural learning. Cases in point would be considering whether the central bias is an emergent property of a statistical “learning-how” mechanism, or whether architectures operating on selecting “glimpses” (e.g., Mnih et al., 2014) would benefit from procedurally learnt stable fixation policies. Of particular theoretical as well as practical interest is that, in our model, learning a simple procedure rendered an advantage in finding the target without any explicit “teaching signal” informing the network about target identity or location. This finding may hold potential for the domain of machine learning and artificial intelligence, namely, by employing learning strategies requiring less supervision by (e.g., developing loss functions or training schemes) focusing on “learning how” instead of “learning that”, directly leveraging the task structure itself rather than the outcome of single trials. Generally, we believe it profitable tested whether some tasks, particularly involving statistical learning, may be explained and successfully modelled as instances of procedural learning.

Acknowledgments

This research was supported by a German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) grant *GE 1889/4-2* to T.G. There are no conflicts of interest. We thank Matt Peterson for sharing the data of his 2001-study with us.

References

Bar, M. (2004). Visual objects in context. Nature Reviews Neuroscience, 5(8), 617–629.

Beesley, M., Vadillo, M. A., Pearson, D. & Shanks, D. R. (2015). Pre-exposure of repeated search configurations facilitates subsequent contextual cuing of visual search. Journal of Experimental Psychology: Learning, Memory, and Cognition, 41(2), 348-362.

Behan, M., & Kime, N. M. (1996). Intrinsic circuitry in the deep layers of the cat superior colliculus. Visual Neuroscience, 13(6), 1031-1042.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. Cognitive Psychology, 14, 143-177.

Bogaerts, L., Siegelman, N., Christiansen, M. H., & Frost, R. (2022). Is there such a thing as a ‘good statistical learner’?. Trends in cognitive sciences, 26(1), 25-37.

Brockmole, J. R., & Henderson, J. M. (2006). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. The Quarterly journal of experimental psychology, 59 (7), 1177–1187.

Buswell, G. T. (1935). How people look at pictures: A study of the psychology of perception in art. Chicago: University of Chicago Press.

Canosa, R. L., Pelz, J. B., Mennie, N. R., & Peak, J. (2003). High-level aspects of oculomotor control during viewing of natural-task images. Proceedings of SPIE Human Vision and Electronic Imaging VIII, 5007, 240–251

Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive psychology, 4(1), 55–81.

Chua, K. P., & Chun, M. M. (2003). Implicit scene learning is viewpoint dependent. Perception & Psychophysics, 65(1), 72-80.

Chun, M. M. (2000). Contextual cueing of visual attention. Trends in cognitive sciences, 4(5), 170–178.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spa- tial attention. Cognitive psychology, 36(1), 28–71.

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. Psychological Science, 10, 360-365.

Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. Nature Neuroscience, 2, 844-847

Clarke, A. D., & Tatler, B. W. (2014). Deriving an appropriate baseline for describing fixation behaviour. Vision research, 102, 41-51.

Fahimi, R., & Bruce, N. D. (2021). On metrics for measuring scanpath similarity. Behavior Research Methods, 53, 609-628.

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. Trends in cognitive sciences, 10(8), 382-390.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of move- ment. Journal of experimental psychology, 47(6), 381.

Geng, J. J., & Behrmann, M. (2005). Spatial probability as an attentional cue in visual search. Perception & Psychophysics, 67(7), 1252-1268.

Gerstner, W., Kistler, W. M., Naud, R., & Paninski, L. (2014). Neuronal dynamics: From single neurons to networks and models of cognition. Cambridge University Press.

Geyer, T., Seitz, W., Zinchenko, A., Müller, H.J., & Conci, M. „Why are acquired search-guiding context memories resistant to updating?“ Front. Psychol. | doi: 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.

Geyer, T., Zehetleitner, M., & Müller, H. J. (2010). Contextual cueing of pop-out visual search: when context guides the deployment of attention. Journal of Vision, 10(5), 1-11.

Geyer, T., Zinchenko, A., Seitz, W., Balik, M., Müller, H. J., & Conci, M. (2023). Mission impossible? Spatial context relearning following a target relocation event depends on cue predictiveness. Psychonomic Bulletin & Review, 1-8.

Golomb, J. D., & Kanwisher, N. (2012). Higher level visual cortex represents retinotopic, not spatiotopic, object location. Cerebral Cortex, 22(12), 2794-2810.

Goujon, A., Didierjean, A., & Thorpe, S. (2015). Investigat- ing implicit statistical learning mechanisms through contextual cueing. Trends in cognitive sciences, 19(9), 524–533.

Greene, A. J., Gross, W. L., Elsinger, C. L., & Rao, S. M. (2007). Hippocampal differentiation without recognition: an fMRI analysis of the contextual cueing task. Learning & memory, 14(8), 548-553.

Hannula, D. E., & Greene, A. (2012). The hippocampus reevaluated in unconscious learning and memory: at a tipping point?. Frontiers in human neuroscience, 6, 80.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. Nature Reviews Neuroscience, 11(7), 523–532.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency- based visual attention for rapid scene analysis. IEEE Transactions on pattern analysis and machine intelligence, 20(11), 1254–1259.

Izhikevich, E. M. (2003). Simple model of spiking neurons. IEEE Transactions on neural networks, 14(6), 1569–1572.

Jekel, C. F., Venter, G., Venter, M. P., Stander, N., & Haftka, R. T. (2019). Similarity measures for identifying material parameters from hysteresis loops using inverse analysis. International Journal of Material Forming, 12(3), 355-378.

Jiang, Y. V., Swallow, K. M., & Rosenbaum, G. M. (2013). Guidance of spatial attention by incidental learning and endogenous cuing. Journal of Experimental Psychology: Human Perception and Performance, 39(1), 285.

Jiang, Y., Song, J.-H., & Rigas, A. (2005). High-capacity spatial contextual memory. Psychonomic Bulletin & Review, 12(3), 524-529.

Jiang, Y., Wagner, L. C. (2004). What is learned in spatial contextual cuing—configuration or individual locations?. Perception & Psychophysics, 66(3), 454-463.

Johnson, J. S., Woodman, G. F., Braun, E., Luck, S. (2007). Implicit memory influences the allocation of attention in visual cortex. Psychonomic Bulletin & Review, 14, 834-839.

Kasap, B., & van Opstal, A. J. (2017). A spiking neural network model of the midbrain superior colliculus that generates saccadic motor commands. Biological cy- bernetics, 111(3-4), 249–268.

Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. Psychological science, 10(4), 346-352.

Klopf, A. H. (1972). Brain function and adaptive systems: a heterostatic theory (No. 133). Air Force Cambridge Research Laboratories, Air Force Systems Command, United States Air Force.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. TRENDS in Neurosciences, 27(12), 712-719.

Koch, C., & Ullman, S. (1984). Selecting one among the many: A simple network implementing shifts in selective visual attention.

Kunar, M. A., & Wolfe, J. M. (2011). Target absent trials in configural contextual cuing. Attention, Perception, & Psychophysics, 73, 2077-2091.

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, 816-828.

Makovski, T., & Jiang, Y. V. (2010). Contextual cost: When a visual-search target is not where it should be. Quarterly Journal of Experimental Psychology, 63(2), 216-225.

Manginelli, A. A., & Pollmann, S. (2009). Misleading contextual cues: How do they affect visual search?. Psychological Research, 73(2), 212-221.

Manns, J. R., & Squire, L. R. (2001). Perceptual learn- ing, awareness, and the hippocampus. Hippocampus, 11(6), 776–782.

Markram, H., Gerstner, W., & Sjöström, P. J. (2012). Spike-timing-dependent plasticity: A comprehen- sive overview. Frontiers in synaptic neuroscience, 4, 2.

McNaughton, B. L., Barnes, C. A., Gerrard, J. L., Gothard, K., Jung, M. W., Knierim, J. J., ... & Weaver, K. L. (1996). Deciphering the hippocampal polyglot: the hippocampus as a path integration system. The Journal of experimental biology, 199(1), 173-185.

Mnih, V., Heess, N., & Graves, A. (2014). Recurrent models of visual attention. Advances in neural information processing systems, 27.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. Nature, 434(7031), 387-391.

Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. Vision Research, 11(9), 929-IN928.

NVIDIA, Vingelmann, P., & Fitzek, F. H. P. (2020). CUDA, release: 10.2.89. Retrieved from https://developer.nvidia.com/cuda-toolkit

Olson, I. R., & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. Visual cognition, 9(3), 273-302.

Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. Vision Research, 42 (1), 107–123.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems, 32.

Peterson, M. S., & Kramer, A. F. (2001). Attentional guidance of the eyes by contextual information and abrupt onsets. Perception & Psychophysics, 63(7), 1239-1249.

Preston, A. R., & Gabrieli, J. D. (2008). Dissociation be- tween explicit memory and configural memory in the human medial temporal lobe. Cerebral Cortex, 18(9), 2192–2207.

Rabinovich, M. I., Zaks, M. A., & Varona, P. (2020). Sequential dynamics of complex networks in mind: Consciousness and creativity. Physics Reports. doi:10.1016/j.physrep.2020.08.003

Rolls, E. T. (1999). Spatial view cells and the representation of place in the primate hippocampus. Hippocampus, 9(4), 467-480.

Rolls, E. T., Robertson, R. G., & Georges-François, P. (1997). Spatial view cells in the primate hippocampus. European Journal of Neuroscience, 9(8), 1789-1794.

Rosenbaum, D. A., Cohen, R. G., Jax, S. A., Weiss, D. J., & Van Der Wel, R. (2007). The problem of serial order in behavior: Lashley’s legacy. Human movement science, 26(4), 525-554.

Roth, A., & van Rossum, M. C. (2009). Modeling synapses. Computational modeling methods for neuroscientists, 6, 139-160.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. Journal of neurology, neurosurgery, and psychiatry, 20(1), 11. Computational modeling methods for neuroscientists, 6, 139–160.

Seitz, W., Zinchenko, A., Müller, H. J., & Geyer, T. (2023). Contextual cueing of visual search reflects the acquisition of an optimal, one-for-all oculomotor scanning strategy. Communications Psychology, 1(1), 20.

Seitz, W., Zinchenko, A., Müller, H. J., & Geyer, T. (2024). Distraction from long-term memory in visual search: acquired procedural and template-based memory interfere with contextual cueing after target re-location. Visual Cognition, accepted. https://doi.org/10.1080/13506285.2024.2315786

Sewell, D. K., Colagiuri, B., & Livesey, E. J. (2018). Response time modeling reveals multiple contextual cuing mechanisms. Psychonomic bulletin & review, 25(5), 1644–1665.

Shi, Z., Zang, X., Jia, L., Geyer, T., & Mueller, H.J. (2013). Transfer of contextual cueing in full-icon display remapping. Journal of Vision, 13(3), 2.

Šimić, G., Kostović, I., Winblad, B., & Bogdanović, N. (1997). Volume and number of neurons of the human hippocampal formation in normal aging and Alzheimer's disease. Journal of Comparative Neurology, 379(4), 482-494.

Sisk, C. A., Remington, R. W., & Jiang, Y. V. (2019). Mechanisms of contextual cueing: A tutorial review. Attention, Perception, & Psychophysics, 81(8), 2571–2589.

Theeuwes, J., Bogaerts, L., & van Moorselaar, D. (2022). What to expect where and when: How statistical learning drives visual selection. Trends in cognitive sciences. 26(19):860-872

Tseng, Y.-C., & Li, C.-S. R. (2004). Oculomotor correlates of context-guided learning in visual search. Perception & Psychophysics, 66(8), 1363–1378.

Vadillo, M. A., Giménez-Fernández, T., Beesley, T., Shanks, D. R., & Luque, D. (2021). There is more to contextual cuing than meets the eye: Improving visual search without attentional guidance toward predictable target locations. Journal of Experimental Psychology: Human Perception and Performance, 47(1), 116.

Võ, M. L.-H., & Wolfe, J. M. (2013). Differential electrophysiological signatures of semantic and syntactic scene processing. Psychological Science, 24(9), 1816-1823.

Whishaw, I. Q., McKenna, J. E., & Maaswinkel, H. (1997). Hippocampal lesions and path integration. Current opinion in neurobiology, 7(2), 228-234.

Wills, T. J., Lever, C., Cacucci, F., Burgess, N., & O'Keefe, J. (2005). Attractor dynamics in the hippocampal representation of the local environment. Science, 308(5723), 873–876.

Wolfe, J. M. (1994). Guided search 2.0. A revised model of visual search. Psychonomic bulletin & Review, 1(2), 202-238.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), Attention (pp. 13-73). Hove: Psychology Press.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. Nature Human Behaviour, 1(3), 1-8.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. Psychological review, 115(4), 787.

Zelinsky, G.J. (1996). Using Eye Saccades to Assess tge Selectivity of Search Movements. Vision Research, 36 (14), 2177-2187.

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.

Zinchenko, A., Conci, M., Töllner, T., Müller, H. J., & Geyer, T. (2020). Automatic guidance (and misguidance) of visuospatial attention by acquired scene memory: Evidence from an N1pc polarity reversal. Psychological Science, 31(12), 1531-1543.

General Discussion

In summary, three studies have investigated the role of procedural learning - that is, “learning how” in visual search under invariant contexts. Learning a procedural scanning routine may be helpful in bringing the likely relevant aspects of a visual environment into view and enable subsequent (pre-)attentional processing. Hence, procedural learning might be a so far neglected “filter”, even before attentional selection takes place: Wolfe & Horowitz (2017) identify five factors that guide attention - bottom-up salience, top-down feature guidance, scene structure and meaning, search history and reward. These findings are in line with evolutionary mechanisms, as they are cost-effective and avoid unnecessary interference between encoded memories.

While, semantically, our findings coincide with the terms of ‘scene structure’ and ‘search history’, these factors usually refer to the learned syntactic/semantic properties of scenes, such as things adhering to gravity or birds sitting in trees (Biederman, 1976; Biederman et al., 1982), whereas search ‘history’ usually implies a statistical learning of e.g. a region with a high probability of containing the target position (Geng & Behrmann, 2005) or display-specific learning of targets (Chun & Jiang, 1998), the presented studies offer a novel view on how the statistical structure of scenes can be utilized in an agnostic way by learning useful procedures that apply to task-generic demands and make use of the task's specific structure.

Hence, the present work extends the current notions of attentional guidance, which is a procedure that acts ‘blindly’, before attentional selection has taken place bringing potentially relevant portions of a display into view and making them available for (pre-)attentional processing. However, this process itself is not guided by mechanisms of attention, but by procedural routines.

Studies

Study 1 identifies procedural learning and illustrates how procedural learning acts as a motor behind the contextual cueing effect. While contextual ueing has previously been believed to be a display-specific top-down memory effect on attentional guidance (i.e. an instance of “knowing that”), learning of optimized oculomotor scanpaths along the set of the entire display (i.e. an instance of "knowing how") presents an alternative mechanism to bring about the effect.

The central finding of the study is that scanpath similarity is higher when pairs of different participants process identically composed old-context displays, as opposed to scanning itentical new-context displays. Even more importantly, scanpath similarity is increased in individual participants when encountering different repeated displays with different spatial composition compared to non-repeated displays. In other words, the display-generic search procedure is more invariant for the spatially invariant repeated displays, compared to the spatially variable non-repeated displays. This can be thought of a result of tuning the scanning strategy to the bulk of all displays, with repeated displays weighing in more strongly as a direct consequence of repetition, that is, multiple exposures, whereas novel displays are by definition only encountered once.

Study 2 supports the novel view according to which there is guidance and also misguidance of search from experience with repeated display layouts, which arises from both long-term memory for specific distractor-target arrangements and, alongside, long-term memory for generic oculomotor procedures. With this, study 2 thus complements study 1 by demonstrating that learnable display-generic scanning strategies actively aid search even before attentional selection takes place. This happens by learning a procedural strategy that brings locations with a higher task relevance into view earlier, as well as by minimizing costs associated with the encoding and recall of (a large number of) display-specific contextual memories: Faithfully encoding entire display configurations would not only go along with higher costs, but also increase interference between display arrangements. Thus, a procedural strategy is not only cost efficient but is also minimizing potential interference at the same time.

Our findings suggest that procedural learning might play a significant role in various forms of visual statistical learning, such as probability cueing. Here, learning an optimal scanning strategy might lead to findings similar to those of top-down attentional guidance and hence, further research is needed.

Study 3 aimed at implementing a connectionist model of contextual cueing. Using a computational model allowed full control in terms of what kind of learning is enabled or disabled. The current model merely learns to associate fixatated regions of previosly executed saccades as it searches through the displays of a contextual cueing experiment.

Hence, the model can only learn a simplistic procedure and never learns to associate a specific context to a specific target location. In fact, the model is never informed about having found the target at a specific location, and moreover, configurational learning of display-target associations is disabled. Instead, search is terminated upon finding the target and search begins anew with a new display. Nevertheless, the model not only shows a contextual cueing effect in terms of the number of simulated fixations, it also replicates the findings of a higher scanpath similarity for repeated displays versus non-repeated displays within 'participants' of Study 1.

Furthermore, the model can quantitatively predict findings from other studies, such as a biologically plausible number of fixations (e.g., Peterson & Kramer, 2001) and its performance beats the current benchmark models by a large margin. Also, as a consequence of learning, the model develops a central bias as an emergent phenomenon.

These findings illustrate that learning a simplictic oculomotor procedure on self-reinforcing previous oculomotor behavior is sufficient to elicit a contextual cueing effect. Moreover, the learning of such a procedure might give rise to generally observed effects, such as the central bias. Additionally, the findings suggest that enabling a model to leverage task structure might be beneficial not only for visual-search models, but also for the field of machine learning and artificial intelligence by exploring new perspectives on learning and lower demands on labeling.

Hippocampal involvement

Studies 1-3 show that contextual cueing is caused less by display-specific memories but rather by procedural learning. Instead, display-generic oculomotor scanning procedures are optimized with respect to the entire set of displays in which the influence of repeated displays weighs in more strongly.

Searching T/L letter arrays repeatedly aid the visual system in developing new capabilities and learning more effective scanning strategies. This also means that eye movements are not just the consequence of a focused shifting of attention, but that they proactively support the orienting of attention toward objects that will most likely contain relevant information for further processing – before the actual attentional selection takes place. Somewhat at odds with these findings are studies that link the hippocampus - a locus associated with explicit memories to contextual cueing, and hippocampal lesions to an impairment in eliciting the contextual cueing effect (Chun & Phelps, 1999). However, hippocampal activity is not only indicative of explicit, but also implicit motor-sequence learning (e.g., in finger-tapping tasks, Albouy et a., 2013, as well as other, statistical-learning paradigms, such as serial reaction-time task; e.g., Hazeltine, Grafton & Ivry, 1997; Müller et al., 2002), particularly in the initial stages, with a subsequent decrease in activity later on (Albouy et al., 2008). This is consistent with a critical (but over the course of practice diminishing) role of the hippocampus in procedural motor learning (Albouy et al., 2013). Moreover, hippocampal invovement does not necessarily imply the presence of explicit learning and it is possible that the hippocampus supports implicit forms of learning in CC (Chun, 2000).

In fact, there is direct evidence from a contextual cueing study using fMRI by Manelis and Reder (2012), consistent with a procedural-learning account of contextual cueing: With progression of the task, Manelis and Reder (2012) found a significant decrease in functional connectivity between hippocampus and the particular the left superior parietal lobule. One possible interpretation is that encoding the spatial context of a target is a necessary step in forming context-target associations (Manelis & Reder, 2012); however, another possibility is that in the initial blocks, display-generic (i.e. “domain general relational”; see Davici, 2006) information is learned.

Explicit vs. implicit learning

Following a similar argumentation, several studies argue for a unitary memory system in contextual cueing, suggesting that contextual cueing is possibly driven by explicit memories that are below the threshold of awareness (e.g. see Vadillo et al., 2016; Kroell et al., 2019; Meyen et al., 2023). Although these studies do not provide evidence for the absence of other memory systems involved (Meyen et al., 2023), they do not strictly rule out the involvement of other memory systems either. In fact, an analysis of a very large sample size implies that it it unlikely that explicit recognition of distractor-target associations is the motor behind contextual facilitation of search, as even sensitive measures of awareness for display repetition are uncorrelated with a benefit in reaction times for those displays, while the large sample size of about 700 participants makes power issues unlikely (Colagiuri & Livesey, 2016).

One possible reason why procedural learning may not have been considered as a driving force behind the contextual cueing effect in these studies is the fact that it is present in virtually all studies (Jiang & Chun, 2003), and has not been overlooked, but not assumed to contribute to the effect, i.e. the context-dependent difference itself, instead of merely the general decrease in reaction time. Therefore, strong claims about the memory system underlying the CC effect require more evidence than what is currently presented using, e.g., Yes/No recognition tasks that query display-specific knowledge. Based on the results of studies 1-3 presented, one can make the argument that more fine-grained memory tests for contextual cueing should also consider participants’ procedures for solving the search task in particular.

In favor of an implicit learning account of CC, initial studies of attributed the effect to implicit learning (Chun & Jiang, 1998; Chun, 2000; Jiang & Chun, 2003), driven by the observation that learning seemed incidental, with anecdotal reports of difficulties in memorizing displays while performing visual search even when attempting so and no above-chance recognition in a forced-choice test (Chun & Jiang, 1998). Further evidence comes from studies in which, after a learning period of multiple blocks, the target of repeated displays is relocated to a different position, showing poor adaptation to the new target location (Manginelli & Pollmann, 2009), which is consistent with strong retroactive interference typically observed in implicit learning tasks (see Goujon et al., 2015). Beyond these observations, CC conforms to a multitude of implicit learning principles (for a review, see Goujon et al., 2015), such as strong resistance to extinction (Jiang et al., 2005, Zellin et al., 2014) and e.g. an independence of IQ (Merill et al., 2014). Also, CC does not seem to rely on a fully matured hippocampus, suggesting that explicit learning may not be necessary to elicit CC (Jiang et al., 2019).

Here, a procedural learning account may reconcile the debate about the memory systems involved in the effect. While explicit, i.e. display-specific learning of context-target associations may occur in CC, it is conceivable that there is a scanning strategy that is procedural and display-independent. Under this assumption, procedural scanning strategies and display-specific learning would be attributable to two different systems complementing each other.

Relation to scene-based contextual cueing and explicit learning

Interrelated to the question of the memory system/s involved in CC is the choice of stimuli employed. In contrast to the controversial L-vs.-T stimuli, there is wide consensus that CC paradigms employing naturalistic stimuli lead to explicit recognition (Brockmole & Henderson, 2006a, Brockmole & Henderson, 2006b). Evidence using eye-tracking in a visual search task using repeating naturalistic contexts further supports this notion by showing that oculomotor scanpaths almost directly approach the target after some repetitions (Summerfield et al., 2011). In contrast, in conventional (L vs. T) array-based studies, the number of fixations necessary to reach the target remains above 4, even after being exposed to a repeated context 12-16 times (see, e.g., Peterson & Kramer, 2001; Tseng & Li, 2004; study 3 of this thesis).

One possible mediating factor could be something described as "meaningfulness": Scenes can be argued to convey more meaningful to observers than letters. In fact, chess experts have a profoundly higher CC effect scores compared to novices, when chess configurations - which are more meaningful for chess experts than novices - are used as stimuli. Importantly, this effect is stronger for plausible over implausible configurations of chess pieces in experts (Brockmole et al., 2008). Further, the “overshadowing account” provided by Rosenbaum and Jiang (2013) suggests that scene-based information with rather explicit cueing characteristics dominates contextual cueing even when combined with array-based stimuli as well. This finding could be interpreted as evidence for the existence of (at least) two dissociable cognitive systems that contribute to CC. The notion that in less meaningful, array-based scenes, participants predominantly use a procedural “one for all” search strategy, whereas in meaningful scenes explicit recognition is possible, provides a novel and testable account that could resolve the differential findings for varying stimuli.

Summary and outlook

The present studies have shown that procedural learning plays an important role in visual search with L- and T-letter stimuli under invariant spatial contexts. While there is some evidence from empirical and theoretical studies that procedural learning may play a rather central role, particularly when rather "meaningless", L-vs.-T letter arrays are used as stimuli, beyond that, there are evolutionary arguments in favor of optimizing a search procedure rather than forming explicit (albeit maybe weak and "subthreshold") memories about previously encountered displays: A procedural strategy may incur smaller costs in encoding memories, while at the same time avoid interference from multiple, similar memories.

Investigating procedural scanning strategies by means of fixational eye movements may therefore yield new insights into visual search, as they are currently not part of the major theoretical frameworks (such as Guided Search; Wolfe, 2021). This is of particular interest, as procedures may extend beyond eye-movements, i.e. (oculo-)motor behavior in general: In visual search with and without eye-movements, almost identical activations are found in fMRI (De Haan, 2008).

Further, eye tracking may offer specific insights into the process of procedural learning. For instance, eye-movements to different locations may be associated with incurrding different costs (Araujo, 2001) which may play a crucial role in shaping procedural scanning strategies that are display-generic and therefore "agnostic" to the specific yet to be explored content of a display.

Finally, exploring procedural scanning strategies may offer a new way to understand attention: Because simplistic and in mathematical terms precisely definable procedures can mimic the effects of a rather elusive term of attention, they might be one of “the many systems that implement” a “behaviorally relevant selection process” that could replace or at least clarify the currently inconsistent and not too well-defined concepts of (visual) attention (see Hommel et al., 2019).

As a consequence, these findings might pave the way for more comprehensive neuroimaging studies, shedding light on procedural learning and how it integrates with other types of learning, such as explicit learning. Studies such as these will be of particular importance of lifting psychological research away from an artificial notion of a “cognitive” agent evaluating and appraising sensory inputs and then subsequently acting on those, towards understanding cognition and behavior as tightly coupled, interdependent processes, such that behavior itself can be considered e.g. ‘smart’ and not the appraisal of a behavior (see Ryle, 1945).

References

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Re- view, 89, 369–406.

Araujo, C., Kowler, E., & Pavel, M. (2001). Eye movements during visual search: The costs of choosing the optimal path. Vision research, 41(25-26), 3613-3625.

Albouy, G., Sterpenich, V., Balteau, E., Vandewalle, G., Desseilles, M., Dang-Vu, T., ... & Maquet, P. (2008). Both the hippocampus and striatum are involved in consolidation of motor sequence memory. Neuron, 58(2), 261-272.

Albouy, G., King, B. R., Maquet, P., & Doyon, J. (2013). Hippocampus and striatum: Dynamics and interaction during acquisition and sleep-related motor sequence memory consolidation. Hippocampus, 23(11), 985-1004.

Beesley, T., Vadillo, M. A., Pearson, D., & Shanks, D. R. (2015). Pre-exposure of repeated search configurations facilitates subsequent contextual cuing of visual search. Journal of Experimental Psychology: Learning, Memory, and Cognition, 41(2), 348.

Biederman. I. (1976). On processing information from a glance at a scene: some implications for a syntax and semantics of visual processing. UODICS’76. 75–88.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive psychology*, *14*(2), 143-177.

Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: modeling contextual cuing. Journal of Experimental Psychology: Human Perception and Performance, 33(4), 798.

Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006a). Contextual cueing in naturalistic scenes: Global and local contexts. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32(4), 699.

Brockmole, J. R., & Henderson, J. M. (2006b). Short article: Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. Quarterly journal of experimental psychology, 59(7), 1177-1187.

Brockmole, J. R., Hambrick, D. Z., Windisch, D. J., & Henderson, J. M. (2008). The role of meaning in contextual cueing: Evidence from chess expertise. Quarterly Journal of Experimental Psychology, 61(12), 1886-1896.

Brooks, D. I., Rasmussen, I. P., & Hollingworth, A. (2010). The nesting of search contexts within natural scenes: evidence from contextual cuing. Journal of Experimental Psychology: Human Perception and Performance, 36(6), 1406.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. Cognitive psychology, 36(1), 28-71.

Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. Nature neuroscience, 2(9), 844-847.

Chun, M. M. (2000). Contextual cueing of visual attention. Trends in cognitive sciences, 4(5), 170-178.

Clarke, A. D., & Tatler, B. W. (2014). Deriving an appropriate baseline for describing fixation behaviour. Vision research, 102, 41-51.

Colagiuri, B., & Livesey, E. J. (2016). Contextual cuing as a form of nonconscious learning: Theoretical and empirical analysis in large and very large samples. Psychonomic Bulletin & Review, 23, 1996-2009.

Davachi, L. (2006). Item, context and relational episodic encoding in humans. Current opinion in neurobiology, 16(6), 693-700.

De Haan, B., Morgan, P. S., & Rorden, C. (2008). Covert orienting of attention and overt eye movements activate identical brain regions. Brain research, 1204, 102-111.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. Vision Research, 36(12), 1827-1837.

Geng, J. J., & Behrmann, M. (2005). Spatial probability as an attentional cue in visual search. Perception & Psychophysics, 67(7), 1252-1268.

Geng, J. J. (2014). Attentional mechanisms of distractor suppression. Current Directions in Psychological Science, 23(2), 147-153.

Goujon, A., Didierjean, A., & Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. Trends in cognitive sciences, 19(9), 524-533.

Hazeltine, E., Grafton, S. T., & Ivry, R. (1997). Attention and stimulus characteristics determine the locus of motor-sequence encoding. A PET study. Brain: a journal of neurology, 120(1), 123-140.

Henderson, J. M., Weeks Jr, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. Journal of experimental psychology: Human perception and performance, 25(1), 210.

Hoffman, J. E. (2016). Visual attention and eye movements. Attention, 119-153.

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J. H., & Welsh, T. N. (2019). No one knows what attention is. Attention, Perception, & Psychophysics, 81, 2288-2303.

James, W. (1890). The Principles Of Psychology Volume II By William James (1890).

Jiang, Y., & Chun, M. M. (2003). Contextual cueing. Attention and implicit learning, 48, 277.

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.

Jiang, Y., Song, J. H., & Rigas, A. (2005). High-capacity spatial contextual memory. Psychonomic bulletin & review, 12(3), 524-529.

Jiang, Y. V., Swallow, K. M., & Capistrano, C. G. (2013). Visual search and location probability learning from variable perspectives. Journal of Vision, 13(6), 13-13.

Jiang, Y., & Wagner, L. C. (2004). What is learned in spatial contextual cuing—configuration or individual locations?. Perception & Psychophysics, 66, 454-463.

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.

Manelis, A., & Reder, L. M. (2012). Procedural learning and associative memory mechanisms contribute to contextual cueing: Evidence from fMRI and eye-tracking. Learning & Memory, 19(11), 527-534.

Manginelli, A. A., & Pollmann, S. (2009). Misleading contextual cues: How do they affect visual search?. Psychological Research, 73, 212-221.

Merrill, E. C., Conners, F. A., Yang, Y., & Weathington, D. (2014). The acquisition of contextual cueing effects by persons with and without intellectual disability. Research in developmental disabilities, 35(10), 2341-2351.

Meyen, S., Vadillo, M. A., von Luxburg, U., & Franz, V. H. (2023). No evidence for contextual cueing beyond explicit recognition. Psychonomic Bulletin & Review, 1-24.

Miller, J. (1988). Components of the location probability effect in visual search tasks. Journal of Experimental Psychology: Human Perception and Performance, 14(3), 453.

Müller, R. A., Kleinhans, N., Pierce, K., Kemmotsu, N., & Courchesne, E. (2002). Functional MRI of motor sequence acquisition: effects of learning stage and performance. Cognitive Brain Research, 14(2), 277-293.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. Nature, 434(7031), 387-391.

Newell, A. & Rosenbloom, P. Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), Cognitive Skills and their Acquisition, Hillsdale, NJ: Lawrence Erlbaum Associates, 1981.

Peterson, M. S., & Kramer, A. F. (2001). Attentional guidance of the eyes by contextual information and abrupt onsets. Perception & Psychophysics, 63(7), 1239-1249.

Preuschhof, C., Sharifian, F., Rosenblum, L., Pohl, T. M., & Pollmann, S. (2019). Contextual cueing in older adults: Slow initial learning but flexible use of distractor configurations. Visual Cognition, 27(5-8), 563-575.

Rosenbaum, G. M., & Jiang, Y. V. (2013). Interaction between scene-based and array-based contextual cueing. Attention, Perception, & Psychophysics, 75, 888-899.

Ryle, G. (1945, January). Knowing how and knowing that: The presidential address. In Proceedings of the Aristotelian society(Vol. 46, pp. 1-16). Aristotelian Society, Wiley.

Schankin, A., & Schubö, A. (2010). Contextual cueing effects despite spatially cued target locations. Psychophysiology, 47(4), 717-727.

Shaw, M. L., & Shaw, P. (1977). Optimal allocation of cognitive resources to spatial locations. Journal of Experimental Psychology: Human Perception and Performance, 3(2), 201.

Shaw, M. L. (1978). A capacity allocation model for reaction time. Journal of Experimental Psychology: Human Perception and Performance, 4(4), 586.

Sewell, D. K., Colagiuri, B., & Livesey, E. J. (2018). Response time modeling reveals multiple contextual cuing mechanisms. Psychonomic Bulletin & Review, 25, 1644-1665.

Sheliga, B. M., Riggio, L., & Rizzolatti, G. (1995). Spatial attention and eye movements. Experimental brain research, 105, 261-275.

Snoddy, G. S. (1926). Learning and stability: a psychophysiological analysis of a case of motor learning with clinical applications. Journal of Applied Psychology, 10(1), 1.

Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., & Nobre, A. C. (2006). Orienting attention based on long-term memory experience. Neuron, 49(6), 905-916.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. Cognitive psychology, 12(1), 97-136.

Tseng, Y. C., & Li, C. S. R. (2004). Oculomotor correlates of context-guided learning in visual search. Perception & Psychophysics, 66(8), 1363-1378.

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2016). Underpowered samples, false negatives, and unconscious learning. Psychonomic Bulletin & Review, 23, 87-102.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human perception and performance, 15(3), 419.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 0058.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. Psychonomic bulletin & review, 1, 202-238.

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. Psychonomic Bulletin & Review, 28(4), 1060-1092.

Yarbus, A. L., & Yarbus, A. L. (1967). Eye movements during perception of complex objects. Eye movements and vision, 171-211.

Zelinsky, G. J. (1996). Using eye saccades to assess the selectivity of search movements. Vision research, 36(14), 2177-2187.

Zellin, M., von Mühlenen, A., Müller, H. J., & Conci, M. (2014). Long-term adaptation to change in implicit contextual learning. Psychonomic Bulletin & Review, 21, 1073-1079.

Zheng, L., & Pollmann, S. (2019). The contribution of spatial position and rotated global configuration to contextual cueing. Attention, Perception, & Psychophysics, 81, 2590-2596.

Curriculum Vitae

Education

University of Oxford: Research Visit 10/2022 – 03/2023

(Prof. Kia Nobre)

Ludwig-Maximilians-Universität München: PhD studies in Systemic Neurosciences 10/2019 – 04/2024

(Prof. Thomas Geyer & Prof. Hermann Müller)

Technische Universität München: MSc. Neuroengineering, Elite-Master 10/2017 – 04/2020

Technische Universität Chemnitz: BSc. Sensors and Cognitive Psychology 11/2012 – 11/2017

Professional Experience

Prosieben Sat.1 Media SE: Data Science 01/2019 – 07/2019

Publications

Seitz, W., Zinchenko, A., Müller, H. J. & Geyer, T. (in press). Distraction from long-term memory in visual search: acquired procedural and template-based memory interfere with contextual cueing after target re-location. Handling Visual Distraction [Special Issue], *Visual Cognition*.

Seitz, W., Zinchenko, A., Müller, H. & Geyer, T. (2023). Statistical learning in visual search: ‘contextual cueing’ reflects the acquisition of an optimal, ‘one-for-all’ oculomotor scanning strategy. *Communications Psychology, 2023*.

Geyer, T., Seitz, W., Balik, M., Müller, H. J. & Conci, M. Mission Impossible? Spatial context relearning following a target relocation event depends on cue predictiveness. *Psychonomic Bulletin & Review, 2023*.

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, 650245*.

Affidavit

Eidesstattliche Versicherung/Affidavit

Werner Seitz
(Studierende / Student)

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation
Learning how: Procedural learning as a central element of visual statistical learning
selbstständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation Learning how: Procedural learning as a central element of visual statistical learning is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

München / Munich

21. Apr. 2024
(Datum / Date)

Werner Seitz
(Unterschrift / Signature)

Declaration of Contributions

Study 1

WS and AZ conceived and designed the experiments; AZ performed the experiments; WS analyzed the data; HJM contributed to conceptual analysis and discussion; WS, HJM, TG wrote the paper.

Study 2

WS and AZ conceived and designed the experiments; AZ performed the experiments; WS analyzed the data; HJM contributed to conceptual analysis and discussion; WS, HJM, TG wrote the paper.

Study 3

WS conceived and designed the experiments; WS analyzed the data; HJM contributed to conceptual analysis and discussion; WS, AZ, HJM, TG wrote the paper.