Statistical learning of distractor co-occurrences facilitates visual search

Sushrut Thorat[1], Genevieve Quek[1,2], Marius V. Peelen[1]

[1]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands.

[2]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia

Correspondence to:

Marius V. Peelen

Donders Institute for Brain, Cognition and Behaviour

Radboud University

Thomas van Aquinostraat 4, 6525 GD Nijmegen, The Netherlands

Email: marius.peelen@donders.ru.nl

Conflict of interest statement: The authors declare no competing financial interests.

Acknowledgements: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 725970).

Abstract

Visual search depends on the relationship between the target and the distractors, including how the target differs from the distractors and where the target is likely to be amongst the distractors. Whether the statistical structure amongst distractors themselves facilitates search is less well understood. Here, we assessed the benefit of distractor structure using novel shapes whose relationship to each other was learned implicitly during visual search. Participants searched for target items in arrays of shapes that comprised either four pairs of co-occurring distractor shapes (structured scenes) or eight distractor shapes randomly partitioned into four pairs on each trial (unstructured scenes). Across five online experiments (N=1140), we found that after a period of search training, participants were more efficient when searching for targets in structured than unstructured scenes. This structure-benefit emerged independently of whether the position of the shapes within each pair was fixed or varied, and despite participants having no explicit knowledge of the structured pairs they had seen. These results show that learned co-occurrence statistics between distractor shapes increases search efficiency. An increased efficiency of the rejection of regular distractors may contribute to the efficiency of visual search in natural scenes, where such regularities are abundant.

Introduction

Visual search is the task of finding a target object (e.g., a computer mouse on a desk) amongst distractor objects (e.g., other objects on the desk). It is well-established that search-difficulty (as measured by reaction time and/or accuracy) increases linearly with the number of distractors (Wolfe, 1998). Interestingly, this relationship is much weaker for search in natural scenes than for search in artificial arrays comprising randomly arranged objects (Wolfe, Alvarez, et al., 2011). What makes naturalistic visual search so efficient?

An important contribution comes from the information scene context provides about spatial ("where") and featural ("what") target properties (Castelhano & Krzyś, 2020; Oliva & Torralba, 2007; Peelen & Kastner, 2014; Võ et al., 2019; Wolfe, Võ, et al., 2011). For example, the likely location of the target in a scene can be learned and used to facilitate search, both based on recent experience in controlled laboratory experiments ("contextual cueing"; Chun, 2000) and based on long-term daily-life experience (Castelhano & Krzyś, 2020; Võ et al., 2019): when searching for a computer mouse, we start searching to the right of the keyboard and below the monitor. Scene context also provides information about the features that characterize the target (Peelen & Kastner, 2014), or distinguish the target from the distractors (Geng & Witkowski, 2019): we look for a small target far away and a large target nearby (Gayet & Peelen, 2022). Finally, targets are recognized more quickly when embedded in context, reflecting the facilitatory influence of contextual expectations on object recognition (Bar, 2004; de Lange et al., 2018). Thus, our long- and short-term experience with regularities in where and how targets appear in scenes contributes to the efficiency of visual search.

Importantly, however, real-world scenes are additionally characterized by regularities amongst distractors themselves. For example, when searching for the television in a living room, the co-occurrence statistics and spatial arrangements of many distractor objects (e.g., chairs, tables, lamps) is relatively stable. Some of these regularities are consistent across environments and learned across a lifetime (e.g., a typical living room layout), while others are

specific to a particular context and learned more rapidly (e.g., the arrangement of objects in my friend's living room). Previous research found that visual search is easier when distractor objects are arranged in configurations that follow real-world regularities (e.g., lamp above table) than when they are arranged in unfamiliar configurations (e.g., lamp below table; Kaiser et al., 2014). These results may reflect more efficient encoding of familiar object pairs based on long-term experience (Bar & Ullman, 1996; Quek & Peelen, 2020; Stein et al., 2015), facilitating visual search when these objects appear as distractors. Alternatively, visual search may be disrupted when distractor configurations violate higher-level functional and semantic associations (Spaak et al., 2022; Võ & Wolfe, 2013).

Together, the findings reviewed above raise the question of whether statistical regularities amongst distractors contribute to the efficiency of search, independently of target-distractor regularities and independently of long-term semantic knowledge. Interestingly, previous research has shown that statistical regularities between shapes can be learned rapidly (Fiser & Aslin, 2001; Fiser & Lengyel, 2019; Schapiro & Turk-Browne, 2015). For example, when participants passively view displays in which one shape frequently appears together with another shape (always in the same configuration), participants later report higher familiarity for these pairs relative to control pairs (Fiser & Aslin, 2001). Furthermore, such newly-learned shape pairs show object-like behavioral signatures, with attention spreading from one shape to the other (Lengyel et al., 2021), akin to effects of perceptual grouping (Egly et al., 1994; Scholl, 2001). If such regularity-based object grouping occurs amongst distractor objects during visual search, this compression could effectively reduce the distractor numerosity (Zhao & Yu, 2016), thereby enhancing search performance, similar to how perceptual grouping facilitates visual search (Donnelly et al., 1991; Humphreys et al., 1989; Rauschenberger & Yantis, 2006).

To test whether newly-learned statistical regularities amongst distractors contribute to the efficiency of search, here we combined a statistical learning paradigm with a visual search task using novel shapes. The use of novel shapes allowed for full control over co-occurrence

probabilities and low-level stimulus properties. Participants searched for pre-cued target shapes in arrays that consisted of four pairs of co-occurring distractor shapes (structured scenes) or or eight distractor shapes randomly partitioned into four pairs (unstructured scenes). Participants were not informed about the co-occurrences, such that all co-occurrence statistics were learned during the search task itself. To assess if the specific spatial arrangement of co-occurring shapes within the pairs was essential for distractor complexity reduction, the co-occurring shapes either had fixed arrangements (e.g., shape A always appeared above shape B) or their locations within the pairs could be swapped (e.g., shape A could appear above or below shape B).

Across multiple experiments, we found that participants were more efficient in searching for targets in the structured scenes than the unstructured scenes. Interestingly, this pattern was independent of whether the arrangement of co-occurring shapes within the pairs was fixed or not. Finally, unlike previous statistical learning studies where the co-occurring objects were attended (Fiser & Aslin, 2001), here participants were not able to indicate which shapes co-occurred during the visual search experiment, indicating that statistical regularities in the environment facilitate search even when these regularities are not explicitly noticed.

Materials and Methods

Participants

Participants were recruited online using Prolific, received monetary compensation for their participation, and provided informed consent before starting the experiment. The study was approved by the Radboud University Faculty of Social Sciences Ethics Committee (ECSW2017–2306-517). Participants from whom we obtained partial data were excluded from the analysis (~10% dropout rate). For any given experiment requiring a particular number of participants (see below), we first tested around that number of participants balancing the blocking order of scene structure. Then, participants whose overall accuracy and reaction times were above or below 3 standard deviations (SDs) from the means were removed. This was done iteratively until no exclusions happened. Then, more participants were added to get to the desired number and this exclusion process was repeated. In the end, we obtained the desired number of participants for each experiment whose accuracies and reaction times (for correct responses) were within 3 SDs from the means and the blocking order was balanced.

The desired number of participants for the two pilot experiments was 40: Experiment 1A (mean age: 25.3 years, SD = 4.4), Experiment 1B (mean age: 27.1 years, SD = 4.7). The number of participants for the two large-scale replication experiments was 400: Experiment 2A (mean age: 24.1 years, SD = 4.3), Experiment 2B (mean age: 25.6 years, SD = 6.4). Finally, the number of participants for Experiment 3 was 260 (N=260; mean age: 23.9 years, SD = 4.5). This experiment was pre-registered (https://aspredicted.org/blind.php?x=5ne7qa).

Stimuli

The stimulus set consisted of 20 novel shapes (See Figure 1 for examples), a subset of which overlapped with those from seminal statistical learning studies (Fiser & Aslin, 2001, 2005). For each participant, we randomly assigned the 20 shapes to three different sets that were maintained throughout the experiment: 4 shapes were used as search targets (target set), 8

shapes were allocated into 4 co-occurring distractor pairs (structured distractor set), and 8 shapes were used to create 4 random distractor pairs on each new trial (unstructured distractor set). Critically, a shape assigned to the structured set only ever appeared in a vertical pairing with its nominated partner shape. In *fixed* arrangements (Experiments 1A, 2A, 3), the shapes in the structured set appeared in specific vertical arrangements throughout the experiment (e.g., shape A always appeared above shape B). In *free* arrangements (Experiments 1B, 2B), the shapes in the structured set randomly appeared in one of two vertical arrangements across trials (e.g., shape A could appear either above or below shape B). In contrast to these structured conditions, shapes assigned to the unstructured set could be paired with any other shape from the unstructured set and could occupy either the top or bottom position within this random pairing.

The search display was 16em x 28em, where em is the font size on the participant's display. This size was chosen such that the display would approximately extend around 6 degrees of visual angle during typical viewing conditions. We reasoned that participants who used smaller screens had smaller font sizes and were positioned closer to the screen, such that the visual angle subtended by the relevant stimuli were roughly equated across screen sizes. Note, however, that because the study was conducted online, we could not fully control the visual angle subtended by the search display. The experiment was programmed in JavaScript with jsPsych and hosted online on Pavlovia (Open Science Tools Limited, 2021).

Procedure and design

Each trial of the visual search task started with a letter cue (1400 ms) indicating the target shape for that trial (Figure 1B). After a brief (700 ms) delay, a search display with 10 shapes appeared. Participants used the keyboard to indicate whether the target was present on the left ("F" key) or the right ("J" key) side of the display.

Each search display consisted of 4 distractor shape pairs, the target shape, and a foil shape (one of the other three target shapes not currently being searched for) arrayed symmetrically on a 4x4 grid (Figure 1C). The distractor pairs could comprise either the 4 distractor pairs from the structured set or else 4 randomly generated distractor pairs from the unstructured set. Thus, on each trial, participants searched for the target in either a structured or unstructured scene. One distractor pair appeared in each of the 4 columns of the grid, in random horizontal order. The vertical position of the pairs was random, but with the constraint that the locations were mirrored horizontally. The target appeared in one of the remaining locations vertically adjacent to a pair, with the foil (one of the other 3 targets) in the horizontally-mirrored location. The location randomization process ensured that the probability of the target's location was uniform across the entire grid.

Participants completed a total of 12 runs of the visual search task (Figure 1A). Each run consisted of 32 trials (16 trials with structured distractor pairs and 16 trials with unstructured distractor pairs), for a total of 384 trials. Structured and unstructured trials were blocked in the first 9 runs (training runs), but randomly interleaved in the last 3 runs (test runs; Figure 1A). The order of blocking (structured trials first or unstructured trials first) was maintained for a participant throughout the experiment and balanced across participants. We elected to block the structure conditions during training based on evidence that humans appear to learn statistical associations faster when these are presented in a blocked rather than interleaved order (Flesch et al., 2018). All analyses focused on responses in the interleaved test runs to avoid possible block-based differences in arousal or strategy.

The experiment started with three training runs to familiarize participants with the target letter-shape association and to practice the visual search task (phase 1 training; Figure 1A). These runs started with a familiarization block where the letters and their associated target shapes were shown sequentially four times. Next, participants completed six runs of the visual search task to (implicitly) learn the statistical regularities of the structured distractor pairs (phase

2 training; Figure 1A). These runs started with instructions and a reminder of the target letter-shape associations to refresh participants' memory. Finally, participants completed 3 test runs where the structured and unstructured conditions were interleaved.

Familiarity judgment task

After completing the 12 runs of the visual search task, participants additionally completed two runs of a two-alternative-forced-choice (2AFC) familiarity judgment task (Fiser & Aslin, 2001). This component of the experiment aimed to assess participants' explicit knowledge about the shape pairs that had appeared as distractors during the main visual search experiment. On each trial, participants were asked to indicate (or guess) within three seconds which of two shape-pairs had appeared more frequently during the visual search experiment (Figure 1D). One of the pairs was a structured pair, taken directly from the preceding visual search component, while the other was an unstructured pair. We compared the 4 structured pairs with 4 randomly selected unstructured pairs, which were held constant throughout the familiarity judgment task, such that within the task all pairs were presented equally often. The position of the shapes in the pairs was also held constant in the familiarity judgment task. Importantly, in the visual search task, the structured pairs had been presented 14 times more often than the unstructured pairs; if participants noticed these regularities, they should show above-chance performance on the familiarity judgment task (Fiser & Aslin, 2001).

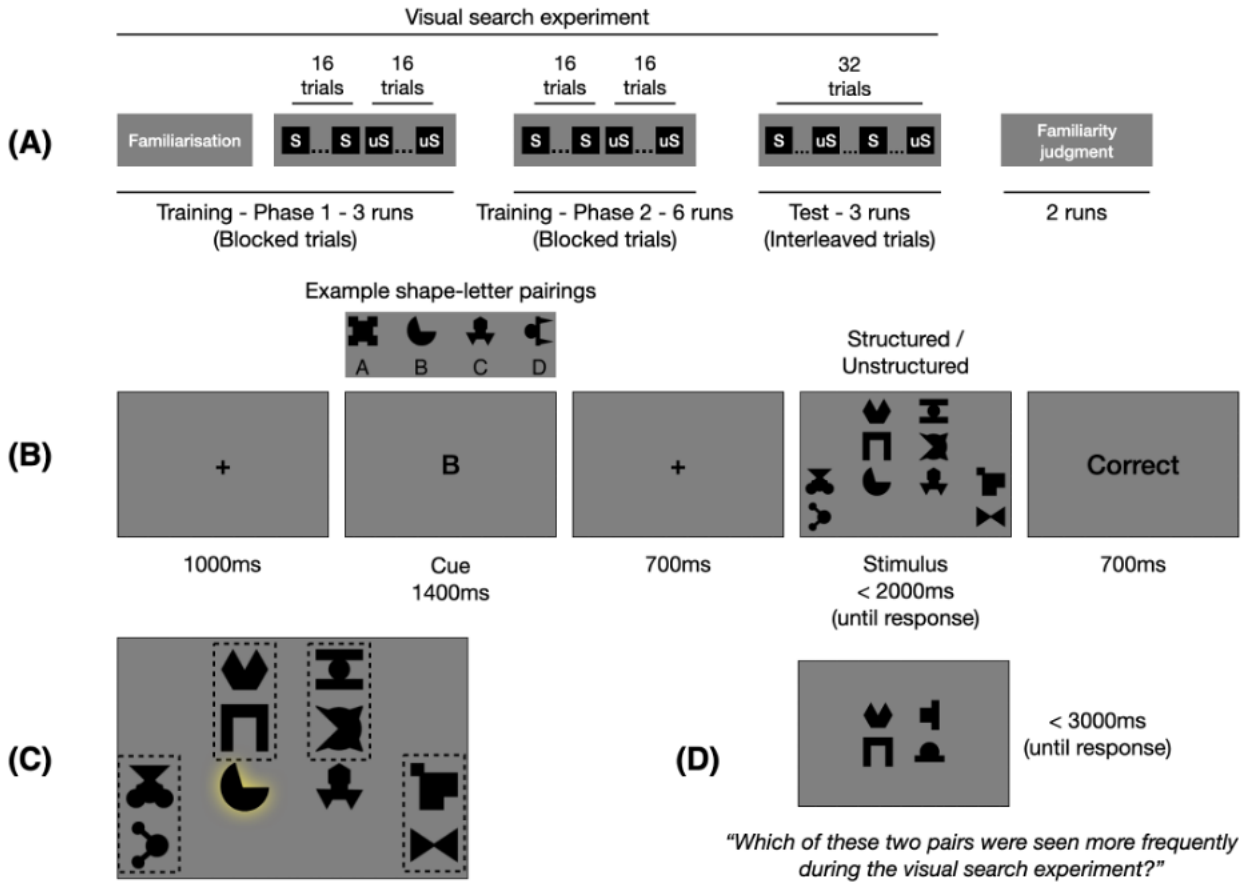

Figure 1: Experimental procedure and design. (A) Schematic outline of the experiment. S = structured scenes; uS = unstructured scenes. Structured and unstructured scenes were blocked during the first nine training runs of the visual search experiment but interleaved during the final three test runs. The visual search experiment was followed by a familiarity judgment task. (B) The trial structure of the visual search experiment. Participants had to search for a target shape cued by its corresponding letter in the subsequent search display and indicate if the target was present on the left or the right part of the display within 2s. (C) An example layout for the visual search display. Ten shapes appeared on each trial: 8 distractors, 1 target (highlighted in yellow, color not shown during the experiment), and 1 foil (which could be a target on other trials). The distractors were presented as 4 pairs. In the structured scenes, the distractors co-occurred in pairs of two (with either fixed arrangements within the pairs or not, in separate experiments). In the unstructured scenes, the distractors were randomly partitioned into four pairs on each trial. Search performance was compared between structured and unstructured displays. (D) A trial of the familiarity judgment task. Participants had to judge which of the two pairs (one taken from the structured scenes and the other from the unstructured scenes) was seen more frequently during the visual search experiment.

Each run contained trials showing the 16 possible combinations between the 4 original structured pairs and the 4 selected unstructured pairs. All analyses reported here concern these trials, as they provide the most sensitive test of familiarity and were included in all familiarity judgment experiments. The familiarity judgment experiments for the fixed arrangement condition (Experiments 2A and 3) additionally included trials in which either the position or the partner was swapped across the set of structured pairs (and, separately, the set of selected unstructured pairs). Results from these trials were not further analyzed, but trials are described here for completeness. Four partner-swapped pairs were constructed, separately for the structured and the unstructured scenes, by swapping the partners of the shapes while maintaining their relative positions in the pairs. Four position-swapped pairs were constructed, separately for the structured and the unstructured scenes, by swapping the positions of the shapes within their pairs. These two manipulations led to 32 additional comparisons between the shapes from the structured and unstructured scenes, for a total of 48 trials per run.

For the first 200 participants in Experiment 2A, these trials were presented in random order and feedback was provided at the end of each run. For the last 200 participants of Experiment 2A and for all participants in Experiment 3, the original 16 comparisons were presented at the beginning of the run, with the other conditions presented interleaved in the remainder of the run. For these participants, feedback was only provided at the end of the second run. Only participants were included who responded at least once to each condition in each run, leaving 368 of the 400 participants in Experiment 2A for the familiarity judgment analysis. Participants not meeting this requirement in Experiment 3 were replaced, such that all 260 participants were included in the familiarity judgment analysis. In Experiment 2B (free arrangement), the first half of the participants (N = 200) did not complete the familiarity judgment task. For the second half of the participants (N = 200), in each run, only the 16 original comparisons between the 4 forced pairs from the structured scenes and 4 forced pairs from the unstructured scenes were shown. All participants responded at least once in each run.

Results

Search performance as a function of learned distractor structure

To test whether the presence of co-occurring distractors facilitated visual search, we evaluated the difference between search performance in structured scenes and unstructured scenes in terms of both accuracy and reaction time (for the correct response trials) in the test runs (i.e., after a period of exposure during training). This difference was termed the *structure-benefit* (indicated by a higher search accuracy or faster reaction times in the structured scenes). We first used a two-sample t-test to assess if the structure-benefit was different across the two arrangement types for co-occurring shapes, i.e., fixed or free within their pairs. When a significant difference was identified, we conducted individual one-sample t-tests for the structure-benefits corresponding to the two types of arrangements within pairs. We used the inverse efficiency score (IES = average reaction time / average accuracy) which includes both the speed and accuracy of the search as a primary metric for the structure-benefit.

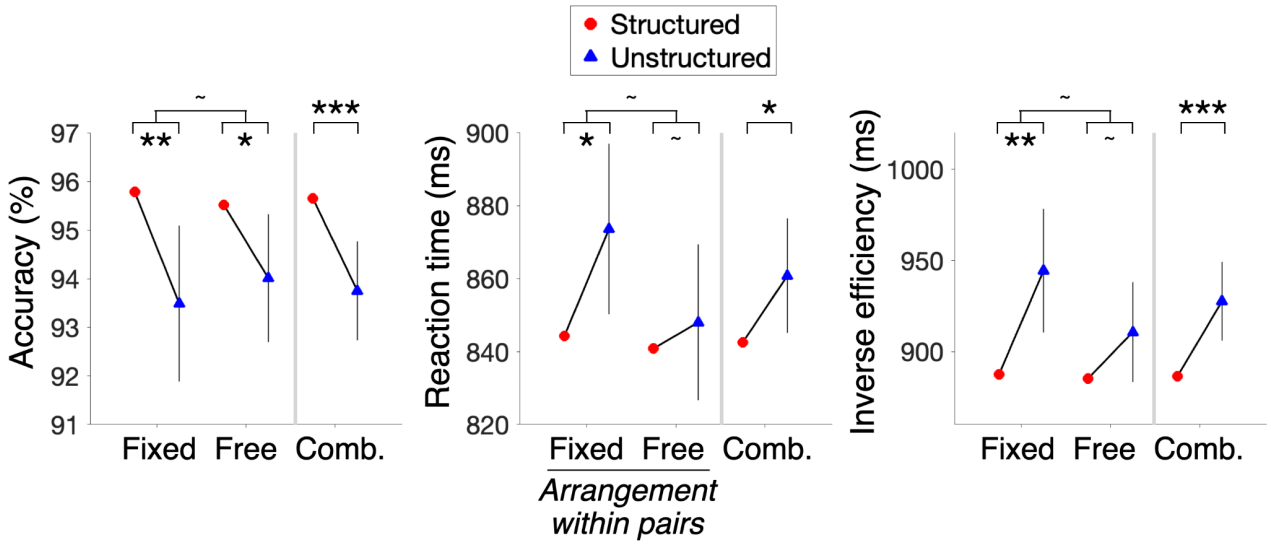

Figure 2: Search efficiency as a function of scene condition: Experiments 1A and 1B. Structure-benefit (increased accuracy or decreased reaction time or decreased inverse efficiency in the structured scenes) was observed in both the experiments with fixed or free arrangements of the co-occurring shapes within their pairs. As no differences were observed between the experiments in either of the measures, the data from the two experiments were combined ('Comb') to accumulate the evidence for the structure-benefit. Error bars indicate 95% confidence intervals for the structure-benefit on each measure, for each

experiment. The asterisks indicate p-values for the t-tests for the corresponding comparisons (*p<0.05, **p<0.01, ***p<0.001, ˜p>0.05).

In Experiments 1A and 1B (N= 40 each; Figure 2), the structure-benefit did not differ across the arrangements within pairs in IES (two-sample t-test: t_{78} = 1.38, p = 0.17, d = 0.21), accuracy (t_{78} = 0.71, p = 0.48, d = 0.11) or reaction time (t_{78} = 1.36, p = 0.18, d = 0.22). Pooling across the arrangement types, there was a highly reliable structure-benefit in IES (one-sample t-test: t_{79} = 3.7, p < 0.001, d = 0.41), which was also reflected in accuracy (t_{79} = 3.6, p < 0.001, d = 0.4) and reaction time (t_{79} = 2.3, p = 0.03, d = 0.26). Thus, these pilot experiments provided evidence that participants searched for targets more efficiently in the context of structured distractor arrays than unstructured distractor arrays, irrespective of the arrangement of pairs in the structured scenes.

Next, we conducted large-sample experiments for the two arrangement types with two goals in mind: First, to ensure that the structure-benefit observed in the pilot experiments was robust (i.e., replicable in a large sample), and second, to measure participants' familiarity with the co-occurring shapes. Here we used one-sided t-tests to test for the existence of structure-benefits, based on the direction of the effect in Experiments 1A and 1B. In these large-sample experiments (N = 400 each; Figure 3), we again observed that the structure-benefit did not differ across arrangement type within pairs in the IES (two-sample t-test: t_{798} = 0.26, p = 0.8, d = 0.01), similarly reflected in accuracy (t_{798} = 1.42, p = 0.16, d = 0.07) and reaction time (t_{798} = 1.18, p = 0.24, d = 0.06). Pooling across arrangement type, there was evidence for structure-benefit in the IES (one-sample, one-sided, t-test: t_{799} = 2.8, p = 0.003, d = 0.1), which was reflected both in accuracy (t_{799} = 2.5, p = 0.006, d = 0.09) and reaction time (t_{799} = 1.8, p = 0.04, d = 0.06). Thus, we found additional, confirmatory, evidence that after a period of exposure to distractor co-occurrence in the search displays, participants performed more efficient search in the structured scenes than the unstructured scenes. Notably, the benefit of distractor co-

occurrence was evident irrespective of whether the co-occurring shapes in the structured scenes occurred in fixed or free arrangements within their pairs.

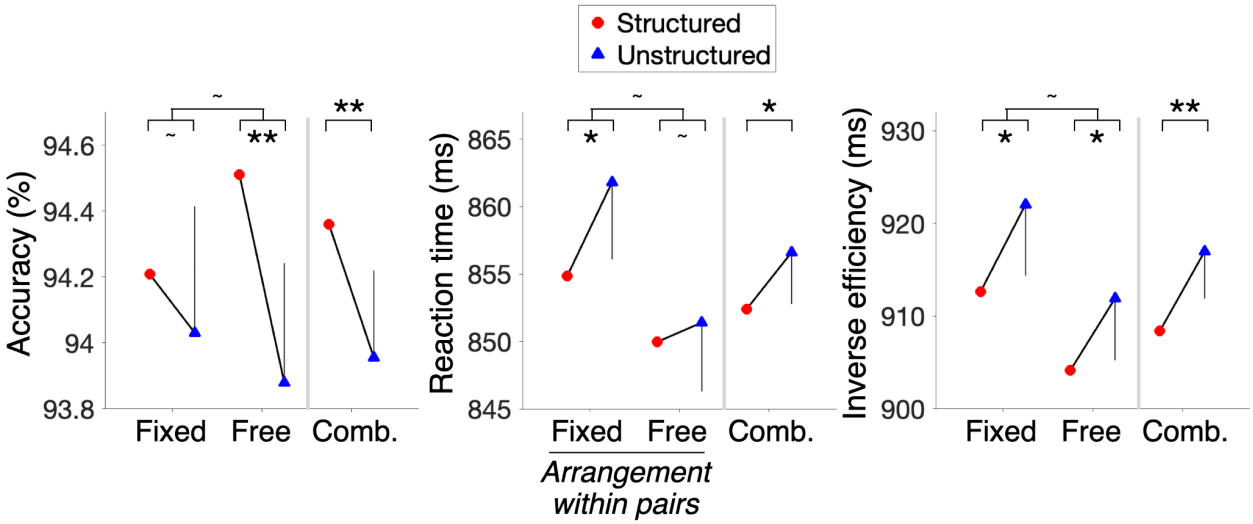

Figure 3: Search efficiency as a function of scene condition: Large-sample experiments 2A and 2B. A structure-benefit (increased accuracy or decreased reaction time or decreased inverse efficiency in the structured scenes) was present for both the fixed and free pair arrangements of co-occurring shapes, replicating the effects from the pilot experiments. As no differences were observed between the experiments in any measure, data from the two experiments were combined ('Comb') to accumulate the evidence for the structure-benefit. Error bars indicate 95% confidence intervals for the structure-benefit on each measure (corresponding to a one-sided t-test), for each experiment. The asterisks indicate p-values for the t-tests for the corresponding comparisons (*p< 0.05, **p<0.01, ***p< 0.001, ~p>0.05).

Familiarity of distractor structure and its relationship with structure-benefits in search

Could participants reliably guess which pairs co-occurred during the visual search task, as has previously been reported in experiments where the pairs were passively viewed (Fiser & Aslin, 2001)? To assess whether this was the case, we included a 2AFC pair familiarity judgment task immediately after the main visual search task (Figure 1D). We defined *familiarity score* as the proportion of responses where the pairs corresponding to the shapes from the structured scenes were selected as more familiar than the pairs corresponding to the shapes from the unstructured scenes.

Familiarity scores did not differ between Experiments 2A and 2B (two-sample t-test: $t_{566} = 0.9$, $p = 0.4$, $d = 0.08$). We therefore pooled the data across the two experiments for maximal power. Surprisingly, familiarity scores did not differ significantly from 0.5 (one-sample t-test: $t_{567} = 1.7$, $p = 0.09$, $d = 0.07$), suggesting that, on average, observers could not guess which shapes co-occurred during the search task.

Although the familiarity score was at chance level at the group level, it could be the case that participants who exhibited a higher structure-benefit in the visual search task were more familiar with the distractor co-occurrences, for example because they had paid more attention to these regularities during the visual search task. To test this, we assessed the correlation between the participants' structure-benefit reflected in IES and their familiarity score. We observed a significant negative correlation when pooling the data of Experiment 2A and 2B ($r = -0.10$, $p = 0.01$). This negative correlation was significant in Experiment 2A ($N = 368$; $r = -0.16$, $p = 0.001$; Figure 4A), but not in Experiment 2B ($N = 200$; $r = 0.02$, $p = 0.7$; Figure 4B). Thus, if anything, participants who had a stronger structure-benefit in the visual search task indicated that the structured pairs were *less* familiar than the unstructured pairs in the familiarity judgment task.

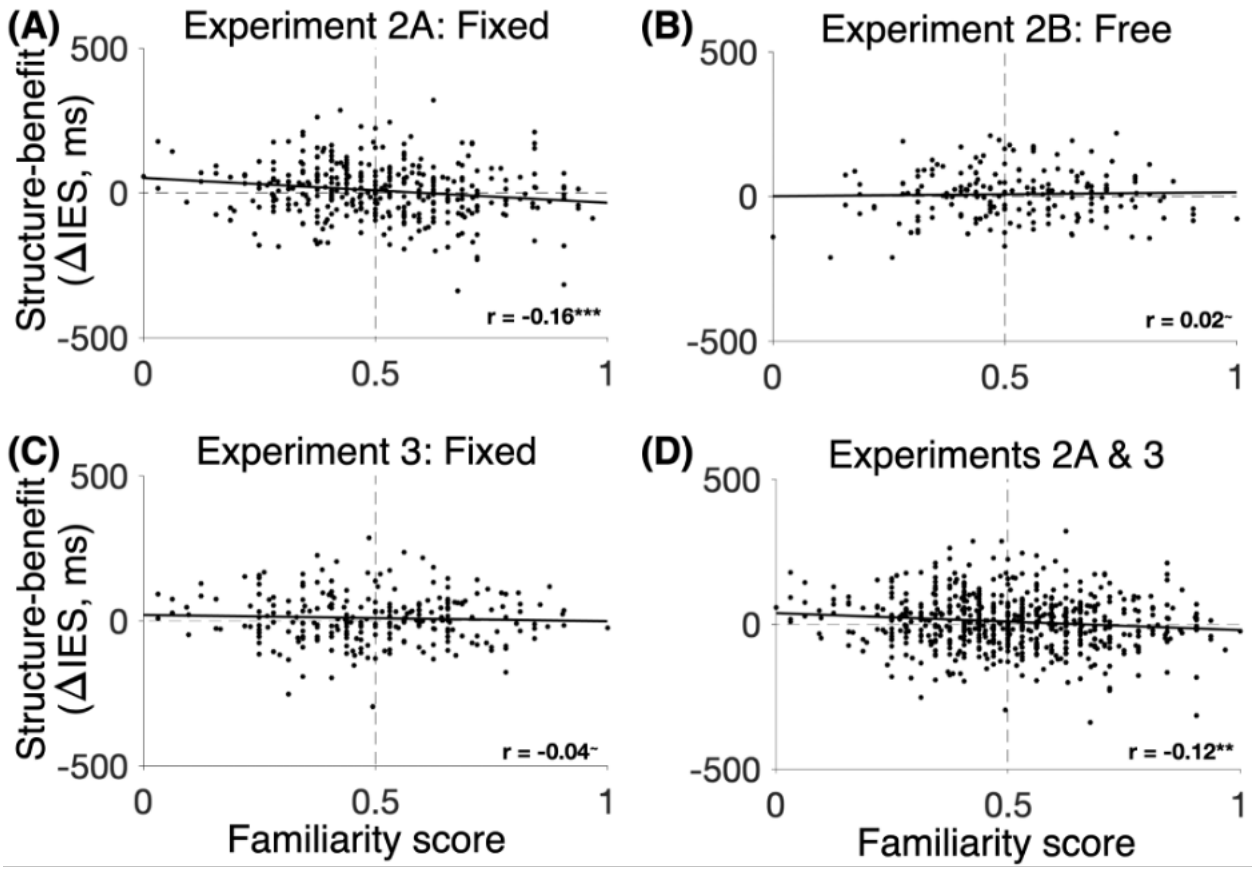

Figure 4: The relationship between the structure-benefit and 2AFC familiarity judgments about the co-occurring distractors. (A) In Experiment 2A, with the fixed arrangement of co-occurring distractors within their pairs, the structure-benefit (in the inverse efficiency score, IES) was negatively correlated with the familiarity scores. (B) In Experiment 2B, with the free arrangement of co-occurring distractors within their pairs, no such correlation was observed. (C) Experiment 3 did not replicate the negative correlation found in Experiment 2A. (D) Pooling across the two experiments, the structure-benefit was negatively correlated with the familiarity scores. The asterisks indicate p-values (*p< 0.05, **p<0.01, ***p< 0.001, ~p>0.05).

To replicate the negative correlation of Experiment 2A, we ran a pre-registered replication of Experiment 2A (Experiment 3; N = 260). As in previous experiments, the test runs of Experiment 3 demonstrated a structure-benefit in IES (one-sample, one-sided t-test: t_{259} = 1.7, p = 0.04, d = 0.11). Mirroring the findings of Experiment 2, the familiarity scores did not differ from 0.5 (one-sample t-test: t_{259} = 0.3, p = 0.79, d = 0.02). However, unlike Experiment 2A, the negative correlation between the familiarity scores and the structure-benefit was not significant in this sample (r = -0.04, p = 0.44; Figure 4C). We wondered if some difference

between the responses in Experiments 2A and 3 could explain the non-replication of the negative correlation. However, there was no difference between the two experiments in either the magnitude of the structure-benefit in IES or the familiarity scores (two-sample t-tests, structure-benefit: $t_{626} = 0.2$, $p = 0.8$, $d = 0.02$; familiarity score: $t_{626} = 0.7$, $p = 0.5$, $d = 0.06$). When pooling the data across Experiments 2A and 3, the negative correlation between the structure-benefit and the familiarity scores remained significant ($N = 628$; $r = -0.12$, $p = 0.003$). Finally, the correlation between the structure-benefit and the familiarity scores across all available data ($N = 828$; Experiments 2A, 2B, 3) was also significantly negative ($r = -0.09$, $p = 0.012$).

Discussion

Across five experiments, we found that statistical co-occurrences between distractor shapes facilitated search performance. The benefit of scene structure arose irrespective of whether the spatial arrangement of co-occurring shapes in the pairs was fixed or not. Surprisingly, the increase in search efficiency was not accompanied by an increase in participants' reported familiarity with the underlying statistical regularities (if anything, these effects were inversely related). These findings indicate that statistical regularities in the environment facilitate search even when these regularities are not explicitly noticed. The more efficient rejection of regular distractors may contribute to the efficiency of visual search in natural scenes, where such regularities are abundant (Kaiser et al., 2019).

How might reliable co-occurrences between distractor items give rise to a visual search benefit? Object grouping has been proposed as a complexity reduction mechanism supporting efficient search (Kaiser et al., 2014, 2019). Under this framework, shapes that consistently co-occur may be represented as a single object, similar to shapes that are grouped based on Gestalt cues (Wagemans et al., 2012). Support for this hypothesis comes from studies showing that fixed arrangements of co-occurring objects produce object-based attention effects (Lengyel et al., 2021). In our study, co-occurring distractor shapes in fixed arrangements produced more efficient search than randomly paired distractor shapes. However, a search benefit was also present (and not statistically different in magnitude) when the co-occurring shapes had *no* fixed arrangement, i.e., could vary freely in their spatial arrangement within the pair. The latter finding does not fit easily with an object grouping account, unless we assume that observers learned two objects, corresponding to the two configurations of the co-occurring shapes.

A possible alternative is that the search benefit reflected bidirectional associations between the shapes. Upon seeing one of the shapes, the representation of the associated shape may be primed, facilitating its recognition and subsequent rejection as a distractor when presented nearby. Such inter-object priming effects could operate weakly but in parallel across

multiple distractor locations. The learning of arbitrary associations has been linked to the hippocampus (Davachi, 2006; Eichenbaum et al., 2007), which can modulate processing in visual cortex regions (Eichenbaum et al., 2007). Accordingly, the effects revealed here may reflect facilitated visual processing of co-occurring shapes due to hippocampus-mediated predictions (Kok & Turk-Browne, 2018). This appears to be a separate mechanism from that observed in previous studies investigating the effects of real-world positional regularities, based on long-term functional and semantic associations between objects (e.g., lamp above table). There, effects of object co-occurrences were specific to familiar spatial configurations (Kaiser et al., 2014; Quek & Peelen, 2020), and may be mediated by representational changes in visual cortex (Kaiser & Peelen, 2018) rather than hippocampus-mediated associations.

The current findings contribute to the statistical learning literature (Fiser et al., 2010) by showing that statistical regularities can be learned when these occur between shapes that have to be ignored. Prior work has proposed that statistical learning is gated by visual attention (Turk-Browne et al., 2005), such that co-occurrences between items are not indicated as familiar post-experiment when the co-occurrences are task-irrelevant. Indeed, unlike studies where participants passively viewed shape combinations (Fiser & Aslin, 2001, 2005), here participants could not discriminate between familiar and unfamiliar pairs post-experiment, even though the familiar pairs had been viewed 14 times more often than the unfamiliar pairs across 9 runs. If anything, the structure benefit observed in visual search was inversely related to participants' familiarity of the shapes. A similar negative relationship between awareness of statistical regularities and the behavioral benefit of these regularities was recently observed in a contextual cueing study, where the regularities concerned target-distractor relations (Spaak & de Lange, 2020).

The negative correlation between structure-benefit and familiarity score suggests that participants who more effectively ignored the regular distractor pairs (thereby showing a greater structure benefit) later judged these pairs to be relatively unfamiliar. This finding may reflect the

effect of inhibitory attention mechanisms, which have previously been found to suppress the visual representation of ignored objects (Seidl et al., 2012) and impair subsequent judgments on these objects (Tipper, 2001). Similarly, here, inhibiting the representations of regular distractor pairs during visual search may have resulted in these objects looking relatively unfamiliar during the explicit familiarity task. It should be noted, however, that the negative correlation between structure-benefit and familiarity score was not reliably observed in the pre-registered replication experiment, such that future studies are needed to confirm this account.

In summary, we find that regularities amongst distractors in the environment can be used to reduce the complexity of a scene, facilitating search for an unrelated target. Together with the encoding of regularities between distractors and targets (e.g., contextual cueing), this may help to explain the efficiency of naturalistic visual search.

References

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. https://doi.org/10.1038/nrn1476

Bar, M., & Ullman, S. (1996). Spatial Context in Recognition. *Perception*, *25*(3), 343–352. https://doi.org/10.1068/p250343

Castelhano, M. S., & Krzyś, K. (2020). Rethinking Space: A Review of Perception, Attention, and Memory in Scene Processing. *Annual Review of Vision Science*, *6*(1), 563–586. https://doi.org/10.1146/annurev-vision-121219-081745

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

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, *16*(6), 693–700. https://doi.org/10.1016/j.conb.2006.10.012

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Donnelly, N., Humphreys, G. W., & Riddoch, M. J. (1991). Parallel Computation of Primitive Shape Descriptions. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(2), 561–570.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*(2), 161–177. https://doi.org/10.1037/0096-3445.123.2.161

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The Medial Temporal Lobe and Recognition Memory. *Annual Review of Neuroscience*, *30*(1), 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

Fiser, J., & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2005). Encoding Multielement Scenes: Statistical Learning of Visual Feature Hierarchies. *Journal of Experimental Psychology: General*, *134*(4), 521–537. https://doi.org/10.1037/0096-3445.134.4.521

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends in Cognitive Sciences*, *14*(3), 119–130. https://doi.org/10.1016/j.tics.2010.01.003

Fiser, J., & Lengyel, G. (2019). A common probabilistic framework for perceptual and statistical learning. *Current Opinion in Neurobiology*, *58*, 218–228. https://doi.org/10.1016/j.conb.2019.09.007

Flesch, T., Balaguer, J., Dekker, R., Nili, H., & Summerfield, C. (2018). Comparing continual task learning in minds and machines. *Proceedings of the National Academy of Sciences*, *115*(44). https://doi.org/10.1073/pnas.1800755115

Gayet, S., & Peelen, M. V. (2022). Preparatory attention incorporates contextual expectations. *Current Biology*, *32*(3), 687-692.e6. https://doi.org/10.1016/j.cub.2021.11.062

Geng, J. J., & Witkowski, P. (2019). Template-to-distractor distinctiveness regulates visual search efficiency. *Current Opinion in Psychology*, *29*, 119–125. https://doi.org/10.1016/j.copsyc.2019.01.003

Humphreys, G. W., Quinlan, P. T., & Riddoch, M. J. (1989). Grouping Processes in Visual Search: Effects With Single- and Combined-Feature Targets. *Journal of Experimental Psychology: General*, *118*(3), 258–279.

Kaiser, D., & Peelen, M. V. (2018). Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *NeuroImage*, *169*, 334–341. https://doi.org/10.1016/j.neuroimage.2017.12.065

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object Vision in a Structured World. *Trends in Cognitive Sciences*, *23*(8), 672–685. https://doi.org/10.1016/j.tics.2019.04.013

Kaiser, D., Stein, T., & Peelen, M. V. (2014). Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proceedings of the National Academy of Sciences*, *111*(30), 11217–11222. https://doi.org/10.1073/pnas.1400559111

Kok, P., & Turk-Browne, N. B. (2018). Associative Prediction of Visual Shape in the Hippocampus. *The Journal of Neuroscience*, *38*(31), 6888–6899. https://doi.org/10.1523/JNEUROSCI.0163-18.2018

Lengyel, G., Nagy, M., & Fiser, J. (2021). Statistically defined visual chunks engage object-based attention. *Nature Communications*, *12*(1), 272. https://doi.org/10.1038/s41467-020-20589-z

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527. https://doi.org/10.1016/j.tics.2007.09.009

Peelen, M. V., & Kastner, S. (2014). Attention in the real world: Toward understanding its neural basis. *Trends in Cognitive Sciences*, *18*(5), 242–250. https://doi.org/10.1016/j.tics.2014.02.004

Quek, G. L., & Peelen, M. V. (2020). Contextual and Spatial Associations Between Objects Interactively Modulate Visual Processing. *Cerebral Cortex*, *30*(12), 6391–6404. https://doi.org/10.1093/cercor/bhaa197

Rauschenberger, R., & Yantis, S. (2006). Perceptual encoding efficiency in visual search. *Journal of Experimental Psychology: General*, *135*(1), 116–131. https://doi.org/10.1037/0096-3445.135.1.116

Schapiro, A., & Turk-Browne, N. (2015). Statistical Learning. In *Brain Mapping* (pp. 501–506). Elsevier. https://doi.org/10.1016/B978-0-12-397025-1.00276-1

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1–2), 1–46. https://doi.org/10.1016/S0010-0277(00)00152-9

Seidl, K. N., Peelen, M. V., & Kastner, S. (2012). Neural Evidence for Distracter Suppression during Visual Search in Real-World Scenes. *Journal of Neuroscience*, *32*(34), 11812–11819. https://doi.org/10.1523/JNEUROSCI.1693-12.2012

Spaak, E., & de Lange, F. P. (2020). Hippocampal and Prefrontal Theta-Band Mechanisms Underpin Implicit Spatial Context Learning. *The Journal of Neuroscience*, *40*(1), 191–202. https://doi.org/10.1523/JNEUROSCI.1660-19.2019

Spaak, E., Peelen, M. V., & de Lange, F. P. (2022). Scene Context Impairs Perception of Semantically Congruent Objects. *Psychological Science*, *33*(2), 299–313.

Stein, T., Kaiser, D., & Peelen, M. V. (2015). Interobject grouping facilitates visual awareness. *Journal of Vision*, *15*(8), 10. https://doi.org/10.1167/15.8.10

Tipper, S. P. (2001). Does Negative Priming Reflect Inhibitory Mechanisms? A Review and Integration of Conflicting Views. *The Quarterly Journal of Experimental Psychology Section A*, *54*(2), 321–343. https://doi.org/10.1080/713755969

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The Automaticity of Visual Statistical Learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Võ, M. L.-H., Boettcher, S. E., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology*, *29*, 205–210. https://doi.org/10.1016/j.copsyc.2019.03.009

Võ, M. L.-H., & Wolfe, J. M. (2013). Differential Electrophysiological Signatures of Semantic and Syntactic Scene Processing. *Psychological Science*, *24*(9), 1816–1823. https://doi.org/10.1177/0956797613476955

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Wolfe, J. M. (1998). What Can 1 Million Trials Tell Us About Visual Search? *Psychological Science*, *9*(1), 33–39. https://doi.org/10.1111/1467-9280.00006

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics*, *73*(6), 1650–1671. https://doi.org/10.3758/s13414-011-0153-3

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84. https://doi.org/10.1016/j.tics.2010.12.001

Zhao, J., & Yu, R. Q. (2016). Statistical regularities reduce perceived numerosity. *Cognition*, *146*, 217–222. https://doi.org/10.1016/j.cognition.2015.09.018