



Real-world object categories and scene contexts conjointly structure statistical learning for the guidance of visual search

Ariel M. Kershner¹ · Andrew Hollingworth¹

Accepted: 12 March 2022 / Published online: 14 April 2022
© The Psychonomic Society, Inc. 2022

Abstract

We examined how object categories and scene contexts act in conjunction to structure the acquisition and use of statistical regularities to guide visual search. In an exposure session, participants viewed five object exemplars in each of two colors in each of 42 real-world categories. Objects were presented individually against scene context backgrounds. Exemplars within a category were presented with different contexts as a function of color (e.g., the five red staplers were presented with a classroom scene, and the five blue staplers with an office scene). Participants then completed a visual search task, in which they searched for novel exemplars matching a category label cue among arrays of eight objects superimposed over a scene background. In the *context-match condition*, the color of the target exemplar was consistent with the color associated with that combination of category and scene context from the exposure phase (e.g., a red stapler in a classroom scene). In the *context-mismatch condition*, the color of the target was not consistent with that association (e.g., a red stapler in an office scene). In two experiments, search response time was reliably lower in the context-match than in the context-mismatch condition, demonstrating that the learning of category-specific color regularities was itself structured by scene context. The results indicate that categorical templates retrieved from long-term memory are biased toward the properties of recent exemplars and that this learning is organized in a scene-specific manner.

Keywords Visual search · Statistical learning · Categorical cuing

Introduction

To perform most real-world activities, people must find and attend to objects that match current goals. Over the last 20 years or so, it has become clear that the guidance of attention to relevant objects is driven not only by stimulus salience and top-down templates, but also by the history of previous selective actions, i.e., *selection history* (Awh et al., 2012; Failing & Theeuwes, 2018; Le Pelley et al., 2016). Core phenomena of this type include *inter-trial effects* (Kristjansson et al., 2002; Li & Theeuwes, 2020; Talcott & Gaspelin, 2020), *reward learning* (Anderson et al., 2011; Hickey et al., 2010), *learned distractor rejection* (Gaspelin et al., 2015; Stilwell et al., 2019; Wang & Theeuwes, 2018), and target *probability cuing* (Geng & Behrmann, 2005; Jiang et al., 2013).

These phenomena show that the human visual system tracks recent statistical regularities predicting the properties that are likely to be associated with task-relevant objects, and that this learning can play a major role in where, and to what objects, attention is directed. However, to be of any practical use in real-world visual search, such learning must be *structured*, because the visual world is itself structured by elements such as scene context and object category. As an example of contextual structure, learning that targets in a kitchen have tended to appear near the sink may predict the location of the next target in the kitchen, but it does not provide much information about the likely location of targets when the context changes to a park. Similarly, for target category structure, learning that recent car targets have tended to be red may help predict the color of the next car, but it does not provide much predictive value when the target category changes to a shoe or a cat.

In the literature on attention guidance by learning and history, there has been extensive work on the structural role of scene context in statistical learning of target properties, broadly collected under the term “contextual cuing” (for a review,

✉ Ariel M. Kershner
ariel-kershner@uiowa.edu

¹ Department of Psychological and Brain Sciences, The University of Iowa, Iowa City, IA 52242, USA

see Sisk et al., 2019). Most of this work has focused on contextual structure in the learning of target position regularities (e.g., Brockmole et al., 2006; Chun & Jiang, 1998), though a smaller group of studies has focused on the learning of surface feature properties, such as object shape (Chun & Jiang, 1999) or rewarded color (Anderson, 2015).

In contrast with this extensive literature, there has been relatively little work conducted to understand how target object *category* structures the acquisition of recent statistical properties to guide visual search. Zelinsky and colleagues pioneered work on the role of object category in visual search, but this has tended to focus on the role of mature category representations rather than on the learning of recent statistical regularities. Using real-world images of teddy bears as targets, Yang and Zelinsky (2009) showed that visual search could be guided, visually, to targets that were defined only by their category “teddy bear.” One plausible mechanism by which this occurs is through retrieval of long-term visual representations of teddy bears (either as individual exemplars or as a category prototype), which then functions as a template to guide attention towards targets with similar visual properties in the search display. Consistent with this view, further work on categorical search has shown that attention is guided toward objects in the search array that share visual features with the target category (Alexander & Zelinsky, 2011), especially typical features of that category (Maxfield et al., 2014), and that attention is guided best to the target when it is cued at the basic level, presumably because visual variability increases at the superordinate levels (e.g., all chairs have legs but not all furniture has legs) (Yu et al., 2016).

Recently, Bahle et al. (2021) examined how the learning of new statistical regularities biases the expression of this type of category-specific template representation. The experiments were divided into two sessions, an exposure session and a visual search session. In the former, participants viewed six photographs of objects from each of 40 familiar real-world categories (e.g., “cat,” “chair”). The objects were presented individually, and participants simply categorized each as “natural” or “man-made.” Critically, the exemplars from a category had a similar color (e.g., all six chairs were black). In the search session, participants completed a categorical search task (Yang & Zelinsky, 2009). They were shown a category label cue on each trial (e.g., “chair”) and searched through an object array for any category member. Critically, the color of the category member in the search array either matched (e.g., black chair) or mismatched (e.g., brown chair) the color of the category exemplars from the exposure session. Search was reliably faster in the match condition, indicating that participants had acquired color regularities from the exposure session, that these regularities were organized by object category, and that category-specific learning influenced the formation of the visual template guiding search. In analogy to the term “contextual cuing,” where recent statistical regularities are organized by

context, Bahle et al. (2021) termed these processes “categorical cuing,” because category-specific learning cued the probable features of the target object, facilitating search. In general, the results indicate that the long-term category representations guiding visual search are surprisingly malleable and sensitive to recent statistics. Such sensitivity could be implemented either by preferential retrieval of recent exemplars (in an exemplar-based model of category structure) or by modification of a summary representation of the category (in a prototype model).

The effects in Bahle et al. (2021) were further notable because: (1) the bias toward the properties of recent exemplars was observed for highly familiar, over-learned categories; (2) there was a relatively large set of structural units over which learning occurred (40 categories and 40 colors); (3) the learning specifically influenced the guidance of attention, with the effect primarily attributable to differences in the time required to orient attention and gaze to the target; and (4) learning transferred across tasks, from a superordinate-level classification task to a visual search task. Furthermore, category-specific learning was extended to multiple recent colors within each category. That is, match effects were observed when participants were exposed to exemplars of two different colors in each category; search was more rapid for either exposed color relative to a third, novel color.

Categorical and contextual structure in the learning of recent statistical regularities have been thus far studied separately, but it is plausible that they will interact in visual search: the learning of category-specific regularities could itself be structured by search context. For example, one might observe that highlighters in Clyde’s office tend to be green, whereas highlighters in Jenn’s office tend to be yellow, leading to the formation of search templates that differ on the dimension of color when searching for a highlighter in one office versus the other. Addressing this issue is theoretically important, because it helps distinguish between an account of statistical learning effects on visual search in which different sources of learning are applied independently versus an account in which they are dependent. Moreover, evidence for dependency would illuminate the nature of the memory representations’ function in generating learning and selection history effects, indicating that information about recent contexts and target features are stored in a bound, episodic format. Consistent with this possibility is evidence that reward learning effects in visual search are applied in a scene-specific manner (Anderson, 2015). In sum, the present research question advances understanding of how the multiple structural constraints inherent in real-world environments are combined to guide visual search.

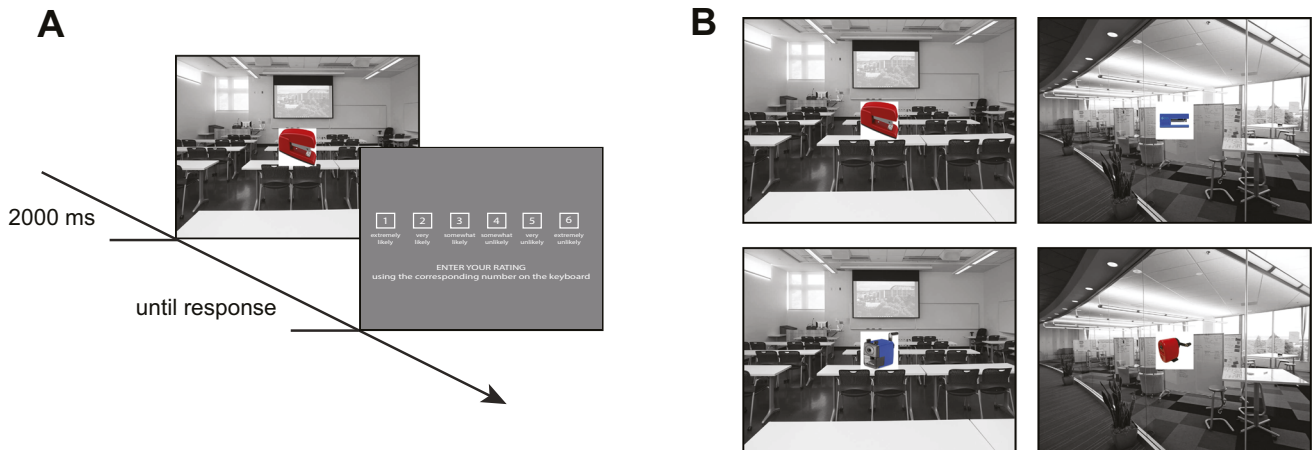
Experiment 1

In Experiment 1, we investigated the possible joint constraint of context and category in the learning and application of

statistical regularities guiding visual search (Fig. 1). In an exposure session, participants viewed 420 object exemplars:

five in each of two different colors for each of 42 different real-world categories. Each object was presented against a

Exposure Session



Visual Search Session

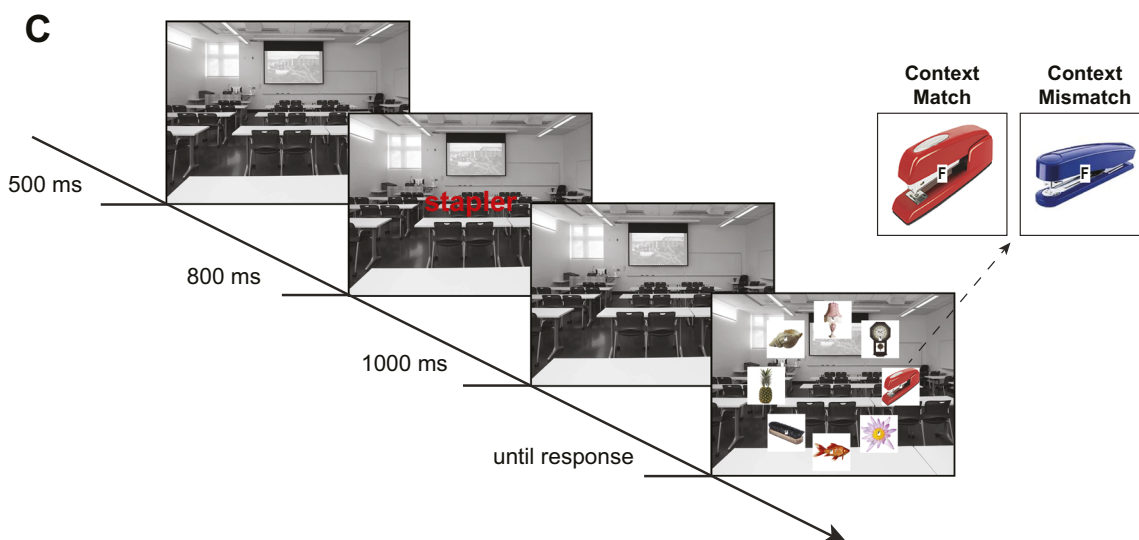


Fig. 1 Overview of method and design of Experiment 1. **a** Participants first completed an exposure session, in which they viewed 420 objects: five object exemplars in each of two colors in each of 42 different categories. The objects were presented against scene backgrounds for 2 s each. The participants completed a Plausibility-Rating task, in which they rated how likely it would be to encounter an object of that type in a scene of that type on a scale of 1 (extremely likely) to 6 (extremely unlikely). **b** In the exposure session, two categories were paired that had exemplars with the same two possible colors (e.g., red or blue staplers or pencil sharpeners). These two categories were paired with two different scene background photographs in which each object type might plausibly appear (e.g., classroom and office). The assignment of object colors to scene backgrounds was complementary. For example, in the exposure

session red staplers appeared against the classroom background and blue staplers against the office background. This assignment was reversed for sharpeners: blue against the classroom and red against the office. **c** Participants then completed a visual search session. On each trial, they first saw a scene background for 500 ms, then a text cue describing the target category for 800 ms, followed by a 1 s delay and a search array of eight objects. They searched for the object that matched the category label and reported the orientation of a superimposed letter “F”. The target object in the search array either matched or mismatched the category-specific color of exemplars associated with that background during the exposure session. Note that the category label was always presented in red font color and did not cue the color of the target object.

scene background photograph. To ensure that participants attended to the relationship between object and scene, their task in the exposure session was to rate the plausibility that an object of that type would be found in a scene of that type.

The associations between category-specific colors and scenes in the exposure session were structured as follows. Two categories were paired that had exemplars with the same two possible colors (e.g., red or blue staplers and red or blue pencil sharpeners). These two categories were matched with two different scene background photographs in which each object type might plausibly appear (e.g., classroom and office). The assignment of object colors to scene backgrounds was complementary. For example, red staplers appeared against the classroom background and blue staplers against the office background. This assignment was reversed for sharpeners: blue against the classroom and red against the office. Thus, each scene background was associated with exemplars of both colors, but from different categories.

Participants then completed a visual search session, in which the targets were new exemplars from the object categories used in the exposure phase. They were cued with a category label (e.g., “stapler”) displayed against a scene context background. Then, they searched through an array of eight objects to find the target and report the orientation of a superimposed letter. We manipulated the consistency between the scene background and the target color. In the *context-match condition*, the target color was consistent with the color associated with that combination of category and scene background from the exposure session (e.g., a red stapler target presented against the classroom background). In the *context-mismatch condition*, the color of the target was not consistent with that association (e.g., a red stapler target presented against the office background).

If the statistical learning of recent, category-specific color regularities is organized by scene context, when participants view the search target label presented against a scene background, they should tend to instantiate a search template that is biased toward the color of items from that category previously associated with that context, leading to more efficient guidance, and thus lower RT, in the context-match condition than in the context-mismatch condition.

Method

Participants Participants (18–30 years old) were recruited from the University of Iowa undergraduate subject pool and received course credit. All participants reported normal or corrected-to-normal vision. Human subjects’ procedures were approved by the University of Iowa Institutional Review Board. We collected data from 60 participants to ensure sufficient power to detect a small-to-medium-sized effect in the central contrast of interest. Seven participants

were replaced for failing to meet an a priori criterion of 85% accuracy in the search task. Participant gender was not collected.

Apparatus Due to novel coronavirus restrictions, the experiment was conducted online. It was programmed with OpenSesame software (Mathôt et al., 2012) and converted to Javascript for web-based delivery on a JATOS server maintained by the University of Iowa. Because participants completed the experiment using their own computers, we report stimulus size in absolute pixel values.

Stimuli. The stimulus set comprised 504 object images and 42 scene backgrounds. In addition, there were 150 distractor objects (75 artifact, 75 natural) for the search session that did not overlap with the experimental categories. Most stimuli were adapted from the set used in Bahle et al. (2021). Additional object and scene background images were acquired using Google image search and existing photo databases, such as Adobe Stock images. Each object image was sized to fit within a 150×150 pixel square and was presented against a white background within that square region. There were 42 object categories (22 natural and 20 artifact) and 12 exemplars in each category, six in each of the two colors per category (see Appendix Tables 2 and 3 for a complete list of categories, colors, and scene contexts). The colors for each category were chosen so that there was significant color variability across categories. For each participant, five of the six exemplars from each color in each category were randomly chosen for the exposure session. The final exemplar was assigned to the search session.

Exposure session. For the exposure session, object categories were paired, and each category within a pair had the same possible two colors. Colors were then assigned in a complementary fashion to two scene backgrounds (e.g., red staplers and blue sharpeners against the classroom background; blue staplers and red sharpeners against the office background). There were two possible configurations of this type for each pair of categories, and this was chosen randomly for each pair for each participant. In this design, since each scene was associated with the two possible colors, any effect of color match in the search session must have been mediated by object category. Scene context backgrounds ($1,024 \times 768$ pixels) were presented in grayscale to avoid interactions with the target color manipulation. The object exemplar was presented centrally, superimposed over the background image.

Search session. For the search session, eight objects were presented on a virtual circle (radius of 300 pixels), again superimposed over a scene context background. The location of the first object was selected randomly within a range of 1° to 45° , with the remaining objects each offset by 45° around the virtual circle. All arrays contained one target item matching the category label cue. Seven distractor objects were

chosen randomly without replacement from the set of 150 distractors. Each search array contained a total of four artifacts and four natural objects. For example, if the target was an artifact, three artifacts and four natural objects were chosen from the set of distractors. Target and distractor locations were also chosen randomly. A small, black letter “F” on a white background (Arabic font, approximately 16×22 pixels) was superimposed centrally on each object. The orientation of the “F” (facing left or facing right) was chosen randomly for each object. The target F was quite small, typically requiring fixation of the target object to discriminate its orientation. This was designed so that the guidance of attention would be implemented with overt shifts of gaze, which has been demonstrated to increase sensitivity to differences in attention guidance (Hollingworth & Bahle, 2020). The cue that appeared before each search array described the category of the target object (e.g., “stapler”) and was presented in red, Arabic font.

Procedure. Upon initiating the experiment, participants provided informed consent and received instructions. They were told that they would complete two sub-experiments. They then received instructions for the exposure session. Note that they did not receive instructions for the search session until after completing the exposure session. Thus, during the exposure session, they were not aware that they would subsequently perform a search task.

For the exposure session, the trial began with a screen instructing the participant to “Press Spacebar” to start the trial. After doing so, there was a 200-ms delay, followed by the object stimulus displayed against the scene background for 2,000 ms. Participants then saw a response screen asking them to rate how likely it would be to encounter an object of that type in a scene of that type on a scale of 1 (*extremely likely*) to 6 (*extremely unlikely*). (Note that, although each background was chosen as a plausible context for the object category, it was not necessarily the case that there would be a high probability of encountering the object there. For example, a bear could plausibly appear in a forest scene, but encountering a bear in any given forest is unlikely. In contrast, encountering a chair in a living room scene is very likely.) They entered the corresponding number on the keyboard.

In the exposure session, participants completed five blocks of 84 trials. In each block, they viewed one exemplar in each of the two colors for each of the 42 categories. Trials in a block were randomly intermixed. In total, there were ten exposures per category (five for each of the two colors per category). For the plausibility-rating task, mean plausibility across the categories was 2.64 (SD = 0.31).

Participants then completed the search session. Each trial began with a centrally presented “Press Spacebar” screen. Once pressed, there was a 200-ms delay before a scene background was presented for 500 ms. Then, a category label cue was centrally presented over the scene background (e.g.,

“stapler”) in red font for 800 ms, which indicated the category of the search target in the upcoming search display. The use of a category label cue required participants to retrieve a representation of the target category from memory as a template to guide visual search. Once the cue was removed, the scene background was presented alone for 1,000 ms. Finally, the search display was presented over the scene background. Participants were instructed to find the cued object and report the orientation of the “F” superimposed on it, and to do so as quickly and as accurately as possible. Participants pressed the “P” key to indicate a right-facing “F” (normal) and the “Q” key to indicate a left-facing “F” (mirror reversed).

Response terminated the search display. A smiley emoticon was displayed for 200 ms following a correct response, and a frowny emoticon was displayed for 500 ms following an incorrect response.

The search session began with instructions indicating the change in task. Participants first completed ten trials of practice using target object categories and scene backgrounds not used in the exposure session. Then, they completed one experimental block of 168 search trials. Each of the 42 categories was the target of search four times. Two trials per category were in the *context-match* condition, in which the color-category-background association from the exposure session was retained (e.g., a red stapler against the classroom and a blue sharpener against the office). Two other trials were in the *context-mismatch* condition, in which the color-background associations were reversed. Trials in the block were randomly intermixed. Each of the exemplars in the search phase was repeated once (e.g., the same red stapler exemplar was the target against the classroom in the context-match condition and against the office in the context-mismatch condition). This reduced possible variability across conditions, potentially increasing sensitivity to the effect of context match. The entire experiment lasted approximately 1 h. Participants were encouraged to take short breaks between exposure blocks and between the exposure and search sessions.

Results

Search accuracy For the visual search task, mean accuracy was 95.36% correct. The arcsine square root transformed values did not differ as a function of context match, $F(1, 59) = 1.06$, $p = .308$, $\text{adj } \eta_p^2 = .001$.

Manual response time (RT) The critical measure was mean RT in the search task as a function of context match condition. The analysis was limited to correct search trials. We also used a two-step RT trimming procedure. First, RTs shorter than 250 ms (not plausibly based on target discrimination) or longer than 6,000 ms were eliminated. Next, RTs more than 2.5

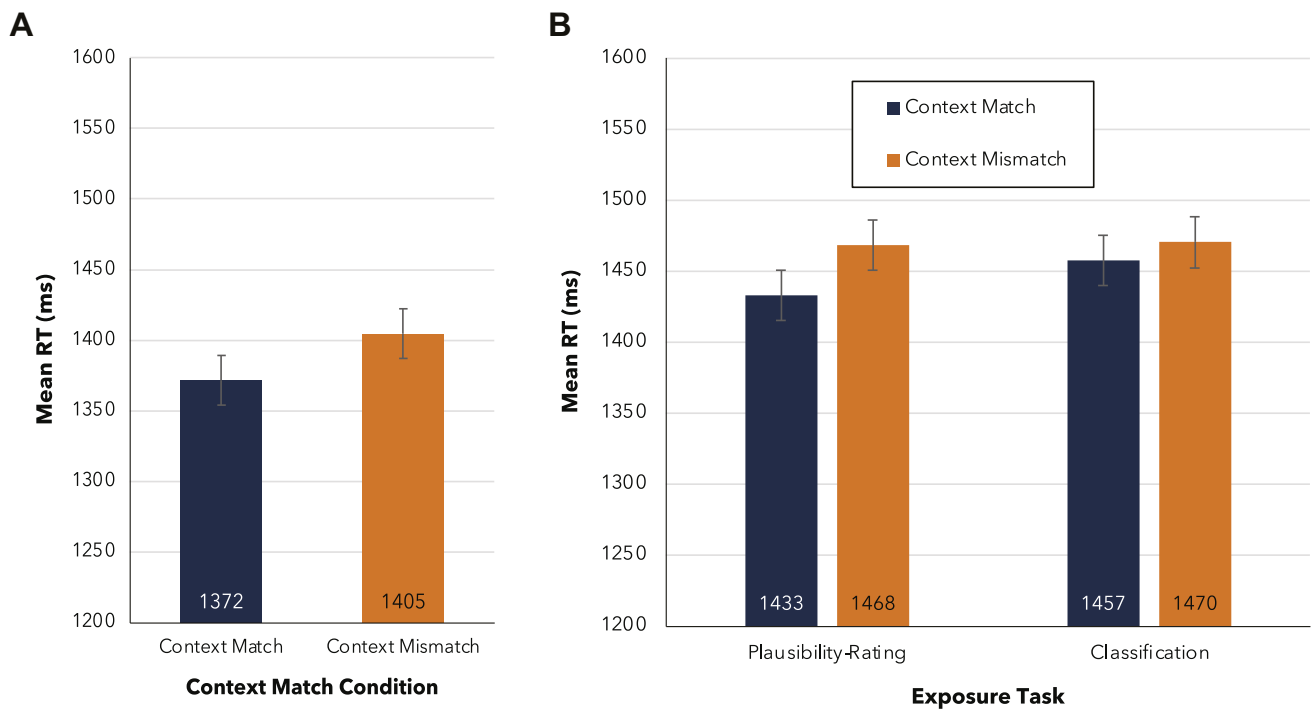


Fig. 2 Visual search results for Experiment 1 (a) and Experiment 2 (b). Mean search response time (RT) as a function of context match condition. Errors bars are condition-specific, within-subject 95% confidence intervals (Morey, 2008)

standard deviations from the participant’s mean in each condition were eliminated. A total of 8.02% of trials was eliminated. The results are reported in Fig. 2, collapsing across object type. The full set of marginal means is reported in Table 1.

Analysis 1 ANOVA. We analyzed the RT data with a 2 (context match: match, mismatch) × 2 (object type: artifact, natural) repeated-measures ANOVA, treating participant as a random effect. We included object type as a factor to examine potential differences in learning and context as a function of superordinate category, though we did not develop predictions for

this factor, as previous work has shown equivalent categorical cuing for artifacts and natural objects (Bahle et al., 2021). Adjusted η_p^2 values accompany each test (Mordkoff, 2019), correcting for the positive bias inherent in standard η_p^2 . There was a reliable main effect of context match, with lower mean RT on context-match (1,372 ms) compared with context-mismatch (1,405 ms) trials, $F(1, 59) = 6.48, p = .014, \text{adj } \eta_p^2 = .084$. There was also a reliable effect of object type, with lower mean RT for natural objects (1,371 ms) than for artifacts (1,412 ms), $F(1, 59) = 10.1, p = .002, \text{adj } \eta_p^2 = .132$. These factors did not interact, $F(1, 59) = 0.48, p = .492, \text{adj } \eta_p^2 = -0.009$.

Table 1 Marginal means for response time and search accuracy from Experiments 1 and 2

| Experiment 1 | | |
|--------------------------|-------------------------------|-------------------------------|
| | Match | Mismatch |
| Artifact | 1,340 ms (SE = 27.34); 95.34% | 1,424 ms (SE = 24.10); 94.87% |
| Natural | 1,351 ms (SE = 23.72); 95.02% | 1,391 ms (SE = 22.46); 95.87% |
| Experiment 2 | | |
| Plausibility-Rating Task | | |
| Artifact | 1,449 ms (SE = 25.13); 96.58% | 1,493 ms (SE = 26.55); 95.63% |
| Natural | 1,418 ms (SE = 18.85); 96.48% | 1,446 ms (SE = 25.21); 96.10% |
| Classification Task | | |
| Artifact | 1,487 ms (SE = 20.45); 96.21% | 1,512 ms (SE = 23.84); 96.46% |
| Natural | 1,431 ms (SE = 22.11); 95.80% | 1,434 ms (SE = 24.29); 96.63% |

Analysis 2 Mixed effects. In a complementary analysis of the RT data, we sought to draw both population inferences (from the participant sample) and inferences about the population of real-world categories (from the sample of categories). Thus, we employed a linear mixed-effects approach with a cross-classified random-effects structure, simultaneously treating participant and category item as random effects (Baayen et al., 2008). In addition, treating category item as a random effect increased our confidence that the observed results were robust not only across the set of participants but also across the set of categories.

The fixed-effects structure included context match condition and object type (natural, artifact). We then determined the random-effects structure best supported by the data. We began with the maximal random-effects structure and then simplified the model in the manner recommended by Matuschek et al. (2017), removing random-effects components that did not significantly improve model fit (via likelihood ratio test) or that produced critical failures in model convergence. The final random-effects structure included an intercept for participant, an intercept for category, and a slope for object type by participant.

Analyses were implemented with the lme4 package (version 1.1-26) in R (version 4.0.3). Degrees of freedom for the statistical tests were estimated using the lmerTest package (version 3.1-3).

There was a reliable main effect of context match condition, with lower RT on context match compared with context mismatch trials, $F(1, 9,116) = 9.13, p = .003$. There was no reliable main effect of object type, $F(1, 42.4) = 0.99, p = .326$, and no reliable interaction between object type and context match, $F(1, 9,114) = 0.48, p = .491$. Thus, the mixed-effects results support those from the ANOVA with respect to the context-match effect, and allow inferences from this sample of categories to the population of categories.

Discussion

In Experiment 1, we demonstrated that the learning of category-specific color regularities was itself structured by scene context. When searching for an object type in a scene, participants selectively retrieved, and instantiated as a template, properties of recent exemplars from that category which had appeared in that particular scene. Thus, the two sources of structure in the learning of object regularities, scene contexts and object categories, are dependent.

Experiment 2

The design of Experiment 1 meant that the two colors within a category were associated with backgrounds from *different* scene categories (e.g., red staplers with a classroom and blue

staplers with an office). In Experiment 2, we sought to associate the colors with different exemplars within a scene category. For example, red staplers in the exposure session appeared against classroom 1 and blue staplers against classroom 2. In the search session, the target object color either matched (e.g., red staplers against office 1) or mismatched (e.g., red staplers against office 2) the color-scene association. This allowed us to examine whether the structure imposed by scene context operates at the level of scene exemplars or at the level of scene categories. If the former, then we should replicate the results of Experiment 1. If the latter, then no match effect should be observed, as both colors within an object category were associated with the same scene category.

In addition to this primary goal, we sought to examine the effect of attention in the learning of object-category-to-scene associations. In the search session, one group of participants completed the plausibility-rating task used in Experiment 1, which required attending to the relationship between object and scene. A second group of participants simply classified each object as “man-made” or “natural,” which did not require attention to the background or to the relationship between object and background. Previous work has shown that attention to the relationship between two entities is often required to form an association (Gwinn et al., 2019; Rosas et al., 2013; Sisk et al., 2019)

Method

Participants We collected data from 120 participants, 60 in each exposure session task. Twelve participants were replaced for failing to meet an a priori criterion of 85% accuracy in the search task.

Apparatus Experiment 2 was also conducted online using the same apparatus.

Stimuli. The object stimulus set was comprised of 504 object images, 84 scene backgrounds, and the same set of 150 distractors as used in Experiment 1. Additional scene context images were acquired so that each category was assigned to one type of scene context (e.g., staplers to offices, sharpeners to classrooms), and each color was assigned to a different scene context exemplar (e.g., red staplers to office 1 and blue staplers to office 2). The viewpoints and general composition of the two backgrounds were chosen to be quite similar. Finally, some category colors were replaced to increase color variability. The complete set of object categories, colors, and backgrounds is listed in the Appendix Tables 2 and 3. Note that, unlike Experiment 1, each scene background was associated with only one color. Thus, this design cannot eliminate the possibility that, during search, scene context facilitated search for a particular color in general (rather than in a category-specific manner). However, the results of Experiment 1 render this possibility unlikely.

Procedure. For the exposure session, the plausibility-rating task was the same as in Experiment 1. For the classification task, participants were asked to classify the exemplar as either “Man-made” or “Natural.” They viewed a response screen similar to that for the plausibility-rating task, but with the options “1” for man-made and “6” for natural. For the plausibility-rating task, mean plausibility across the categories was 2.11 (SD = 0.48). For the classification task, mean accuracy was 96% (SD = 0.09).

Next, participants completed one experimental block of 168 search trials with the same trial structure as in Experiment 1.

Results

Search accuracy For the visual search task, mean accuracy after the classification exposure task was 96.3% correct and after the plausibility-rating task was 96.2% correct. A 2 (exposure task) \times 2 (context match) repeated-measures ANOVA was conducted over the arcsine square root transformed probabilities. There was no main effect of match, $F(1, 118) = .096$, $p = .757$, $\text{adj } \eta_p^2 = -.008$, or exposure task, $F(1, 118) = .012$, $p = .914$, $\text{adj } \eta_p^2 = -.008$. There was a reliable interaction between task and match, $F(1, 118) = 4.77$, $p = .031$, $\text{adj } \eta_p^2 = .031$. For the plausibility-rating task, there was a numerical trend toward higher accuracy in the context-match condition (96.5%, SD = 0.2%) than in the context-mismatch condition (95.9%, SD = 0.2%), $F(1, 59) = 2.29$, $p = .136$, $\text{adj } \eta_p^2 = .021$. For the classification task, there was a numerical trend toward higher accuracy in the context-mismatch condition (96.5%, SD = 0.2%) than in the context-match condition (96.0%, SD = 0.2%), $F(1, 59) = 2.53$, $p = .117$, $\text{adj } \eta_p^2 = .025$.

Manual RT The RT data were trimmed using the same procedure as in Experiment 1. A total of 7.83% of trials was eliminated. The results are presented in Fig. 2, collapsing across object type. The full set of marginal means is reported in Table 1.

Analysis 1: ANOVA. We analyzed the RT data with a 2 (exposure task) \times 2 (context match) \times 2 (object type) mixed-factor ANOVA. Again, we did not develop predictions for the object type factor. There was a reliable main effect of context match, with lower mean RT on context match (1,446 ms) compared with context mismatch (1,469 ms) trials, $F(1, 118) = 7.65$, $p = .007$, $\text{adj } \eta_p^2 = .053$. There was no reliable effect of exposure task, $F(1, 118) = 0.09$, $p = .771$, $\text{adj } \eta_p^2 = -.008$, and no reliable interaction between exposure task and context match, $F(1, 118) = 1.37$, $p = .244$, $\text{adj } \eta_p^2 = .003$. There was a reliable effect of object type, with lower mean RT for natural objects (1,432 ms) compared with artifacts (1,486 ms), $F(1, 118) = 46.96$, $p < .001$, $\text{adj } \eta_p^2 = .279$. Object type did not interact with exposure task or context match, $F(1, 118) = 3.36$, $p = .069$, $\text{adj } \eta_p^2 = .019$; $F(1, 118) = 1.43$, $p = .235$, $\text{adj } \eta_p^2 =$

.004, respectively, nor was there a three-way interaction, $F(1, 118) = 0.04$, $p = .835$, $\text{adj } \eta_p^2 = -.008$. In planned follow-up tests, the match effect was statistically reliable following the plausibility-rating task, $F(1, 59) = 7.89$, $p = .007$, $\text{adj } \eta_p^2 = .103$, but not the classification task, $F(1, 59) = 1.04$, $p = .312$, $\text{adj } \eta_p^2 = .001$.

Analysis 2: Mixed effects. The fixed-effects structure included the factorial combination of exposure task and context match condition. The final random-effects structure included an intercept for participant and an intercept for category. There was a reliable main effect of context match condition, with lower RT on context-match compared with context-mismatch trials, $F(1, 18,531) = 10.29$, $p = .001$. There was no reliable main effect of exposure task, $F(1, 118) = 0.08$, $p = .778$, and no reliable interaction between exposure task and context match, $F(1, 18,531) = 2.27$, $p = .132$. Object type did not produce a reliable main effect, $F(1, 40) = 2.20$, $p = .146$, it did not produce reliable two-way interactions with either exposure task or context match, $F(1, 18,531) = 3.78$, $p = .052$ and $F(1, 18,531) = 1.44$, $p = .229$, respectively, and there was no reliable three-way interaction, $F(1, 18,531) = 0.054$, $p = .800$.

In planned follow-up analyses, we examined the effect of context match separately for the plausibility-rating and classification tasks. There was a reliable main effect of context match in the former, $F(1, 9,263) = 11.42$, $p < .001$, but not in the latter, $F(1, 9,229) = 1.40$, $p = .236$.

Discussion

In Experiment 2, we replicated the context-match effect when the two colors within an object category were associated with different exemplars from the same scene category (rather than from different scene categories, as in Experiment 1). Thus, the results confirm that individual scene exemplars structure the acquisition of statistical regularities within object categories and that this structure influences the feature values instantiated in a categorical search template. The secondary goal of Experiment 2 was to examine the role of attention during exposure in the learning of structured statistical regularities. The “classification task” did not require attention to the relationship between object and scene background. There was no reliable context match effect in this condition, but there was a numerical trend, and there was no reliable interaction between exposure task and context match. Thus, although the results are broadly consistent with a role for attention in learning, they do not support strong conclusions on this specific question.

General discussion

Our previous work has shown that statistical learning of the surface feature properties of recently observed objects is

organized by real-world object categories, influencing visual search in a category-specific manner (Bahle et al., 2021). Such learning is also structured by scene and array context (Anderson, 2015; Chun & Jiang, 1999), consistent with the larger literature on contextual cuing. In two experiments, we demonstrated that these two forms of structure operate in a dependent manner. Visual search was influenced by within-category color regularities, and this category-level learning was contingent on the scene context in which the exemplars appeared.

The first key finding was that object category templates were biased toward the properties of recently viewed exemplars rather than depending solely on more generalized knowledge acquired over extensive experience. That is, although red may not be a frequent color for cars given one's overall experience with cars, it is possible to quickly set up a bias toward red items when searching for a car if the last few car exemplars have been red (see also Bahle et al., 2021). Note that, unlike Bahle et al., there was no baseline condition in which the target color matched neither of the exposed colors. However, the context effects observed here allow the same inference: It would not have been possible to observe a context effect if search were not guided by the color of the recent exemplars observed in that context. In addition, since the learning effects in Bahle et al. specifically influenced the guidance of attention (as assessed by eye movement measures), we can be confident that the present differences in RT were largely attributable to differences in the guidance of attention and gaze (rather than to other processes, such as post-selection target confirmation or response execution).

The second key finding was that category-specific biases were episodic in the sense of being structured by scene context. That is, the structures imposed by object category and scene context are not independent of each other; rather, category-level learning is organized by scene context. This dependency in category learning likely reflects the fact that the properties of real-world category members often vary systematically as a function of context (e.g., yellow taxis are typical in New York, whereas black taxis are typical in London). Of course, categorical search for real-world, overlearned categories will depend heavily on relatively stable representations acquired over a lifetime of experience (Yang & Zelinsky, 2009). However, the functional expression of the category representation is biased by local changes in the statistical distribution of features and to changes in context.

The incorporation of both category and contextual constraints may arise through the underlying format of the memory representation. The properties of category exemplars are likely to be stored as part of a bound, episodic representation of a scene (e.g., Hollingworth, 2006). Exemplar retrieval would then depend on the scene context that cues the previous episode (Anderson, 2015; Anderson & Britton, 2019; Bramaio et al., 2017; Godden & Baddeley, 1975; Hardt et al., 2010; Richardson & Spivey, 2000). In turn, a bias to retrieve

exemplars associated with the current scene would, in the present design, tend to lead to retrieval of exemplars of one color and not the other, producing the present effects. Although this account places exemplar retrieval at the heart of the observed results, we do not consider the data as mediating between competing exemplar (e.g., Medin & Schaffer, 1978; Nosofsky, 1987) and prototype (e.g., Minda & Smith, 2001; Rosch, 1975) theories of categorization. For example, the results could be accommodated by a prototype model assuming that retrieval of a small number of highly accessible exemplars can influence the use of the category in addition to that derived from a more stable summary representation (e.g., Allen & Brooks, 1991).

Currently, there is conflicting evidence concerning whether learning of and guidance by statistical regularities is driven by implicit or explicit memory. In the contextual cuing literature, learning was initially thought to be implicit, but there is evidence that the magnitude of the effect correlates positively with explicit awareness (Annac et al., 2019; Vadillo et al., 2016), although this correlation is not always observed (Colagiuri & Livesey, 2016). In addition, contextually specific guidance effects are observed both when participants are aware of the associations (e.g., Brockmole & Henderson, 2006) and when awareness is much more limited (e.g., Chun & Jiang, 1998). Here, we focused on the guidance process itself rather than on questions of implicit versus explicit memory, and thus we did not include a test probing explicit memory. Moreover, such a test would have needed to have been administered between the exposure and search sessions, because the associations changed in the search session. This would have delayed and potentially contaminated the transfer of learning across tasks, because test items instantiating different associations would have been necessary. The issue of awareness could be addressed more directly in a modified version of the categorical cuing paradigm that implements a repeated search design (similar to contextual cuing), where explicit memory for category-color consistencies could be assessed at the end of the experiment. The advantage of the current, two-session design is that it demonstrates cross-task transfer that is often absent in other forms of statistical learning.

Finally, we observed a reliable context match effect in the plausibility-rating task, when participants needed to attend to the association between the scene context and the category color during the exposure phase. No reliable context match effect was observed in the classification task, when attending to the relationship was not required to complete the exposure task. The between-task interaction did not reach reliability, limiting our ability to draw strong conclusions about a difference in the context match effect as a function of attention. In the reward learning literature, there is some evidence that context-specific learning depends on attending to the association between context and reward value (Gwinn et al., 2019). Our results are suggestive that attention may play a role in the context-specific learning of category-specific regularities, but this remains an open question.

Appendix

Table 2 Target object types, categories, colors, and corresponding scene contexts used in Experiment 1

| Category 1 | Category 2 | Context 1 | Context 2 | Color 1 | Color 2 |
|-------------|---------------|---------------|------------------|---------|---------|
| Horse | Dog | Stable | Yard | Brown | Black |
| Bed Frame | Leather Chair | Empty Room | Living Room | Black | Brown |
| Bean | Onion | Grocery Store | Vegetable Garden | Yellow | Red |
| Watch | Backpack | Library | Locker | Black | Yellow |
| Bell Pepper | Pear | Farm | Fridge | Green | Yellow |
| Apple | Grape | Carnival | Kitchen | Green | Red |
| Snake | Frog | Water | Pond | Green | Brown |
| Potato | Mushroom | Pantry | Factory | Red | Brown |
| Dress Shirt | Perfume | Closet | Makeup Area | Purple | Green |
| Cup | Pot | Dining Room | Stove | Black | Grey |
| Cat | Laptop | House | Electronic Store | Black | Grey |
| Shoe | Hairbrush | Foyer | Bathroom | Red | Blue |
| Sharpener | Stapler | Classroom | Office | Blue | Red |
| Car | MP3 | Parking Lot | Farmers Market | Blue | Red |
| Rat | Rabbit | Alley | Flower Garden | Black | Brown |
| Crab | Beetle | Grass | Tree | Blue | Red |
| Bird | Butterfly | Birdhouse | Sky | Red | Blue |
| Dress | T-Shirt | Bedroom | Dresser | Blue | Yellow |
| Camera | Hat | Art Studio | Mall | Black | Blue |
| Tricycle | Leaf | Driveway | Street | Yellow | Red |
| Bear | Squirrel | Forest | Mountain | White | Brown |

Table 3 Target object types, categories, colors, and corresponding scene contexts used in Experiment 2

| Artifact/ Natural | Target | Set 1 | Set 2 | Scene Context | |
|----------------------|-------------|-----------|--------|-------------------|-------------|
| Natural | Apple | Green | Red | Carnival | |
| | Bean | Yellow | Red | Grocery Store | |
| | Bear | Black | Brown | Forest | |
| | Beetle | Green | Red | Tree | |
| | Bell Pepper | Green | Yellow | Farm | |
| | Bird | Brown | Blue | Birdhouse | |
| | Butterfly | Blue | Orange | Sky | |
| | Cat | Black | Orange | House | |
| | Cherry | Black | Red | Farmer's Market | |
| | Crab | Blue | Red | Water | |
| | Dog | Black | Brown | Yard | |
| | Frog | Brown | Green | Pond | |
| | Grape | Green | Red | Kitchen | |
| | Horse | Black | Brown | Stable | |
| | Leaf | Green | Red | Street | |
| | Mushroom | Brown | Red | Factory | |
| | Onion | Red | Yellow | Vegetable Garden | |
| | Pear | Yellow | Green | Fridge | |
| | Potato | Brown | Red | Pantry | |
| | Rabbit | Brown | Black | Flower Garden | |
| | Rat | Brown | Black | Alley | |
| | Snake | Brown | Green | Grass | |
| | Artifact | Backpack | Black | Yellow | Locker |
| | | Bed Frame | Black | Brown | Empty Room |
| | | Camera | Black | Purple | Art Studio |
| | | Car | Blue | Red | Parking Lot |
| Cup | | Black | Green | Dining Room | |
| Dress | | Blue | Yellow | Bedroom | |
| Dress Shirt | | Purple | Green | Closet | |
| Hairbrush | | Blue | Red | Bathroom | |
| Hat | | Blue | Brown | Mall | |
| Laptop | | Black | Red | Electronics Store | |
| Leather Chair | | Black | Brown | Living Room | |
| MP3 Player | | Blue | Red | Recording Studio | |
| Perfume | | Red | Purple | Make-up Store | |
| Pot | | Black | Red | Stove | |
| Sharpener | | Blue | Red | Classroom | |
| Shoe | | Red | Blue | Foyer | |
| Stapler | | Blue | Green | Office | |
| T-Shirt | | Red | Yellow | Dresser | |
| Tricycle | | Yellow | Blue | Driveway | |
| Watch | | Black | Gold | Library | |

Code availability The codes for the experiments reported here are available upon request.

Funding This material is based upon work supported by a National Science Foundation Graduate Fellowship under Grant No. DGE-1945994.

Data availability The data and materials for the experiments reported here are available upon request.

Declarations

Conflicts of interest There are no conflicts of interest to report.

Ethics approval Approval was granted by the University of Iowa Institutional Review Board.

Consent to participate All participants provided informed consent before participating.

Consent for publication Not applicable.

References

- Alexander, R. G., & Zelinsky, G. J. (2011). Visual similarity effects in categorical search. *Journal of Vision*, *11*(8), 1–15. <https://doi.org/10.1167/11.8.9>
- Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, *120*(1), 3–19. <https://doi.org/10.1037/0096-3445.120.1.3>
- Anderson, B. A. (2015). Value-driven attentional priority is context specific. *Psychonomic Bulletin & Review*, *22*(3), 750–756. <https://doi.org/10.3758/s13423-014-0724-0>
- Anderson, B. A., & Britton, M. K. (2019). Selection history in context: Evidence for the role of reinforcement learning in biasing attention. *Attention, Perception, & Psychophysics*, *81*(8), 2666–2672. <https://doi.org/10.3758/s13414-019-01817-1>
- Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(25), 10367–10371. <https://doi.org/10.1073/pnas.1104047108>
- Annac, E., Pointner, M., Khader, P. H., Müller, H. J., Zang, X., & Geyer, T. (2019). Recognition of incidentally learned visual search arrays is supported by fixational eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(12), 2147–2164. <https://doi.org/10.1037/xlm0000702>
- Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. <https://doi.org/10.1016/j.tics.2012.06.010>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bahle, B., Kershner, A. M., & Hollingworth, A. (2021). Categorical cuing: Object categories structure the acquisition of statistical regularities to guide visual search. *Journal of Experimental Psychology: General*, *150*(12), 2552–2566. <https://doi.org/10.1037/xge0001059>
- Bramao, I., Karlsson, A., & Johansson, M. (2017). Mental reinstatement of encoding context improves episodic remembering. *Cortex*, *94*, 15–26. <https://doi.org/10.1016/j.cortex.2017.06.007>
- Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*(1), 99–108. <https://doi.org/10.1080/13506280500165188>
- Brockmole, J. R., Castelhana, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 699–706. <https://doi.org/10.1037/0278-7393.32.4.699>
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. <https://doi.org/10.1006/cogp.1998.0681>
- Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*(4), 360–365. <https://doi.org/10.1111/1467-9280.00168>
- 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*(6), 1996–2009. <https://doi.org/10.3758/s13423-016-1063-0>
- Failing, M., & Theeuwes, J. (2018). Selection history: How reward modulates selectivity of visual attention. *Psychonomic Bulletin & Review*, *25*(2), 514–538. <https://doi.org/10.3758/s13423-0171380-y>
- Gaspelin, N., Leonard, C. J., & Luck, S. J. (2015). Direct evidence for active suppression of salient-but-irrelevant sensory inputs. *Psychological Science*, *26*(11), 1740–1750. <https://doi.org/10.1177/0956797615597913>
- Geng, J. J., & Behrmann, M. (2005). Spatial probability as an attentional cue in visual search. *Perception & Psychophysics*, *67*(7), 1252–1268. <https://doi.org/10.3758/bf03193557>
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*(3), 325–331.
- Gwinn, R., Leber, A. B., & Krajbich, I. (2019). The spillover effects of attentional learning on value-based choice. *Cognition*, *182*, 294–306. <https://doi.org/10.1016/j.cognition.2018.10.012>
- Hardt, O., Einarsson, E. O., & Nader, K. (2010). A bridge over troubled water: reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual Review of Psychology*, *61*, 141–167. <https://doi.org/10.1146/annurev.psych.093008.100455>
- Hickey, C., Chelazzi, L., & Theeuwes, J. (2010). Reward changes salience in human vision via the anterior cingulate. *Journal of Neuroscience*, *30*(33), 11096–11103. <https://doi.org/10.1523/jneurosci.1026-10.2010>
- Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 58–69. <https://doi.org/10.1037/0278-7393.32.1.58>
- Hollingworth, A., & Bahle, B. (2020). Eye tracking in visual search experiments. In S. Pollmann (Ed.), *Neuromethods: Spatial Learning and Attention Guidance* (pp. 23–35). Springer. https://doi.org/10.1007/978-1-4939-9303-0_3
- Jiang, Y. V., Swallow, K. M., Rosenbaum, G. M., & Herzig, C. (2013). Rapid acquisition but slow extinction of an attentional bias in space. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 87–99. <https://doi.org/10.1037/a0027611>
- Kristjansson, A., Wang, D. L., & Nakayama, K. (2002). The role of priming in conjunctive visual search. *Cognition*, *85*(1), 37–52. [https://doi.org/10.1016/S0010-0277\(02\)00074-4](https://doi.org/10.1016/S0010-0277(02)00074-4)
- Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., & Wills, A. J. (2016). Attention and associative learning in humans: An integrative review. *Psychological Bulletin*, *142*(10), 1111–1140. <https://doi.org/10.1037/bul0000064>
- Li, A. S., & Theeuwes, J. (2020). Statistical regularities across trials bias attentional selection. *Journal of Experimental Psychology: Human Perception and Performance*, No Pagination Specified-No Pagination Specified. <https://doi.org/10.1037/xhp0000753>

- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. <https://doi.org/10.3758/s13428-011-0168-7>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- Maxfield, J. T., Stalder, W. D., & Zelinsky, G. J. (2014). Effects of target typicality on categorical search. *Journal of Vision*, *14*(12). <https://doi.org/10.1167/14.12.1>
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. <https://doi.org/10.1037/0033-295X.85.3.207>
- Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 775–799. <https://doi.org/10.1037/0278-7393.27.3.775>
- Mordkoff, J. T. (2019). A simple method for removing bias from a popular measure of standardized effect size: Adjusted partial eta squared. *Advances in Methods and Practices in Psychological Science*, *2*(3), 228–232. <https://doi.org/10.1177/2515245919855053>
- Morey, R. C. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. <https://doi.org/10.20982/tqmp.04.2.p061>
- Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(1), 87–108. <https://doi.org/10.1037/0278-7393.13.1.87>
- Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: looking at things that aren't there anymore. *Cognition*, *76*, 269–295.
- Rosas, J. M., Todd, T. P., & Bouton, M. E. (2013). Context change and associative learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(3), 237–244. <https://doi.org/10.1002/wcs.1225>
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192–233. <https://doi.org/10.1037/0096-3445.104.3.192>
- Sisk, C. A., Remington, R. W., & Jiang, Y. (2019). Mechanisms of contextual cueing: A tutorial review. *Attention, Perception, & Psychophysics*, *81*(8), 2571–2589. <https://doi.org/10.3758/s13414-019-01832-2>
- Stilwell, B. T., Bahle, B., & Vecera, S. P. (2019). Feature-based statistical regularities of distractors modulate attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(3), 419–433. <https://doi.org/10.1037/xhp0000613>
- Talcott, T. N., & Gaspelin, N. (2020). Prior target locations attract overt attention during search. *Cognition*, *201*. <https://doi.org/10.1016/j.cognition.2020.104282>
- 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. C., & Theeuwes, J. (2018). Statistical regularities modulate attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 13–17. <https://doi.org/10.1037/xhp0000472>
- Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, *49*(16), 2095–2103. <https://doi.org/10.1016/j.visres.2009.05.017>
- Yu, C. P., Maxfield, J. T., & Zelinsky, G. J. (2016). Searching for Category-Consistent Features: A Computational Approach to Understanding Visual Category Representation. *Psychological Science*, *27*(6), 870–884. <https://doi.org/10.1177/0956797616640237>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.