AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2021, Vol. 150, No. 12, 2552–2566
https://doi.org/10.1037/xge0001059

Categorical Cuing: Object Categories Structure the Acquisition of Statistical Regularities to Guide Visual Search

Brett Bahle[1], Ariel M. Kershner[2], and Andrew Hollingworth[2]
[1] Department of Psychology, University of California, Davis
[2] Department of Psychological and Brain Sciences, The University of Iowa

Recent statistical regularities have been demonstrated to influence visual search across a wide variety of learning mechanisms and search features. To function in the guidance of real-world search, however, such learning must be contingent on the *context* in which the search occurs and the *object* that is the target of search. The former has been studied extensively under the rubric of *contextual cuing*. Here, we examined, for the first time, *categorical cuing*: The role of object categories in structuring the acquisition of statistical regularities used to guide visual search. After an exposure session in which participants viewed six exemplars with the same general color in each of 40 different real-world categories, they completed a categorical search task, in which they searched for any member of a category based on a label cue. Targets that matched recent within-category regularities were found faster than targets that did not (Experiment 1). Such categorical cuing was also found to span multiple recent colors within a category (Experiment 2). It was observed to influence both the guidance of search to the target object (Experiment 3) and the basic operation of assigning single exemplars to categories (Experiment 4). Finally, the rapid acquisition of category-specific regularities was also quickly modified, with the benefit rapidly decreasing during the search session as participants were exposed equally to the two possible colors in each category. The results demonstrate that object categories organize the acquisition of perceptual regularities and that this learning exerts strong control over the instantiation of the category representation as a template for visual search.

Keywords: visual search, categorical search, statistical learning

How is our attention oriented to behaviorally relevant stimuli in the world? Traditional theories of visual attention held that attention is guided by two dichotomous mechanisms (Egeth & Yantis, 1997). First, attention is attracted to physically salient stimuli and events, such as a uniquely colored item against a relatively uniform background (Theeuwes, 1992) or an object that creates a unique onset transient (Hollingworth et al., 2010; Yantis & Jonides, 1984). Second, attention is guided by observer goals (Desimone & Duncan, 1995; Folk et al., 1992; Wolfe, 1994). It is possible, for example, to strategically limit attention and gaze to those items that match a particular feature value (Williams, 1967; Zelinsky, 1996). In the last 20 years, however, it has become increasingly evident that there are several additional forms of guidance that do not fit conveniently within this structure (Awh et al., 2012), classified as effects of learning and history that influence the guidance of attention in a manner that is largely independent of stimulus salience and observer goals. These include *intertrial effects*, in which recently relevant feature values and locations tend to attract attention and gaze (Kristjansson et al., 2002; Li & Theeuwes, 2020; Talcott & Gaspelin, 2020), *reward learning*, in which stimuli previously associated with reward continue to recruit attention (Anderson et al., 2011; Hickey et al., 2010), *learned distractor rejection*, in which features and locations consistently associated with distraction become less distracting with experience (Gaspelin et al., 2015; Stilwell et al., 2019; Wang & Theeuwes, 2018), and a group of *spatial learning phenomena*, in which learned associations between spatial contexts and target locations reliably guide attention (Chun & Jiang, 1998; Geng & Behrmann, 2005; Jiang et al., 2013).

These effects of learning and history demonstrate that the human visual system is sensitive to recent statistical regularities predicting the properties (appearance, location) that are likely to be associated with task-relevant objects. However, to support real-world visual search, it is not enough to learn the statistical regularities of search targets *in general*, because the guidance of attention to targets is strongly contingent on (at least) two forms of structure. The first is the type and identity of the *context* in which the search occurs; the relevant statistical regularities are those within a particular contextual type (e.g., kitchens) or exemplar (Grandma's kitchen). For example, learning the locations associated with reward in a kitchen does not strongly generalize to rewarded locations on a freeway or in a church. The second form of structural

This article was published Online First April 8, 2021.
Brett Bahle https://orcid.org/0000-0002-8965-7003
This research was supported by a National Institute of Health Grant (R01EY017356) to Andrew Hollingworth.
Correspondence concerning this article should be addressed to Brett Bahle, Department of Psychology, University of California, Davis, 267 Cousteau Place, Davis, CA 95618, United States. Email: brettbahle@gmail.com

organization is the type and identity of the *object* that is the target of search; the relevant statistical regularities are those associated with a particular object type (e.g., cars) or particular object exemplar (e.g., my car). Learning the recent statistics of the appearance of cars does not necessarily predict the appearance of the search target when one is looking for a bucket or for a cat (to choose just two examples).

In the literature on attention guidance by learning and history, there has been considerable work on the structural effects of search context, with research examining context specificity both in the learning of target location (e.g., Brockmole et al., 2006; Chun & Jiang, 1998) and, to a lesser extent, in the learning of target appearance (Anderson, 2015; Chun & Jiang, 1999). In fact, work on context-specific learning of statistical regularities has been broadly collected under the term *contextual cuing* (for a recent review, see Sisk et al., 2019). Moreover, contextual cuing occurs both when the context is defined by consistent spatial information and when the context is defined by consistent identity information (Goujon et al., 2009; Makovski, 2016; but see Makovski, 2018). Thus, it is well established that scene context structures statistical learning for the purpose of visual search.

However, there has been little or no investigation into the structural effects of *target object category* on statistical learning in visual search. Existing work on the statistical learning of visual categories has tended to focus, instead, on the learning of transition probabilities among categories, such as learning that birds tend to follow apples in sequences of images or that forest scenes tend to follow kitchen scenes (Brady & Oliva, 2008; Otsuka et al., 2014; Otsuka et al., 2013). Such work probes category-to-category associations, rather than role of categories in structuring the acquisition of perceptual regularities, and this existing work does not directly apply to visual search processes.

Thus, in the present study, we conducted an initial investigation into *categorical cuing*: the role of object categories in organizing the learning of statistical regularities for the purpose of guiding visual search. Given that this is a broad topic, we made several decisions to focus the scope of investigation. First, we examined how statistical learning is organized by existing category structure rather than examining the formation of new object categories. For categorical cuing to be functional in real-world search, it would need to operate through modification of already well-established category representations, as it is rarely the case that we search for unfamiliar object types. To this end, we investigated how the recent statistical properties of exemplars from a category influence search for any member of that category. Note that the main categorization literature provides little direction in understanding this type of learning. Studies that have investigated well-established categories have tended to do so independently of new learning (e.g., work in the tradition of Rosch, 1975), and studies that have examined category learning have tended to do so for novel categories (e.g., work in the tradition of Medin & Schaffer, 1978). Thus, in addition to probing mechanisms of search guidance, the present study has the potential to inform understanding of a key type of category learning: dynamic modification of existing real-world category representations. Second, we probed the effects of categorical regularities acquired in a task that did not involve visual search (rather than through repeated visual search, as is typical in the visual search literature), because our real-world exposure to object statistics does not always come in the course of visual search. Finally, we focused on the learning of one specific surface feature property, color, rather than other possible object features, such as shape or location, since color is a strong cue controlling the guidance of attention during visual search (Alexander et al., 2019; Beck et al., 2012; Williams, 1967; Zelinsky, 1996).

The structure of the general approach is illustrated in Figure 1. Each experiment began with an exposure session in which participants completed a simple categorization task. They were shown six exemplars from each of 40 categories (20 artifact, 20 natural) for 2 s and categorized each object as "artifact" or "natural." The exemplars from a given category always appeared in the same general color (e.g., each backpack was a novel black exemplar, and each bunch of grapes was a novel red exemplar). After completing the exposure session, participants performed a categorical visual search task (Yang & Zelinsky, 2009) for new exemplars from the same categories. Because their search target on each trial was cued with a category label, such as "backpack," participants were required to retrieve from memory a representation of category appearance (Solomon & Barsalou, 2004) to guide attention. Such retrieval of category information to guide search is known to be sensitive to category structure (Maxfield et al., 2014). We leveraged this property of categorical search to probe whether the guidance of attention is sensitive to the category-specific statistics of recently exposed exemplars. The target object from the cued category either matched the color of recent exemplars from that category (e.g., black backpack, red grapes) or mismatched (e.g., brown backpack, green grapes), and search time provided a measure of the extent to which recent, category-specific statistics biased the instantiation of the search template. Participants performed four blocks of search, searching for each category twice in each block (once in the match and once in the mismatch condition). This allowed us to then examine the potential reduction of the category bias as participants were exposed equally to the two possible colors over the course of multiple searches.

In Experiment 1, we implemented the basic method as illustrated in Figure 1. The primary dependent measure of search efficiency was manual reaction time (RT) as a function of within-category color match. In Experiment 2, we tested participants' ability to learn a more complex distribution of statistical regularities by including multiple colors per category in the exposure phase. In Experiment 3, to ensure that the effect of color match was influencing the guidance of attention to the target object (instead of postselection operations), gaze position was monitored, and the primary dependent measure was elapsed time until fixation of the target object. In Experiment 4, we tested whether a match effect would still arise in a simple categorization task that eliminated the need for visual search. In the second session of Experiment 4, participants first saw the category label, followed by a single exemplar that was or was not a category member. Categorization RT for color-matching and mismatching category members was the primary dependent measure.

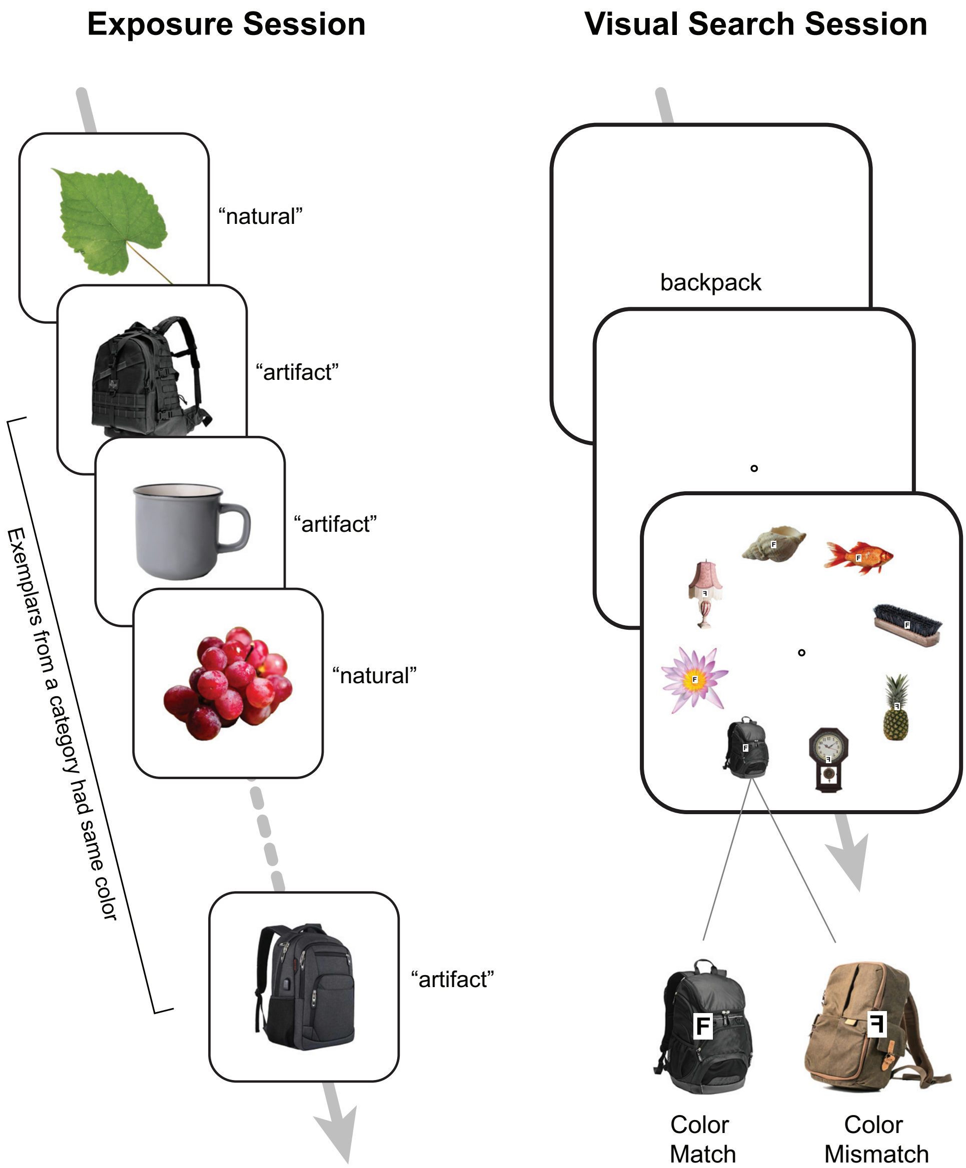
Experiment 1

Method

Participants

Participants in all four experiments were recruited from the University of Iowa community, were between 18 and 30 years of age,

Figure 1
Overview of Method and Design of Experiment 1



Note. Participants first completed an exposure session, in which they viewed six exemplars in each of 40 different categories (20 artifact and 20 natural) for 2 s. They categorized each object as "man made" or "natural." Exemplars from a category had the same general color. Participants then completed a visual search session. On each trial, they first saw a word 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 general color of exemplars from that category presented during the exposure session. Note that the images in this figure are not drawn to scale: objects in the exposure and visual search sessions were displayed at the same size. See the online article for the color version of this figure.

and received course credit for their participation. All participants reported normal or corrected-to-normal vision. Each participant completed only one of the experiments. All human subjects' procedures were approved by the University of Iowa Institutional Review Board.

The present experiments were not based on an existing effect that could support an a priori power analysis. Thus, we used a relatively large sample of 40 in Experiment 1 to ensure sufficient power to detect a medium sized effect. For the type of within-subjects contrast of interest here, a sample of 40 has 80% power to detect an effect of $\eta_p^2 = .18$. Five participants were replaced because accuracy in the search task fell below an a priori criterion of 85% correct. Of the final 40 participants, 27 were female.

Apparatus

Stimuli were presented on an LED monitor (resolution: 1,280 × 960 pixels) with a refresh rate of 100 Hz. Viewing distance was 77 cm, maintained by a forehead rest. Manual responses were collected with a USB button box. The experiment was controlled by E-prime software (Schneider et al., 2002).

Stimuli

There were 40 common categories in the experiment: 20 artifact and 20 natural (see the Appendix A for a complete list of categories). Most categories were defined at the basic level (e.g., "bear"); a few were defined at the subordinate level (e.g., "running shoe"). Within each category, there were 20 exemplar photographs. These were gathered from a variety of sources, including existing object databases and Internet searches. Ten of the exemplars appeared in one general color, and 10 appeared in a different general color. For example, in the "car" category, 10 cars were blue, and 10 were red (see the Appendix A for a complete list of the two colors for each category). The colors for each category were chosen so that there was always a high degree of variability in the particular colors a participant viewed in the exposure session, ensuring that any observed color effects on search were driven by color consistency within a category rather than color consistency across all exposed items.

For each participant, one of the two colors in each of the forty categories was randomly chosen for the exposure session. Six of the exemplars in the chosen color appeared in the six blocks of the exposure session (e.g., a participant would see six different blue cars in the exposure session). The remaining four exemplars in the chosen color appeared in the match condition of the search session, one in each of the four blocks of search. Thus, participants always searched for a novel exemplar; they never searched for an exemplar that they had viewed in the exposure session. Four exemplars from the other color in each category were selected randomly to appear in the mismatch condition in the four blocks of visual search. The sequence of presented exemplars in each category was determined randomly for each participant.

For both phases of the experiment, object stimuli subtended on average $3.08° \times 3.08°$, presented against a white background with a central, black fixation disk. In the exposure session, each trial contained one object image presented centrally. In the search session, the eight objects were presented on a virtual circle around central fixation with a radius of 7.40° visual angle. The location of the first object was selected randomly within a range of 1° to 45°. The remaining objects were each offset by 45° around the virtual circle. Every search display contained one member of the cued category (i.e., all trials were target-present trials). The distractor objects on each trial were chosen from a set of 150 distractor images. Each distractor image came from a different category (75 artifact, 75 natural) that did not overlap with the 40 experimental categories. Each array had four artifacts and four natural objects. For example, if the target was an artifact, there were three artifact distractors (chosen randomly without replacement) and four natural object distractors (also chosen randomly without replacement). The assignment of objects to locations was determined randomly, and thus target location was chosen randomly. A small, black letter "F" on a white background (Arial font, subtending $.25° \times .41°$) was superimposed centrally on each array object, with the orientation of the "F" (facing left or facing right) selected randomly. The cue that appeared before each search array was a word presented in Arial font describing the category of the target object (e.g., "backpack").

Procedure

After arriving for the experiment session, participants were informed that they would participate in two separate experiments. They were given instructions for the first experiment, which was the exposure phase. The trial began with a centrally presented "Press Button" screen that remained visible until the participant pressed a pacing button. After a 400-ms delay, an object image was displayed for 2 s (equating the exposure duration for all stimuli). Participants pressed the left button if the object was from a naturally occurring category or the right button if the object was from an artifact category. Speed was not stressed in the instructions except that the response should be made within the 2-s presentation duration. The participants received "Incorrect!" feedback in red font for 2 s if they made an incorrect categorization or did not respond within 2 s. Participants completed six blocks of 40 trials. Within each block, they saw one exemplar from each of the 40 categories, randomly intermixed. They were required to take a short break between blocks.

After completing the exposure phase, participants received instructions for the second experiment, which was the search phase. Each trial again began with a centrally presented "Press Button" screen. Once the participant pressed a pacing button, there was a 400-ms delay, followed by the category cue label presented centrally for 800 ms. After cue offset, there was a 1,000-ms blank delay before the presentation of the search display, which remained visible until response. Participants searched for the object matching the category label and reported the orientation of the "F" superimposed upon it, using the right button to report a right-facing "F" (i.e., standard) and the left button to report a left-facing "F" (i.e., mirror-reversed). Participants were instructed to make this response as quickly and as accurately as possible. Incorrect responses were followed by "Incorrect!" feedback in red, Arial font for 2 s. Participants completed four blocks of 80 trials. Within each block, they searched for each category twice, once with a matching color exemplar (i.e., an exemplar with the same general color as the exemplars from that category viewed in the exposure session) and once with a mismatching color exemplar (i.e., an exemplar with the color not viewed in the exposure session for that category). Within a block, trials were randomly intermixed. Participants were required to take a short break between blocks.

The entire experiment lasted approximately 40 min. There was a gap of approximately 5 min between the end of the exposure session and the first trial of the search session.

Data Processing

In the exposure session, categorization accuracy did not reliably differ as a function of object type (artifact/natural), with mean accuracy of 98.2% for artifacts and 97.8% for natural objects. Mean search accuracy in the main session was 94.6% correct. There was no reliable effect of match condition, no reliable effect of Block, and no reliable interaction. Incorrect search trials were eliminated from the RT analyses. In addition, for correct trials, RTs more than 2.5 *SD* from the participant's mean in each condition were eliminated. The pattern of results was not influenced by RT trimming in any experiment in this study. A total of 8.3% of trials was eliminated from the RT analyses.

Results and Discussion

The primary analysis concerned the speed of visual search when the target exemplar either matched the color of exemplars in that category from the exposure session (match condition) or did not match that color (mismatch condition). In addition, we examined how the match effect changed over multiple blocks of search, in which participants saw exemplars in both possible colors from a category equally often. Finally, we conducted these analyses separately for the artifact category items and for the natural category items.

The mean RT data were entered into a 2 (match, mismatch) $\times$ 4 (block) repeated-measures ANOVA. The results are reported in Figure 2A. First, there was a main effect of match. Mean search RT was reliably shorter in match condition (1,152 ms) than in the mismatch condition (1,241 ms), $F(1, 39) = 61.3$, $p < .001$, adj $\eta_p^2 = .601$.[1] Second, the main effect of Block was not reliable, $F(3, 117) = .55$, $p = .649$, adj $\eta_p^2 = -.011$. Finally, there was a reliable interaction between these factors, $F(3, 117) = 4.81$, $p = .003$, adj $\eta_p^2 = .087$, indicating a reduction in the match effect as block number increased. Planned contrasts were consistent with this pattern. There was a reliable match effect in Block 1, $F(1, 39) = 45.1$, $p < .001$, adj $\eta_p^2 = .525$; Block 2, $F(1, 39) = 24.0$, $p < .001$, adj $\eta_p^2 = .365$; and Block 3, $F(1, 39) = 15.3$, $p < .001$, adj $\eta_p^2 = .264$; but the effect only reached trend level in Block 4, $F(1, 39) = 3.96$, $p = .054$, adj $\eta_p^2 = .069$. In sum, visual search was strongly influenced by the recent color statistics of natural categories, and this effect then diminished with repeated exposure to both colors within a category.

One possible concern with the color manipulation is that color differences were sometimes correlated with subordinate category differences within the set of natural categories. For example, bears with a black color were drawn from the black bear species, and bears with a brown color were drawn from the brown bear and grizzly bear species. Thus, participants may have instantiated subordinate category representations for some categories rather than encoding the statistics of color directly, and they may have been more likely to populate the search template with a particular subordinate category representation depending on exposure conditions. In addition, subordinate category covariation among features (e.g., additional morphological regularities in the category of grizzly bears, such as head shape or posture) could plausibly have influenced performance for some natural categories. Such concerns do not extend to the artifact categories, however, as we purposefully chose arbitrary color differences for artifacts. To assess the effects of category type, we added artifact/natural as a factor in the ANOVA. Artifact/Natural did not produce a main effect, nor did it interact with the other factors. Moreover, the same pattern of statistical significance as in the main analysis was observed when the analysis was limited to artifact categories, with a reliable effect of match, $F(1, 39) = 36.5$, $p < .001$, adj $\eta_p^2 = .470$, no reliable effect of block, $F(3, 117) = .49$, $p = .687$, adj $\eta_p^2 = -.013$, and a reliable interaction, $F(3, 117) = 2.87$, $p = .039$, adj $\eta_p^2 = .045$. The same pattern was also observed for natural categories, with a reliable effect of match, $F(1, 39) = 45.3$, $p < .001$, adj $\eta_p^2 = .525$, no reliable effect of block, $F(3, 117) = .96$, $p = .417$, adj $\eta_p^2 = -.001$, and a trend-level interaction, $F(3, 117) = 2.66$, $p = .052$, adj $\eta_p^2 = .040$. The results are plotted separately for artifact and natural categories in Figure 2B and 2C.

A second possible concern is that the match effect could have been generated by general color priming from the overall set of items viewed in the exposure session rather than from within-category regularities. This is unlikely, as the exposed colors varied widely and often had opposite values in hue or lightness (see the Appendix A). In addition, the match effect (lower RT in the match than in the mismatch condition) was numerically observed for 38 (18 artifact, 20 natural) of the 40 category items. Thus, we can be confident that the match effect was caused by the learning of *within-category* color regularities, observed broadly across the set of categories.

To summarize the results, participants were initially much faster to respond to the search target when its color matched the color of recently exposed exemplars. Specifically, in Block 1 of search, when participants had recently viewed six exemplars in one color from each category, they were, on average, 134 ms faster to respond to the search target in the match condition than in the mismatch condition. These category-specific biases developed despite exposure to a wide range of individual colors across a large set of 40 categories. Moreover, the biases were rapidly modified by the properties of exemplars in the search session, diminishing substantially over the course of four blocks of search in which both colors in a category appeared equally often.

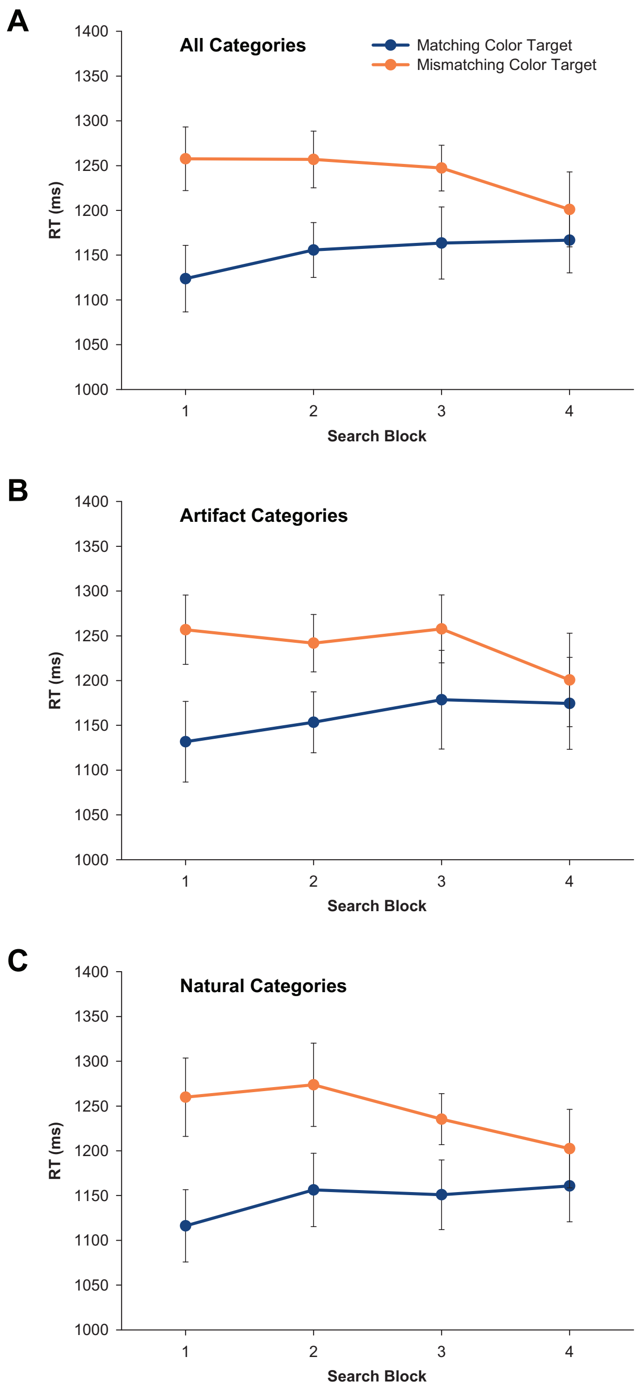
Together, the results suggest that category-specific statistics are acquired simultaneously across a wide range of real-world categories, and that these statistics exert strong control over the specific instantiation of the category representation as a template for visual search. More generally, the results indicate that although real-world category representations may be relatively stable over longer time scales, the functional expression of the category representation at any specific time is likely to be significantly influenced by the statistics of recently viewed exemplars.

Experiment 2

In Experiment 1, exposed exemplars from a category shared a single color. However, real-world statistical regularities will almost always be more complex than the predominance of a single feature value, and consistencies among exemplars will likely change as a function of context, time of day, and so forth. In Experiment 2, we tested whether category-specific learning can span multiple exposed colors in each category, as it would need to in order to support real-world guidance of visual search. The method of Experiment 2 was the same as in Experiment 1, except participants saw exemplars with two different colors in each category during the exposure session (e.g., black backpacks and brown backpacks). In the search session, target objects following the category cue either appeared in one of the exposed colors (match condition) or in a third color (e.g., yellow backpacks) that had not been presented during the exposure session (mismatch condition).

[1] We report adjusted η_p^2, which removes the positive bias inherent in standard η_p^2 (Mordkoff, 2019).

Figure 2
Experiment 1 Results



Note. Panel A displays mean search RT as a function of match condition and search block. Panels B and C plot the same data for artifact and natural category items separately. Error bars are condition-specific, within-subject 95% confidence intervals (Morey, 2008). See the online article for the color version of this figure.

Method

Participants

The observed effect size for the main effect of match in Experiment 1 was adj $\eta_p^2 = .601$, indicating that $N = 7$ would be necessary to achieve 80% power. Given that we expected the match effects to be reduced with multiple exposed colors in each category, we used a substantially larger $N = 24$. One participant was replaced for failing to achieve 85% accuracy. Of the final 24 participants, 17 were female.

Apparatus

The apparatus was the same as in Experiment 1.

Stimuli and Procedure

The stimuli and procedure were the same as in Experiment 1 with the following exceptions. Some changes to the categories were necessary to ensure that three plausible colors were available for each. There were a total of 36 categories: 18 artifact and 18 natural. For each category, there were 27 exemplar images, nine in each of the three colors. A full list of the categories and the three colors associated with each can be found in the Appendix B. For each participant, colors were randomly assigned to conditions using the same method as in Experiment 1.

In the exposure session, participants completed six blocks of 72 trials. Each block presented one exemplar from each of the 36 categories in each of the two colors. In the search phase, participants completed three blocks of 108 trials. In each block, they searched for each category three times, twice in the match condition (with each of the exposed colors appearing once) and once in the mismatch condition. The entire experiment lasted approximately 50 min.

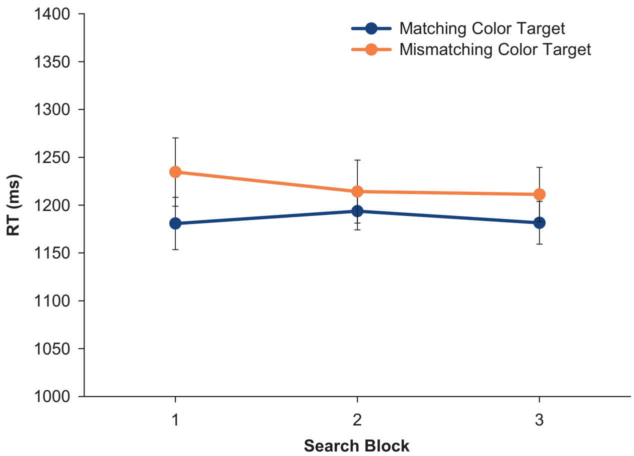
Data Processing

In the exposure session, mean categorization accuracy (artifact/natural) was reliably higher for artifacts (98.7% correct) than for natural objects (98.0%), $t(23) = 2.17$, $p = .040$, adj $\eta_p^2 = .134$, though accuracy was clearly very high for both object types. Mean search accuracy in the main session was 95.9% correct. There was no reliable effect of match condition, no reliable effect of block, and no reliable interaction. Incorrect search trials were eliminated from the RT analyses. In addition, for correct trials, RTs more than 2.5 *SD* from the participant's mean in each condition were eliminated. A total of 6.7% of trials was eliminated from the RT analyses.

Results and Discussion

The mean RT data were entered into a 2 (match, mismatch) × 3 (block) repeated-measures ANOVA. The results are reported in Figure 3. First, there was a main effect of match. Mean search RT was reliably shorter in match condition (1,185 ms) than in the mismatch condition (1,220 ms), $F(1, 23) = 10.2$, $p = .004$, adj $\eta_p^2 = .276$. The was no main effect of block, $F(2, 46) = .27$, $p = .767$, adj $\eta_p^2 = -.031$, nor an interaction between these factors, $F(2, 46) = 1.16$, $p = .32$, adj $\eta_p^2 = .007$. The magnitude of the match effect did not reliably differ for artifacts and natural objects, $F(1, 23) = 4.03$, $p = .057$, adj $\eta_p^2 = .112$, and there was no interaction between

Figure 3
Experiment 2 Results



Note. Mean search RT as a function of match condition and search block. Error bars are condition-specific, within-subject 95% confidence intervals (Morey, 2008). See the online article for the color version of this figure.

artifact/natural and block, $F(2, 46) = .194$, $p = .825$, adj $\eta_p^2 = -.035$. The match effect was observed numerically for 28 (13 artifact, 15 natural) of the 36 category items.

In sum, the match effect was observed despite recent exposure to two different colors within each category. Unlike Experiment 1, there was no evidence that the magnitude of the guidance effect diminished over the course of three blocks of search. This was plausibly caused by the fact that matching-color exemplars appeared twice as often in the search blocks as novel color exemplars and that there were only three blocks of search (instead of four, as in Experiment 1).

Experiment 3

In Experiments 1 and 2, inferences were drawn from manual RT. It can be difficult from end-of-trial measures such as RT to identify when during the search process a manipulation is exerting influence. Recent statistical regularities could have influenced the guidance of attention to the target; they could also have influenced the time required to confirm the identity of the target once it had been attended. To ensure that a substantial proportion of the effect on manual RT was due to guidance processes per se, we replicated the main features of Experiment 1 but monitored participants' gaze position throughout each search trial. If the effect derives, to a significant degree, from guidance processes, we should observe more rapid oculomotor orienting to target objects on match trials compared with mismatch trials.

Method

Participants

As discussed above, the effect size in Experiment 1 indicated that $N = 7$ would be necessary to achieve 80% power. As a conservative approach, we used $N = 16$. Of the 16 participants, 12 were female.

Apparatus

The apparatus was identical to Experiment 1, except participants' right eye position was monitored during the search phase using an SR Research EyeLink 1000 eye tracker sampling at 1000 Hz.

Stimuli and Procedure

The stimuli and procedure were identical to Experiment 1 with the following exceptions. At the beginning of the search phase, participants were instructed that their gaze position would be monitored, and the eye tracker was calibrated. The eye tracker was recalibrated between search blocks and as needed throughout the search session if the participant's gaze position deviated by more than approximately .75° from the central fixation disk. Each trial of search was initiated by the experimenter, who pushed a silent button upon visual confirmation that the participant was maintaining central fixation. The target "F" stimuli, which were the same as in Experiment 1, were sufficiently small that we expected the target object to be fixated before response on the large majority of trials.

Data Processing

Saccades were defined by a combined velocity (30°/s) and acceleration (8000°/s^2) threshold. Fixation position data were analyzed with respect to a region of interest defined around the target object. The region was rectangular and extended approximately .3° beyond the edges of the target objects. An *entry* into the target region was defined as one or more consecutive fixations within that region.

In the exposure session, categorization accuracy did not reliably differ as a function of object type (artifact/natural), with mean accuracy of 98.8% for artifacts and 98.0% for natural objects. Mean search accuracy in the main session was 97.9% correct. There was no reliable effect of match condition, no reliable effect of block, and no reliable interaction. Trials were removed from the analyses of search efficiency if the participant did not fixate the target before response, if the response was incorrect, or, for the remaining trials, if RT was more than 2.5 *SD* from the participant's condition mean. A total of 7.3% of trials was eliminated from the search efficiency analyses.

Results and Discussion

Each trial was parsed into two periods based on the eye tracking record (Malcolm & Henderson, 2009; Zelinsky & Sheinberg, 1997). The first, termed *search time*, was defined as the time from onset of the search array to the beginning of the first fixation on the target region for the entry that immediately preceded the response. The second, termed *decision time*, was defined as the time from the end of the search time period to the manual response. On 4.7% of trials, the participant's gaze entered the target region, exited the target region, and then reentered later, followed by the response. These trials were included in the search time analysis, and the search time measure for these trials was the time from array onset to the beginning of the *last* entry into the target region. However, these 4.7% of trials were removed from the decision times analysis, as it is possible that decision processes spanned these multiple entries. The sum of the two periods on a trial is RT. Search time will generally (though not necessarily exclusively) reflect processes involved in directing attention to the target (Hollingworth & Bahle, 2020). Decision time will generally (though not necessarily exclusively) reflect processes involved in confirming that the fixated object matches the target category, discriminating "F" orientation, and executing the manual response. If recent category-specific statistics influence the formation of the template that guides visual search, then we should observe a match effect on the search time measure.

Search Time

Mean search time results are presented in Figure 4A. The data were entered into a 2 (match, mismatch) × 4 (block) repeated-measures ANOVA. First, there was a main effect of match, $F(1, 15) = 16.8$, $p = .001$, adj $\eta_p^2 = .497$, with shorter mean search time in the match condition (614 ms) than in the mismatch condition (678 ms). There was no effect of block, $F(3, 45) = .51$, $p = .680$, adj $\eta_p^2 = -.032$. Finally, there was a trend-level interaction between these two factors, $F(3, 45) = 2.43$, $p = .078$, adj $\eta_p^2 = .082$, consistent with a reduction in the size of the match effect across blocks. Planned follow-up tests indicated that the color-match effect was statistically reliable in both Block 1, $F(1, 15) = 7.69$, $p = .014$, adj $\eta_p^2 = .295$, and Block 2, $F(1, 15) = 26.1$, $p < .001$, adj $\eta_p^2 = .611$, but not in Block 3, $F(1, 15) = 3.65$, $p = .075$, adj $\eta_p^2 = .142$, or Block 4, $F(1, 15) = .189$, $p = .670$, adj $\eta_p^2 = -.053$. The magnitude of the match effect did not reliably differ for artifacts and natural objects, $F(1, 15) = 1.64$, $p = .220$, adj $\eta_p^2 = .039$, and there was no interaction between artifact/natural and block, $F(3, 45) = 1.20$, $p = .320$, adj $\eta_p^2 = .012$. The match effect was observed numerically for 32 (15 artifact, 17 natural) of the 40 category items.

Thus, after being exposed to exemplars with a consistent color within a category, participants were faster to orient their attention to the target when it matched that color compared with when it did not, and this effect was observed consistently across the 40 categories.

Decision Time

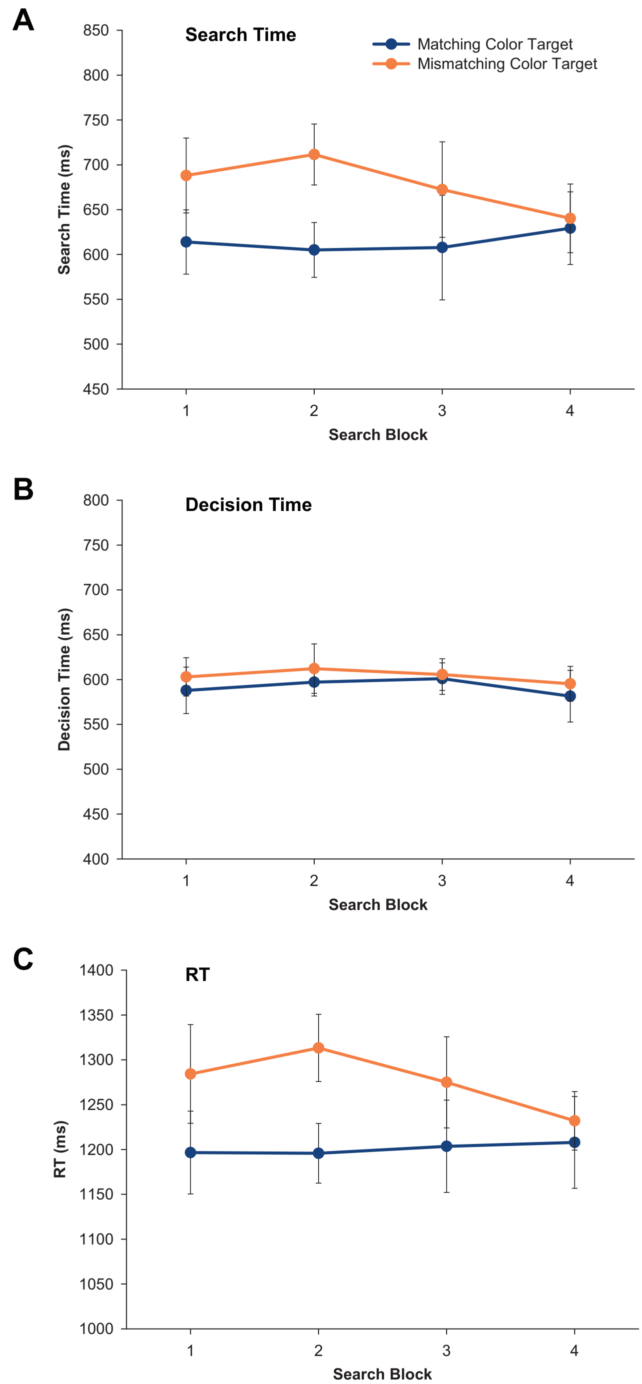
The decision time data are presented in Figure 4B. There was no main effect of match, $F(1, 15) = 2.91$, $p = .109$, adj $\eta_p^2 = .107$, although the numerical trend was toward shorter mean decision time in the match condition (591 ms) than in the mismatch condition (604 ms). There was also no effect of block, $F(3, 45) = .76$, $p = .524$, adj $\eta_p^2 = -.015$, and no interaction, $F(3, 45) = .19$, $p = .905$, adj $\eta_p^2 = -.053$.

Manual RT

To compare the Experiment 3 results directly with those from Experiment 1, we also analyzed mean manual RT (Figure 4C). There was a reliable main effect of match, $F(1, 15) = 22.0$, $p < .001$, adj $\eta_p^2 = .567$, with shorter mean RT in the match condition (1,201 ms) than in the mismatch condition (1,276 ms); there was no effect of block, $F(3, 45) = .71$, $p = .549$, adj $\eta_p^2 = -.018$; and there was a trend level interaction between these two factors, $F(3, 45) = 2.32$, $p = .088$, adj $\eta_p^2 = .076$, consistent with a reduction in the size of the match effect across blocks. Thus, the RT results from Experiment 3 replicated those of Experiment 1.

In sum, when RT was decomposed via the eyetracking record into search time and decision time measures, the bulk of the match

Figure 4
Experiment 3 Results



Note. Panel A displays mean search time (elapsed time until the first fixation of the target preceding the response) as a function of match condition and search block. Panel B displays mean decision time (time from target fixation until response), and Panel C displays mean manual RT. Error bars are condition-specific, within-subject 95% confidence intervals (Morey, 2008). See the online article for the color version of this figure.

effect was observed on search time, suggesting that the primary influence was on the guidance of attention to the target. The results are consistent with our conclusion that recent statistical regularities influence the instantiation of the category-specific template guiding search.

Experiment 4

In Experiment 3, the search time measure primarily reflected the guidance of attention to the target. Such guidance processes were likely to be based both on low-level visual similarity between template and target and on partial or full categorization of the target object before foveation. To isolate categorization processes, in Experiment 4 we eliminated the visual search component. After completing the same exposure phase as in Experiment 1, participants performed a simple categorization task. They were first shown a category label, then a single object exemplar at the center of the screen. The object was either a member of the category (*same* condition) or not (*different* condition). When the object matched the category label, its color either matched or mismatched the general color of the exemplars in the exposure phase. Faster and/or more accurate categorization in the match compared with the mismatch condition would indicate that the basic processes involved in assigning objects to categories were influenced by the statistical properties of recently viewed exemplars.

Method

Participants

As discussed above, the RT data from Experiment 1 suggested that $N = 7$ would be sufficient for 80% power to observe a match effect. However, given that we eliminated the visual search component in Experiment 4 potentially altering the processes contributing to the match effect, we used a substantially larger $N = 24$ participants. Of the 24 participants, 20 were female.

Procedure

The exposure phase was identical to Experiment 1. For the categorization phase, participants first saw a category label (e.g., "backpack") for 800 ms, followed by a 1,000-ms blank period, after which one object was presented at central fixation. On same trials (50% of all trials), the object was a member of the category (always a novel exemplar). On different trials, it was an object drawn randomly from the set of 150 distractors used in the search task of Experiment 1. Participants pressed the left response button if the object was from the cued category and the right response button if it was from a different category. They were instructed to make this response as quickly and as accurately as possible.

There were two categorization blocks of 160 trials each. In each block, participant saw the label for each of the 40 categories four times: twice in the same condition and twice in the different condition. For trials in the same condition, one trial was in the match condition and one was in the mismatch condition. Trials from the different conditions were randomly intermixed.

Results and Discussion

In the exposure session, accuracy on the artifact/natural task did not reliably differ as a function of object type, with mean accuracy of 98.7% for artifacts and 98.2% for natural objects. The key results came from categorization accuracy and RT in the main session, as follows.

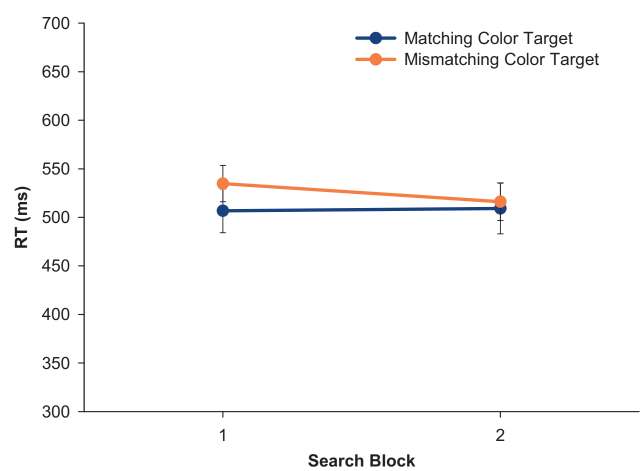
Categorization Accuracy

Overall categorization accuracy was 96.3% correct. Mean accuracy was reliably higher for the Different condition (96.9%) than for the Same condition (95.7%), $F(1, 23) = 4.59$, $p = .043$, adj $\eta_p^2 = .130$. The mean accuracy data from the same condition were entered in a 2 (match, mismatch) × 2 (block) repeated-measures ANOVA. There was no main effect of match, $F(1, 23) = .038$, $p = .846$, adj $\eta_p^2 = -.042$, with 95.5% correct in the match condition and 95.3% correct in the mismatch condition. There was a reliable main effect of block, $F(1, 23) = 10.1$, $p = .004$, adj $\eta_p^2 = .274$, with mean accuracy increasing from Block 1 (94.2%) to Block 2 (96.6%). The two factors did not interact, $F(1, 23) = .182$, $p = .674$, adj $\eta_p^2 = -.035$.

Categorization RT

The RT analysis was limited to correct trials. In addition, trials with RTs more than 2.5 *SD* from the participant's mean in each condition were removed from analysis (1.5% of correct trials). The results are presented in Figure 5. The same condition data were entered into a 2 (match, mismatch) × 2 (block) repeated-measures ANOVA. First, there was a main effect of match, $F(1, 23) = 5.66$, $p = .026$, adj $\eta_p^2 = .162$. There was no effect of block, $F(1, 23) = .512$, $p = .481$, adj $\eta_p^2 = -.021$, but there was a trend-level interaction between these two factors, $F(1, 23) = 4.03$, $p = .057$, adj $\eta_p^2 = .112$. Planned contrasts indicated that the match effect was statistically reliable in Block 1, $t(23) = 3.65$, $p = .001$, adj $\eta_p^2 = .339$, but not in Block 2, $t(23) = .68$, $p = .498$, adj $\eta_p^2 =$

Figure 5
Experiment 4 Results



Note. Mean categorization RT for same-category items as a function of match condition and block. Error bars are condition-specific, within-subject 95% confidence intervals (Morey, 2008). See the online article for the color version of this figure.

−.014. The magnitude of the match effect did not reliably differ for artifacts and natural objects, $F(1, 23) = .771$, $p = .389$, adj $\eta_p^2 = -.010$, and there was no interaction between artifact/natural and block, $F(1, 23) = .738$, $p = .399$, adj $\eta_p^2 = -.011$. The Block 1 match effect was observed numerically for 34 (16 artifact, 18 natural) of the 40 category items.[2]

In sum, participants were initially faster to categorize items that matched the color of exemplars from that category in the exposure session. Thus, recent category-specific statistical regularities influence basic processes involved in assigning exemplars to categories. Similar to the effect on visual search in Experiments 1 and 3, color biases were substantially reduced across blocks of categorization as participants were equally exposed to items of both colors within a category.

General Discussion

The results of the present study demonstrate that the influence of statistical regularities on visual search is structured by the category of the target object. Participants were exposed to a wide range of exemplars of different colors, with color consistency maintained only at the within-category level. When searching for an exemplar of one of those categories based on a category label cue, the guidance of visual search was strongly influenced by within-category color regularities, with substantially faster search for exemplars that matched the color of recently viewed exemplars in that category. This basic categorical cuing effect was observed even when within-category color variation was introduced during the exposure session (Experiment 2). The bulk of the effect during search was on the process of orienting attention and gaze to the target object (Experiment 3). However, even when the task did not require visual search (Experiment 4), regularities among recent exemplars from a category influenced basic categorization efficiency, illustrating that visual search guidance and visual categorization are similarly sensitive to recent statistical regularities.

This work identifies a novel form of structure in the literature on learning and history in visual search. Just as statistical learning is structured by scene context, statistical learning is structured by target object category. Category-structured statistical learning may be a particularly robust form of learning, as demonstrated by the following findings. First, participants simultaneously learned feature regularities across a very large number of structural units (categories). Specifically, they acquired information about the colors of exemplars from 40 different real-world object categories, and in Experiment 2, they learned two colors in each of 36 object categories. This raises the possibility that, under real-world conditions, people track perceptual statistics across all encountered object categories. This is plausible given evidence of extremely large memory capacity for the perceptual details of real-world objects (Brady et al., 2008; Hollingworth, 2004). Second, statistical regularities were acquired relatively efficiently. The basic categorical cuing effect was established after six exposures. It was then substantially reduced across just a few exposures to an even distribution of the two colors during the search blocks. Thus, the “acquisition window” over which recent category-specific statistics influence search may be quite narrow. Finally, category-specific statistical learning generalized across two substantially different tasks. The exposure task involved superordinate-level categorization of single objects; the search task involved visual search through eight-item object arrays and report of a superimposed letter. This type of robust generalization across stimulus and task differences is quite rare in the literature on learning and visual search (and in the literature on statistical learning more generally). For example, contextual cuing and related effects tend to be strongly stimulus and viewpoint specific (Brockmole & Henderson, 2006a; Chua & Chun, 2003), and generalization tends to be limited to structurally similar visual search tasks (Jiang et al., 2015). Together, these properties suggest that categorical cuing is likely to be a robust and pervasive form of structured learning.

The two major structural factors highlighted thus far have been the scene/spatial context in which the search occurs and the type of object that is the target of search. However, there are likely to be additional dimensions that have yet to be (or have been only minimally) studied, such as task and temporal context. In the former case, there is initial evidence that the learning of spatial regularities is quite strongly task specific, with poor generalization even to other, quite similar search tasks (Jiang et al., 2015). Temporal structure has not been studied to our knowledge, but there are likely to be regularities in the properties of real-world search targets that become relevant at, for example, different times of day. In addition, there is a distinction in the type of guidance that can be implemented from observed regularities: the learning of target location regularities can support the spatial guidance of attention, and the learning of surface-feature regularities can support feature-based guidance of attention. The combination of these multiple factors and the type of learning (spatial, featural) creates a potentially elaborate structure for the acquisition of target regularities, much of which has received little or no attention from researchers.

An issue that has arisen in the literature on statistical learning and search guidance is whether or not the effects are based on implicit memory (Chun & Jiang, 1998; Vadillo et al., 2016). This could be applied in two different ways to the present study. First, we can ask whether participants were aware of the color consistencies during the exposure session. If so, we can then ask whether they used this knowledge to guide search strategically in the search session. On the first of these questions, we do not see this as critical for drawing inferences from the present study. In the literature on contextual cuing, guidance effects are observed for repeated target locations in real-world scenes when participants are aware of the contingencies (e.g., Brockmole & Henderson, 2006b), and similar effects are observed for abstract stimulus arrays under conditions where awareness is much more limited (e.g., Chun & Jiang, 1998). It may be of interest to test awareness of regularities in future studies, but here we are concerned with the guidance effect itself. On the second question, we think it is unlikely that participants guided attention strategically based on statistical regularities observed in the exposure session, or, at least, it is unlikely that they persisted in such a strategy. In Experiments 1 and 3, using the standard method, reliable effects of match were observed in both Blocks 2 and 3. Block 2 started after 80 Block 1 trials, in which the search target was equally likely to match or mismatch the exposed color (and Block 3 after 160 trials). Yet, for Experiment 1, 31 of the 40 participants showed a match effect in Block 2 and 34 of 40 in Block 3. In

[2] Categorization RT for “different” trials were not of theoretical interest, since there was no systematic relationship between the cued category and the to-be-categorized object. Mean correct RT was 538 ms following an artifact cue label and 552 ms following a natural object cue label.

Experiment 3, 13 of 16 showed a match effect on search time for Block 2 and 12 of 16 for Block 3. If participants were aware of within-category color distributions, it is unlikely that they would persist in strategic search for exposed colors despite such a large number of trials disconfirming the utility of that strategy. This also reduces the possibility that the effects were driven by local priming based on color repetition within a category.

The present data also highlight the fact that, to understand visual search processes, we must understand how search templates are formed from long-term memory (LTM) representations, and particularly from representations of real-world categories. When research on visual search began, it was primarily a *tool* for understanding fundamental aspects of vision and attention (e.g., what constitutes a feature in vision and how are features integrated?). Visual search is now treated as an important human behavior to be studied in its own right: how do people find the objects they need in complex scenes? In the earlier tradition, the process of forming a search template was not of central interest. Paradigms were designed so that the target of search remained the same across an entire experiment or was visually presented immediately before each search trial commenced (very frequently both). In contrast, real-world search behavior is characterized by frequently changing targets, and there is no helpful experimenter to hold up a picture of, say, a pen, immediately before you search for one. Instead, information to form the search template must be retrieved from LTM, and this will occur dynamically as task goals change and different objects become relevant (Land & Hayhoe, 2001).

The LTM information used to form the template could either concern a specific target exemplar (when only a favorite blue pen will do) or the target category (when any pen will do). Such template representations formed from LTM will never provide a perfect match to the visual properties of the object in the scene (Castelhano et al., 2008; Malcolm & Henderson, 2009; Vickery et al., 2005; Wolfe et al., 2004), particularly as the template representation is likely to draw from multiple encounters with an exemplar (in the case of exemplar-specific search) and from multiple different exemplars in a category (in the case of categorical search). This latter form of search (Yang & Zelinsky, 2009) is extremely common in real-world behavior and also provides a key test bed for examining how long-term knowledge is translated into a visual search template, because extensive research on the structure of categorical knowledge can inform models of template formation, and evidence from categorical search tasks can, in turn, inform our understanding of real-world category representations and categorization mechanisms. Foundational work on categorical search has been conducted by Zelinsky and colleagues, who demonstrated that, despite the necessary imprecision of categorical search templates (relative to the target object as it appears in the scene), attention can be efficiently guided to category members (Yang & Zelinsky, 2009), and this guidance is graded by target typicality (Maxfield et al., 2014), drawing a direct link between template guidance and the known properties of category structure (e.g., Rosch et al., 1976). Moreover, evidence from categorical search has been used to infer the visual differences that are functional in defining real-world categories (Yu et al., 2016).

In the present study we demonstrated that categorical search templates, rather than always reflecting typical values within a category established over extensive experience, can be strongly biased to instantiate visual properties of recently viewed exemplars. In addition, the effects observed in the cued visual search task generalized to a cued visual categorization task (Experiment 4), in which participants saw a category label cue and then responded to indicate whether a photograph of a single exemplar was a member of that category or not. In both cases, participants had time to generate a representation of the category before the appearance of the visual stimuli, creating a template for search guidance, in the search task, or for comparison with the test stimulus, in the simple categorization task. This is in an inductive use of the category, as participants generated a prediction about the future appearance of category members based on a general category cue. Note that this differs from the typical use of category information in the literature on categories and concepts, in which participants first view a stimulus item and then decide in which of several categories it belongs (a stimulus categorization task). In the search task, we can be confident that the inductive use of the category was functional in generating the match effect, because match had a large effect on the guidance of gaze to the target (requiring a predictive template) and a much smaller effect on the confirmation of the target category once the object had been fixated. Thus, statistical regularities among recently viewed exemplars in a category appear to have a substantial influence on predictions about the appearance of future exemplars.

Exemplar effects have been observed in the standard categorization literature, but under substantially different circumstances than implemented here. They are characteristically found in learning paradigms (Allen & Brooks, 1991; Medin & Schaffer, 1978; Regehr & Brooks, 1993; Thibaut & Gelaes, 2006), in which participants categorize a small set of highly controlled, novel stimuli, presented numerous times over the course of the experiment, into a relatively small number of categories. More closely related to the present method, exemplar effects have been observed in the specialized domain of medical diagnosis, in which similarity to previously viewed individual cases exerted a strong effect on the categorization of subsequent cases (Brooks et al., 1991). These studies employed standard stimulus categorization tasks. As discussed above, the present exemplar effects were observed in the inductive use of the category, suggesting that exemplar effects generalize across visual categorization and category-based induction (see discussion in Murphy, 2002). In addition, the exemplar effects observed here were for common, everyday objects belonging to a relatively large number of categories that should have been highly familiar to all participants. Specifically, exemplar effects were found in a domain where they might be least expected: overlearned real-world categories. Although our knowledge and application of real-world categories seems stable, the functional expression of the category may be instead quite variable, depending on the statistics of recently viewed exemplars. In sum, the current results indicate that exemplar effects are likely to be pervasive across a range of categorization tasks, exposure conditions, and category types.

It is important to note that although we found exemplar effects, we do not necessarily interpret these 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, as they could be plausibly accommodated by either approach. In particular, a prototype approach could accommodate the present results by weighting more heavily recently observed features in a summary representation or by adding the assumption that similarity to 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). What is clear is that to account for the present results, both prototype and exemplar theories would need to incorporate strong dependence on recently observed properties and/or exemplars.

Finally, variability in category use based on recently observed properties/exemplars has practical implications for understanding high-stakes visual search tasks, such as those found in radiology and baggage screening. These are often categorical search tasks, in that the observer is monitoring for the presence of any target that belongs to a class (e.g., cancerous lesions, weapons). Biases created by recently viewed exemplars could have a substantial influence on search guidance in these domains. In fact, a phenomenon of this type has been identified in radiology: After detecting a first, benign lesion within an image, sensitivity to additional, cancerous lesions is substantially reduced, a phenomenon termed *satisfaction of search* (e.g., Berbaum et al., 1991). One cause of this phenomenon is a bias in *perceptual set*: Detection of the first lesion biases the search template toward visual and categorical properties of that lesion, causing readers to preferentially attend to similar image regions and to miss dissimilar, cancerous lesions. Indeed, both in radiology (Mello-Thoms, 2006) and in traditional search tasks (Cain et al., 2013), subsequently detected targets have higher similarity to the first target than do subsequently missed targets. The present results raise the possibility that this type of suboptimal search could be understood within the broader context of exemplar effects in the formation of categorical search templates.

Conclusion

Here, we examined how statistical learning of the surface feature properties of real-world objects is organized by object category and the effect of this learning on the instantiation of a categorical template for visual search. The work identifies a novel and pervasive form of structure in the literature on learning and history in visual search. More generally, it suggests that common, real-world categories are surprisingly labile, with the functional implementation of the category strongly dependent on the properties of recently observed exemplars.

References

Alexander, R. G., Nahvi, R. J., & Zelinsky, G. J. (2019). Specifying the precision of guiding features for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(9), 1248–1264. https://doi.org/10.1037/xhp0000668

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., 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

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

Beck, V. M., Hollingworth, A., & Luck, S. J. (2012). Simultaneous control of attention by multiple working memory representations. *Psychological Science*, *23*(8), 887–898. https://doi.org/10.1177/0956797612439068

Berbaum, K. S., Franken, E. A., Dorfman, D. D., Rooholamini, S. A., Coffman, C. E., Cornell, S. H., Cragg, A. H., Galvin, J. R., Honda, H., & Kao, S. C. (1991). Time course of satisfaction of search. *Investigative Radiology*, *26*(7), 640–648. https://doi.org/10.1097/00004424-199107000-00003

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes - Extracting categorical regularities without conscious intent. *Psychological Science*, *19*(7), 678–685. https://doi.org/10.1111/j.1467-9280.2008.02142.x

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

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

Brockmole, J. R., & Henderson, J. M. (2006b). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*(1), 99–108. https://doi.org/10.1080/13506280500165188

Brockmole, J. R., Castelhano, 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

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General*, *120*(3), 278–287. https://doi.org/10.1037/0096-3445.120.3.278

Cain, M. S., Adamo, S. H., & Mitroff, S. R. (2013). A taxonomy of errors in multiple-target visual search. *Visual Cognition*, *21*(7), 899–921. https://doi.org/10.1080/13506285.2013.843627

Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*, *15*(4), 795–801. https://doi.org/10.3758/pbr.15.4.795

Chua, K. P., & Chun, M. M. (2003). Implicit spatial learning is viewpoint-dependent. *Perception & Psychophysics*, *65*(1), 72–80. https://doi.org/10.3758/BF03194784

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

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*, 269–297. https://doi.org/10.1146/annurev.psych.48.1.269

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030–1044. https://doi.org/10.1037/0096-1523.18.4.1030

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

Goujon, A., Didierjean, A., & Marmeche, E. (2009). Semantic contextual cuing and visual attention. *Journal of Experimental Psychology: Human*

Perception and Performance, *35*(1), 50–71. https://doi.org/10.1037/0096-1523.35.1.50

Hickey, C., Chelazzi, L., & Theeuwes, J. (2010). Reward changes salience in human vision via the anterior cingulate. *The Journal of Neuroscience*, *30*(33), 11096–11103. https://doi.org/10.1523/jneurosci.1026-10.2010

Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 519–537. https://doi.org/10.1037/0096-1523.30.3.519

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/10.1007/7657_2019_30

Hollingworth, A., Simons, D. J., & Franconeri, S. L. (2010). New objects do not capture attention without a sensory transient. *Attention, Perception & Psychophysics*, *72*(5), 1298–1310. https://doi.org/10.3758/app.72.5.1298

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

Jiang, Y. V., Swallow, K. M., Won, B. Y., Cistera, J. D., & Rosenbaum, G. M. (2015). Task specificity of attention training: The case of probability cuing. *Attention, Perception & Psychophysics*, *77*(1), 50–66. https://doi.org/10.3758/s13414-014-0747-7

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

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*(25–26), 3559–3565. https://doi.org/10.1016/S0042-6989(01)00102-X

Li, A. S., & Theeuwes, J. (2020). Statistical regularities across trials bias attentional selection. *Journal of Experimental Psychology*, *46*(8), 860–870. https://doi.org/10.1037/xhp0000753

Makovski, T. (2016). What is the context of contextual cueing? *Psychonomic Bulletin & Review*, *23*(6), 1982–1988. https://doi.org/10.3758/s13423-016-1058-x

Makovski, T. (2018). Meaning in learning: Contextual cueing relies on objects' visual features and not on objects’ meaning. *Memory & Cognition*, *46*(1), 58–67. https://doi.org/10.3758/s13421-017-0745-9

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9*(11), 8–13. https://doi.org/10.1167/9.11.8

Maxfield, J. T., Stalder, W. D., & Zelinsky, G. J. (2014). Effects of target typicality on categorical search. *Journal of Vision*, *14*(12), 1. 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

Mello-Thoms, C. (2006). How does the perception of a lesion influence visual search strategy in mammogram reading? *Academic Radiology*, *13*(3), 275–288. https://doi.org/10.1016/j.acra.2005.11.034

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. D. (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

Murphy, G. L. (2002). *The big book of concepts*. MIT Press.

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

Otsuka, S., Nishiyama, M., & Kawaguchi, J. (2014). Constraint on the semantic flexibility in visual statistical learning. *Visual Cognition*, *22*(7), 865–880. https://doi.org/10.1080/13506285.2014.923548

Otsuka, S., Nishiyama, M., Nakahara, F., & Kawaguchi, J. (2013). Visual statistical learning based on the perceptual and semantic information of objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 196–207. https://doi.org/10.1037/a0028645

Regehr, G., & Brooks, L. R. (1993). Perceptual manifestations of an analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General*, *122*(1), 92–114. https://doi.org/10.1037/0096-3445.122.1.92

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

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-x

Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). *E-Prime user’s guide*. Psychology Software Tools, Inc.

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

Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition*, *32*(2), 244–259. https://doi.org/10.3758/bf03196856

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*. Advance online publication. https://doi.org/10.1016/j.cognition.2020.104282

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606. https://doi.org/10.3758/BF03211656

Thibaut, J. P., & Gelaes, S. (2006). Exemplar effects in the context of a categorization rule: Featural and holistic influences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(6), 1403–1415. https://doi.org/10.1037/0278-7393.32.6.1403

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

Vickery, T. J., King, L. W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*(1), 81–92. https://doi.org/10.1167/5.1.8

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

Williams, L. G. (1967). The effects of target specification on objects fixated during visual search. *Acta Psychologica*, *27*, 355–360. https://doi.org/10.1016/0001-6918(67)90080-7

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/bf03200774

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research*, *44*(12), 1411–1426. https://doi.org/10.1016/j.visres.2003.11.024

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

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 601–621. https://doi.org/10.1037/0096-1523.10.5.601

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

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

Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel-serial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(1), 244–262. https://doi.org/10.1037/0096-1523.23.1.244

Appendix A

Complete Stimuli List for Experiments 1, 3, and 4

Category type	Label cue	Color1	Color2
Artifact	Backpack	Black	Light brown
Artifact	Baseball cap	Blue	Tan
Artifact	Bed	Black	White
Artifact	Camera	Blue	Purple
Artifact	Car	Blue	Red
Artifact	Cooking pot	Black	Red
Artifact	Dress	Blue	Yellow
Artifact	Hairbrush	Blue	Red
Artifact	Laptop	Black	Silver
Artifact	Leather chair	Black	Brown
Artifact	Men’s dress shirt	Purple	Yellow
Artifact	MP3 player	Black	Red
Artifact	Mug	Black	Gray
Artifact	Pencil sharpener	Blue	Red
Artifact	Perfume bottle	Green	Purple
Artifact	Running shoe	Black	Blue
Artifact	Stapler	Blue	Green
Artifact	Tricycle	Yellow	Blue
Artifact	T-shirt	Yellow	Red
Artifact	Wrist watch	Black	Gold
Natural	Apple	Light green	Red
Natural	Bear	Black	Brown
Natural	Beetle	Green	Black
Natural	Bell pepper	Green	Red
Natural	Butterfly	Blue	Orange
Natural	Cat	Black	White
Natural	Cherries	Dark purple	Red
Natural	Crab	Blue-brown	Red
Natural	Dog	Black	Brown
Natural	Frog	Brown	Green
Natural	Grapes	Light green	Red
Natural	Horse	Black	Brown
Natural	Leaf	Green	Yellow
Natural	Mushrooms	Brown	White
Natural	Onions	Purple-red	Yellow-tan
Natural	Pear	Red	Light yellow
Natural	Pile of beans	Black	Dark red
Natural	Potato	Light brown	Purple-red
Natural	Rat	Light brown	White
Natural	Snake	Brown	Green

(*Appendices continue*)

Appendix B

Complete Stimuli List for Experiment 2

Category type	Label cue	Color1	Color2	Color3
Artifact	Backpack	Black	Light brown	Yellow
Artifact	Baseball cap	Blue	Tan	Black
Artifact	Bed	Black	White	Light brown
Artifact	Camera	Blue	Purple	Black
Artifact	Car	Blue	Red	White
Artifact	Chair	Black	Brown	White
Artifact	Cooking pot	Black	Red	Silver
Artifact	Dress	Blue	Yellow	Green
Artifact	Dress shirt	Purple	Green	Blue
Artifact	Hairbrush	Blue	Red	Black
Artifact	Laptop	Black	Silver	Red
Artifact	Mug	Black	Gray	Yellow-green
Artifact	Perfume bottle	Green	Purple	Pink
Artifact	Running shoe	Black	Blue	Red-pink
Artifact	Stapler	Blue	Green	Red
Artifact	Tricycle	Yellow	Blue	Red
Artifact	T-shirt	Yellow	Red	Gray
Artifact	Wrist watch	Black	Gold	Silver
Natural	Apple	Light green	Red	Yellow
Natural	Beans	Black	Dark red	Tan
Natural	Bear	Black	Brown	White
Natural	Beetle	Green	Black	Red
Natural	Bell pepper	Green	Red	Yellow
Natural	Bird	Light brown	Red	Blue
Natural	Butterfly	Blue	Orange	White
Natural	Cat	Black	Orange-brown	Gray
Natural	Dog	Black	Brown	White
Natural	Frog	Brown	Green	Red
Natural	Grapes	Light green	Red	Dark purple
Natural	Horse	Black	Brown	White
Natural	Leaf	Green	Yellow	Red
Natural	Mushrooms	Brown	White	Red
Natural	Onions	Purple-red	Yellow-tan	White
Natural	Pear	Red	Light yellow	Light green
Natural	Rabbit	White	Black	Light brown
Natural	Rat	Light brown	White	Black

Received June 22, 2020
Revision received January 14, 2021
Accepted January 23, 2021 ■