



Target templates in low target-distractor discriminability visual search have higher resolution, but the advantage they provide is short-lived

Jonas Sin-Heng Lau¹ · Hal Pashler¹ · Timothy F. Brady¹

Accepted: 16 November 2020

© This is a U.S. government work and its text is not subject to copyright protection in the United States; however, its text may be subject to foreign copyright protection 2021

Abstract

When you search repeatedly for a set of items among very similar distractors, does that make you more efficient in locating the targets? To address this, we had observers search for two categories of targets among the same set of distractors across trials. Visual and conceptual similarity of the stimuli were validated with a multidimensional scaling analysis, and separately using a deep neural network model. After a few blocks of visual search trials, the distractor set was replaced. In three experiments, we manipulated the level of discriminability between the targets and distractors before and after the distractors were replaced. Our results suggest that in the presence of repeated distractors, observers generally become more efficient. However, the difficulty of the search task does impact how efficient people are when the distractor set is replaced. Specifically, when the training is easy, people are more impaired in a difficult transfer test. We attribute this effect to the precision of the target template generated during training. In particular, a coarse target template is created when the target and distractors are easy to discriminate. These coarse target templates do not transfer well in a context with new distractors. This suggests that learning with more distinct targets and distractors can result in lower performance when context changes, but observers recover from this effect quickly (within a block of search trials).

Keywords Visual search · Search template · Target-distractor discriminability

Introduction

Imagine an airport baggage screener has been screening bags for several hours. She is searching for prohibited items in passengers' carry-on bags as they pass through an x-ray machine. She is regularly finding targets, since the scanning system occasionally inserts prohibited items in the digital images of the bags in order to keep the employee vigilant. The task is known to be mentally demanding, and prone to high error rates (Wolfe, Horowitz, & Kenner, 2005). During her shift, the screener performs visual search on thousands of bags – but they all have something in common. In particular, the “distractors” tend to be the same in all cases, as all of the bags mainly contain clothing and common toiletry items. If a

passenger puts their prohibited item in a bag filled not with clothes, but among unusual objects (e.g., stuffed animals), would this make the screener more likely to miss the target? Has she learned a distractor-specific search template for prohibited items, or a general target template that works equally well for all distractors?

To address this, we need to know both how people's visual search performance improves over time when distractors are repeated across trials (e.g., searching among clothing repeatedly), and whether there is a significant cost for switching the distractors (e.g., from clothing to stuffed animals). This question may be more relevant to study with realistic objects because in traditional paradigms where targets and distractors are simple shapes (Treisman & Gelade, 1980) or letters (e.g., Pashler, 1987; Wolfe, Klempe, & Dahlen, 2000), the feature sets of the targets and distractors tend to be small, and thus there are only a small number of relevant features of the target. This limits the possible influence of the distractors on the learned target template. In addition, visual search in simple displays is often much easier than baggage screening, which could mask performance changes in a new context. Thus, in

✉ Timothy F. Brady
timbrady@ucsd.edu

¹ Department of Psychology, University of California, San Diego, California 92093-0109, USA

the current work we seek to examine this question with more complex real-world objects, which allows us to address the question of how specific the target template is for natural images, and how it is tied to a specific set of distractor images.

The role of distractors in search

The issues at stake in the current work have roots in separate lines of research about visual search. In particular, it is well known that both the difference between the targets and distractors (e.g., Avraham, Yeshurun, & Lindenbaum, 2008; Pashler, 1987) and the heterogeneity of the distractors play an important role in visual search (Duncan & Humphreys, 1989). In fact, distractors play a complicated role in search, both because they affect the target templates that are available to participants and because rejecting distractors may be easiest if they are homogenous. For example, replacing a set of distractors with new distractors that are more easily distinguished from the target can degrade search performance if it is done in a way that makes the distractors more heterogeneous (Rosenholtz, 2001). In general, search performance is best when people know in advance the exact features that define the distractors and the targets (Eckstein, 2011; Schoonveld, Shimozaki, & Eckstein, 2007). In fact, in some situations it has even been shown that participants base their target templates not on the optimal way of detecting the target but on the optimal way of distinguishing the target from the distractors (e.g., Geng, DiQuattro, & Helm, 2017; Navalpakkam & Itti, 2007; Scolari & Serences, 2009; Yu & Geng, 2019).

Learning distractors

In addition to the influence of distractors during a given visual search, our question also depends critically on the extent to which people extract and maintain information about distractors. This question has been the focus of several studies. For example, Wolfe, Klempe, and Dahlen (Experiment 6, 2000) studied the effect of repeated distractors in visual search. In their experiment, participants searched for a target letter among three or five letter distractors. In the repeated condition, the locations and identities of the stimuli were the same on every trial. Participants searched for a different target on each trial, which was indicated to them at the beginning of the trial. Wolfe et al. (2000) found that participants became significantly faster in this condition than in a condition where the identities of target and all distractors changed in every trial. However, even in the repeated condition, participants never became as good as if they simply memorized the display and searched their memory. This suggests that people do make use of distractor information at least under some circumstances, but that even when the distractors are sometimes targets, people

do not form perfect distractor memories or optimally take into account the ways in which search stays the same on every trial.

In another line of research, Won and Geng (2018) asked whether people created a distractor template when searching for a predefined target. They had participants search for a gray patch among distractor patches with a restricted range of colors (e.g., green and blue). After a few blocks of search trials with the same target and distractors, the distractors were switched to some other colors within the same range (e.g., slightly different shades of green and blue) or some outside the range of previously presented colors (e.g., yellow and orange). Participants were faster to identify the target when distractors were within the same range, compared to when they were out of range. The authors claim that people retain information about distractor properties, and use that to aid their visual search. Along the same line, Cheverikov and colleagues (e.g., Chetverikov, Campana, & Kristjansson, 2016, 2017; Chetverikov, Hansmann-Roth, Tanrikulu, & Kristjansson, 2019) have shown that people can learn about the shape of distractor distribution, such as Gaussian or uniform distributions, after a few trials searching for an odd-ball target. Specifically, they had participants search for the most visually salient item among a group of relatively more homogeneous distractors. Targets were not known in advance of each trial. They could be a tilted line among less tilted distractors, or a purple target among green and blue-green distractors, etc. In each learning streak of five to seven trials, participants searched for the visually similar target against the same distractors. Critically, the values of distractors were drawn from a Gaussian or a uniform distribution. In a test trial immediately following the learning streak, a new target was drawn within the range of the previous distractor distribution. The response times in the test trials were found to closely follow the distractor distributions of the learning streak. These experiments suggest that observers learned about the properties of the distractors during visual search.

This question, in the broadest sense, has also been studied under the domain of contextual cueing (Chun & Jiang, 1998, 1999). In a typical contextual cueing paradigm, participants search for a simple pre-defined target (e.g., a T) among visually similar distractors (Ls). Without explicitly communicating to the participants, the locations of the distractors in some trials are predictive of the location of the target. With repeated presentations of the pairing during the training, participants become faster at locating the targets when they are at a learned position relative to the distractors compared to when the target is in a new location. Thus, even with very simple stimuli, information about the distractors is not completely lost. In fact, which shapes tend to be the distractors for a given target is also learned in a contextual cueing-like setting (Chun & Jiang, 1999). Thus, people can associate not only spatial configurations of distractors with the location of targets, but even particular sets of distractors with particular targets.

With real-world objects, it is also known that participants tend to perform reasonably well at incidentally remembering

items that have been presented in search displays, either as targets or as distractors – particularly distractors that closely resemble the targets (e.g., Williams & Henderson, 2005). The effects of repeated distractors on search performance have also been studied using eye tracking (Yang, Chen, & Zelinsky, 2009). Although eye movements are not necessary for most search effects – which also arise in covert search – they can often be a somewhat useful index of how people are moving their attention (Zelinsky, 2008). The researchers had participants search for targets in the same category, while repeating some of the distractors across trials. In the experiment, first fixations were slightly less likely to fall on an old distractor, compared to distractors that had not been repeated as frequently or new distractors. However, this study was not designed to detect the influence of context change: in the study, there was a high variability among the distractors, and the discriminability between the target and distractors was not manipulated.

Categorical target templates

When people search for a target among distractors, a representation of the target is held in working memory to facilitate the task (Bravo & Farid, 2009; Vickery et al., 2005; Wolfe et al., 2004). This target representation can be quite flexible and is able to survive rotation (Reeder & Peelen, 2013). It is variously referred to as the “target template,” “attentional template,” or “search template.”

A number of studies have investigated how the properties of the target templates emerge, and how they are affected by the task. For example, target templates can vary in their specificity or detail. It is perhaps not surprising that more detailed target templates can guide visual search more efficiently (e.g., Malcolm & Henderson, 2010). Wolfe, Horowitz, Kenner, Hyle, and Vasan (2004) suggested that the target template-generation process could be dynamic. When participants are given a text cue prior to a search task, the template generated would be coarse. When a picture cue is used instead, the template becomes more detailed, which can improve search performance.

For example, Bravo and Farid (2009) had participants search for a fish target among coral reef distractors. Each search trial was preceded by a cue. The cue was exactly the same as the search target in some trials. In other trials, the cue could be the same image that was rotated, or a fish image that belonged to the same species. The cue could also be an uninformative word label (i.e., “fish,” when all the targets are always fish). Response times to locate the target decreased as the amount of information contained in the cue increased. The results agree with that of Maxfield, Stalder, and Zelinsky (2014). In the latter study, the researchers presented a text cue to the participants, and had them perform a visual search for categorical objects. Participants were slower when the

targets were low typicality given the text cue. In those situations, participants’ first fixations were also less likely to fall on the target. Hout and Goldinger (2015) found the same pattern in response times with targets of decreased precision. Using eye tracking, they found that participants spent more time scanning the stimulus array and making a decision once fixation fell on the target when the target template was of a low precision.

Highly specific target templates are not always ecologically effective in real-world situations, however. If a target template is general, it can be applicable to many related tasks once it is generated. On the other hand, if the target template is very specific, and it depends entirely on the current context, the target template may become less useful with slight changes to the viewpoint of the target or the search task. In a naturalistic environment, targets are never visually stable. When we are performing a visual search, such as looking for our car key on a cluttered desk, a slight change in perspective or time could make the target look very different (Zelinsky, 2008). Therefore, from daily-life experience, it is likely that the visual system tends to create target templates that are more general rather than highly specific. For example, Bravo and Farid (2012) examined how a target template was generated when participants were asked to search for a fish target among distractors under two conditions. In the first condition, an exact target image was used across trials. In the other condition, fish images of the same species were used as targets. When an exact image was repeated over trials in the training phase, participants generated a very specific target template. During the test when participants were asked to search for different fish targets across trials, their response times were slower than the other group who saw different images of the same species within a block. The latter condition is likely to be more representative of how people search for objects in a real-world scenario. Nevertheless, it echoes the idea that target-template generation is dynamic and task specific. In an experiment that employed simple stimuli (colored rings), Goldstein and Beck (2018) showed that varying search templates across trials lengthened dwell time on distractors, and the process of target verification. It also lengthened the time that participants took to establish the target prior to initiating the search.

As noted previously, some work has also provided evidence that target templates might vary not only in their specificity, but that for cases of fixed distractors, people might build a target template that is not only detailed but also distractor-specific. That is, target templates can be based not on the optimal way of detecting the target but on the optimal way of distinguishing the target from the distractors (e.g., Geng, DiQuattro, & Helm, 2017; Navalpakkam & Itti, 2007; Scolari & Serences, 2009; Yu & Geng, 2019).

Thus, overall, there is ample evidence that using cues before trials can change a person’s target template to be more specific or more general, and that target-template specificity

plays an important role in the speed and generalizability of visual search. There is also evidence from simple stimuli that people might form distractor-specific target templates. However, the role of learning about distractors in shaping the specificity of the target template in a real-world categorical search has not been investigated. In addition, it is unknown whether people form not only detailed but also distractor-specific templates in such scenarios.

Current investigation

In the current study, we had participants perform a demanding visual search task to examine the nature of the target template people form and the extent to which this template becomes distractor-specific. Participants were asked to look for two categories of targets (lanterns and binoculars) among distractors. The same set of distractors was used for a given participant throughout the first half of the experiment. This provided a chance for the participants to learn about the visual features of the distractors, and to adjust their target templates over the course of training. Some groups were trained on distractors that required detailed search templates (e.g., visually similar to the targets, also man-made artifacts; see Fig. 1), and others were trained on distractors that did not require detailed search templates to distinguish from the targets (i.e., not visually similar to the targets, mammals, or plants). Once the participants had this extended experience with the distractors, we replaced them with a new set of distractors. In Experiment 1, the new set of distractors required a detailed search template, as they were quite similar to the targets. In Experiment 2, the new set of distractors did not require a detailed template, as they were dissimilar to the targets. In Experiment 3, the new distractors were from a different category than the learned distractors, but similar visually and semantically to the target, to examine whether the learned template is specific to the particular distractors or only more versus less detailed.

The goal of this set of experiments is twofold. The first half of each experiment allows us to examine whether people learn about the distractors during a visual search task. While encouraging high accuracy in the search task, we could examine how a target template develops. If observers learned to adjust their target templates with respect to the distractors, we would see a decreasing response time over blocks that reflects this learning. Importantly, since target-distractor discriminability was manipulated across groups, we expect the rate of improvement to differ across groups if it arises from template learning rather than solely from overall improvements at the search task. The second half of each experiment – the transfer task – allows us to examine how well the knowledge learned in the training phase can be transferred to a new context. If observers created distractor-specific target templates, we should expect a sizable drop in performance once a new set of distractors was introduced. Otherwise, we should expect stable performance before and after the switch.

Stimulus validation

Part 1: Multi-dimensional scaling with human observers

In our studies, we aimed to enquire about the extent to which people create a search template that takes into account the context. In other words, the extent to which people adjust the detail stored in the search template according to the distractors they are searching amongst. In particular, we seek to assess whether participants tend to generate a more detailed target over time when the targets are hard to discriminate from the distractors. As targets, we use lanterns and binoculars (Fig. 1). As distractors, we use mammals, plants, artifacts, and sea animals. Thus, this experiment sought to understand the similarity relationship between these categories, and in particular between these categories and the lanterns and binoculars per se. To ensure the distractors in each condition were visually similar compared to the targets (or not similar, as appropriate), we performed a similarity experiment and then a multi-dimensional scaling (MDS) analysis on the distractor images (Hout, Godwin, Fitzsimmons, et al., 2016; Hout, Papesh, & Goldinger, 2013) to assess their similarity. We expected artifacts to be most similar to the search targets. We also sought to validate that our stimulus selection resulted in a variety of visual features within each category.

Methods

Design

There were a total of 18 categories, such as horse, fish, flower, and cell phone. On each trial, participants were shown two images, one from each of two categories. They were asked to judge how visually and semantically similar the pair of images was. To manage the length of the experiment, only one image of each category was shown to a participant. Each image was drawn from a set of multiple images within the category, and each participant saw a different set of images (Fig. 1).

Participants

Participants were recruited online using Amazon's Mechanical Turk (MTurk). Fifty participants were recruited (15 female, 31 male, and four did not disclose their gender). This group of participants acts as independent raters for the images used in Experiments 1 and 2. None of them participated in the visual search experiments. They reported an age between 21 and 62 (mean = 36.4, SD = 11.3) years. The research protocol was approved by the University of California, San Diego's Internal Review Board for research participant protection. All participants provided informed consent before taking part in the experiment, and were compensated for their time.

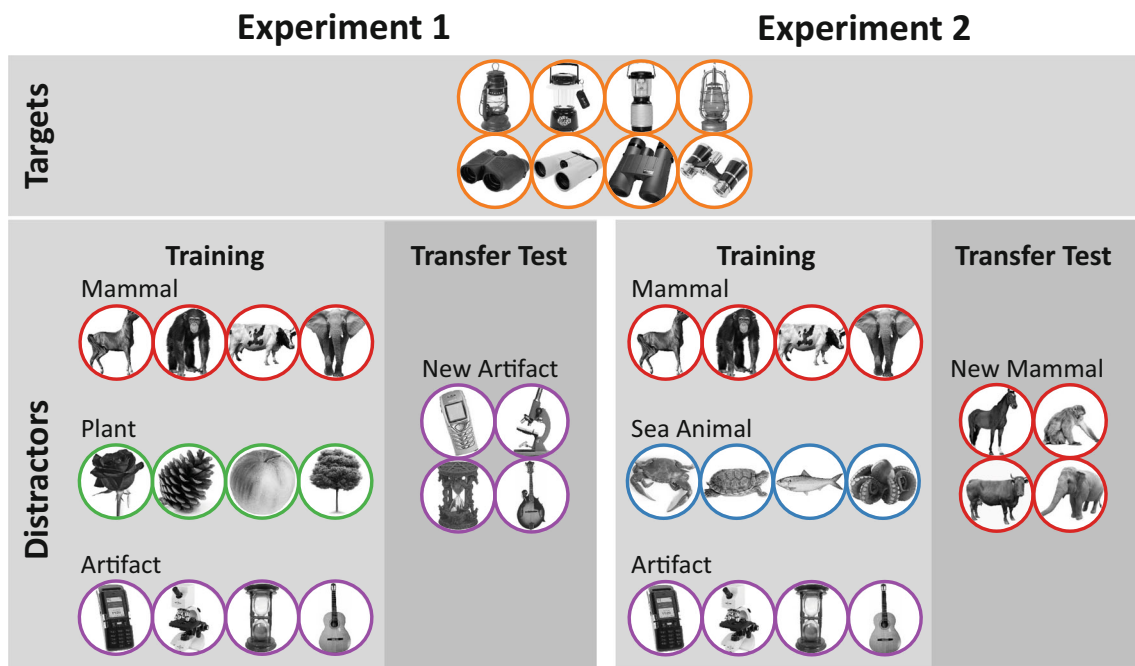


Fig. 1 Structure of Experiment 1 and Experiment 2. Participants were asked to search for any lantern or any pair of binoculars in each trial and indicate whether they had found a lantern or a pair of binoculars. In Experiment 1, across participants we manipulated whether participants searched for the targets among either mammal or plant distractors, which were visually distinct from the targets, or artifact distractors, which were visually more similar to the targets. Once they had sufficient experience with the distractors, a new set of artifact distractors that were visually similar to the targets were introduced, replacing the ones they were

trained on. We intended to assess how well their learned target template transferred to this new search task. The structure of Experiment 2 was identical to that of Experiment 1, except that participants searched for the targets among distractors that were either visually dissimilar to the target mammal or sea animal distractors, or visually similar to the target artifact distractors during training, and among new visually dissimilar to the target mammal distractors during the transfer test. Such an easier transfer test was designed to ask how specific the target template was to a particular set of distractors

Apparatus and stimuli

The experiment was written in JavaScript. The images used in the subsequent visual search experiments were included. Two images of different categories, measuring 100×100 pixels, were shown on each trial.

Procedure

After participants gave their consent for their participation, they were shown the instructions of the experiment. In each trial, two objects in grayscale were shown. The images were set to 70% transparent to match those used in Experiments 1 and 2.

On each trial, a question was shown together with a slide bar, asking participants how visually similar the two images were (Fig. 2). In the instruction, we emphasized that they should judge the pair's visual similarity without taking into account how the two images were conceptually related. The default value was set at the mid-point of the slide bar. Once the participants adjusted the value on the slide bar, a new question showed up with another slide bar. The delay in showing the second question was to encourage participants to perform the task without rushing through the experiment. The second

question asked them to judge the conceptual similarity between the two images. Once they adjusted the value on the second slide bar, a submit button showed up. Clicking on the button led to a new trial, where another pair of images was shown together with visual judgment question and a slide bar with the value reset to the mid-point.

Participants went through five practice trials before the main experiment of 153 trials. Images used in the practice trials came from categories that were not utilized in the main experiments. The order of the practice trials was fixed but that of the actual experiment was randomized. In each trial, the order of the images was randomly determined. The whole experiment took around 20 min to complete.

Results

Each participant gave 153 visual similarity ratings, and another 153 conceptual similar ratings, on the pairs of images. These two sets of ratings were separately analyzed, and also used to form two MDS plots. The data were analyzed with the smacof package in R (De Leeuw & Mair, 2009) to account for individual differences in the subjective ratings. While looking at the scree plots, there were no clear "elbows" that could guide our model selection (Hout et al., 2016). However, our

Fig. 2 Methods of stimulus validation. Two images were first shown on the screen with a question about visual similarity and a slide bar. After the visual similarity judgment was made, another question about conceptual

similarity was shown. Participants were allowed to submit their answers only after adjusting the values on both slide bars

goal was not to fully understand the similarity space, but to validate that the categories were related as we expected. Therefore, we restricted the results to two dimensions to aid visualization. The stresses of the resulting models are 0.365 and 0.326 for the visual similarity and conceptual similarity data. Figure 3 shows the resulting MDS plots for visual (Fig. 3a) and conceptual (Fig. 3b) similarity.

Visual similarity

The artifact stimuli were judged visually most similar to two search targets (average distance = 0.30). With the other categories, plant stimuli were judged to be visually more similar to the targets (0.73) compared to mammal stimuli (0.99) and sea animal stimuli (1.15). For the MDS plot, the categories were color coded for visualization. Stimuli that are closer on the MDS space were judged to be more similar. It can be seen that images of the same category were judged to be more similar to each other, as they formed small clusters. It is also worth noting that all categories are linearly separable from each other.

Conceptual similarity

A similar pattern emerged for conceptual similarity. The artifact stimuli were judged to be conceptually most similar to the search targets (average distance = 0.47). With the other categories, plant stimuli were judged to be semantically more similar to the search targets (0.88) compared to sea animal stimuli (1.13) and mammal stimuli (1.25). Visualizing them

with MDS, images of the same category were judged to be more similar to each other, and they formed tighter clusters compared to the groupings in visual similarity. All stimulus categories were linearly separable from each other.

Discussion

In terms of visual and conceptual similarity, artifacts are most similar to search targets both visually and semantically. In addition, the MDS analysis shows all categories are linearly separable from each other. The stimuli form stronger conceptual groupings compared to visual groupings, as can be seen from the tighter clusters in the conceptual similarity plot. This validated our plan for using the artifacts as more difficult distractors in the visual search experiments. In particular, in the visual search experiments, we expect the search task to be more difficult when observers search for the targets, lanterns and binoculars, among artifact distractors, compared to other types of distractors.

Part 2: Measuring object similarity with a deep convolutional neural network

Due to the large image set involved in the experiments, it was not practical to have human observers to judge the similarity of each pair of images, so we subsampled to get estimates for each category rather than per image, as this is most relevant to our search experiments. However, the heterogeneity of individual images is also informative. Thus, to estimate the pairwise image similarity, we made use of a pre-trained deep

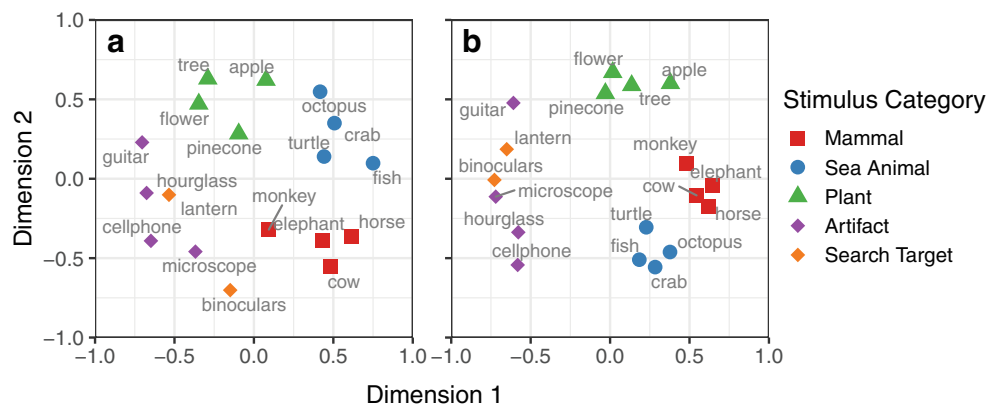


Fig. 3 Multi-dimensional scaling analysis for (a) visual and (b) conceptual similarity. Points closer in space denote stimuli that were judged to be more similar. In general, stimuli fell into clusters that we

defined for the conditions in Experiments 1 and 2. The search targets were more visually and conceptually similar to the artifact stimuli than those in other categories

convolutional neural network (CNN). CNNs are useful metrics of object similarity for predicting memory (Brady & Störmer, 2020), and have consistently been shown to provide some level of match to the human visual system (Yamins et al., 2014). As such models are trained to perform categorization, the high-level features in them contain both semantic and visual information, but are primarily visual in the sense that the network has no broader conceptual knowledge beyond categorical classification. Here we used one particular network, ResNet-18 (ResNet18; He, Zhang, Ren, & Sun, 2016) to extract object features for each image (via the PyTorch instantiation of ResNet-18, pretrained on ImageNet), focusing on the top max-pool layer, the final layer before the classification aspect of the network. For each image utilized in Experiments 1 and 2, ResNet18 produces a feature vector with 512 feature dimensions. We then calculated the correlations between each pair of image feature vectors to form a correlation matrix of all images.

Results and discussion

Figure 4 shows the correlation matrix of the stimuli in Experiments 1 and 2. The tiles with darker shades indicate stronger similarities between a pair of images. The set of small triangles along the diagonal indicates that images within a category (e.g., cellphone) are found to be highly similar to each other. Images that belong to the same superordinate category (e.g., sea animals) are also reasonably similar to each other. The same is true between the search targets and images that belong to the artifact category. The high correlation between the two categories is indicated by the rectangular patch at the lower left-hand corner. Images across superordinate categories are found to be less similar to each other, indicated by the lighter shades, except for a minority of pairs (e.g., apple and guitar).

Overall, the similarity experiments and CNN data validate our selection of artifacts as most similar to our search targets;

suggest our categories are relatively coherent semantically while also featuring more diversity in visual features; and show there are no outliers amongst our objects in terms of visual similarity.

Experiment 1

Experiment 1 was the first visual search experiment. Participants were asked to search for any lantern or any pair of binoculars in each trial and indicate whether they had found a lantern or a pair of binoculars. Across-participants we manipulated whether participants searched for the targets among either mammal or plant distractors, which were visually distinct from the targets, or artifact distractors, which were visually more similar to the targets. Once they had sufficient experience with the distractors, a new set of artifact distractors that were visually similar to the targets were introduced, replacing the ones they were trained on. We intended to assess how well their learned target template transferred to this new search task.

Methods

Design

The experiment had two phases, a training phase and a transfer-test phase. The task for the participants was to search for any lantern or any pair of binoculars among some distractors. In the training phase, participants were randomly assigned into one of the three conditions, in which they encountered a specific set of distractors. One group searched for the targets among pictures of mammals, another group searched among plants, and a third group searched among other artifacts as distractors, respectively. The artifact distractors were both semantically and visually more similar to the targets, creating a more difficult search task for this

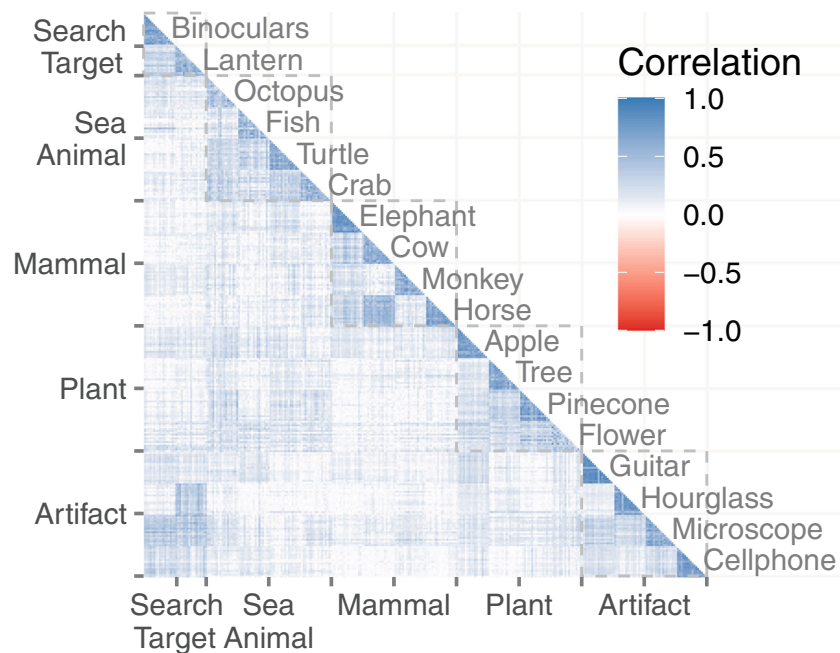


Fig. 4 Correlation matrix with all pairwise similarity measures. Each image was processed with ResNet-18 to extract 512 high-level features. The feature vectors were then correlated with one another. A dark shade of blue denotes high similarity for a pair of images. Images of the same

category (e.g., cellphone) are highly similar to each other. Images of the same superordinate category (e.g., sea animal) are reasonably similar to each other compared to images across superordinate categories. Critically for our purposes, artifacts are by far the most similar to the search targets

group, and an easier target-distractor discrimination task for the mammal and plant distractor groups. The same set of distractors was repeated in each trial for a particular participant.

Accuracy was emphasized during the experiment and therefore was expected to be near ceiling, and so behavioral changes in reaction times across blocks were our main assessment of the learning efficiency during visual search. Changes across conditions indicate the degree of learning for different distractor sets.

In the transfer-test phase, participants in all three conditions searched for the defined targets among a *new* set of artifact distractors. Hence, all participants experienced the same set of visual stimuli as distractors. Behavioral differences between conditions can be attributed to the training participants received. In the “artifact” training condition, the new distractors were distinct exemplars of artifacts but members of this same superordinate category. In the two other conditions, the distractors changed to an entirely new superordinate category (e.g., from plants to artifacts).

Participants

Sixty-nine participants (42 female) were recruited from University of California, San Diego's Psychology Subject Pool. They had a mean age of 20.8 years at the time of participation. All participants gave informed consent prior to taking part in the study. All participants took part in the study for partial course credit.

We were interested in the differences in reaction times between groups, and expected a large effect size for the comparison of the groups (Cohen's $d \approx 0.8$). To achieve a power of 75% for a pairwise t-test, at least 23 participants were required per group. The sample size obtained was based upon this power analysis, which was conducted prior to participant recruitment. Data from three participants were removed with respect to their overall accuracy, with criteria detailed in the *Results* section.

Apparatus and materials

Participants sat in a sound-attenuated room with normal indoor lighting.

The experiment was written with custom JAVA code. Stimuli were shown on a Dell E173FpC 17-in. LCD monitor with 4:3 aspect ratio. At a viewing distance of approximately 60 cm, the monitor's visible area was 31.2° wide and 25.4° tall in visual angle. The resolution of the screen was set to 1,024 × 768. The visible region of the screen was cropped, such that only the central rectangular area of the screen was utilized to show the visual stimuli. The rectangular window measured 970 pixels wide and 680 pixels tall, which translated to 29.63° wide and 21.00° tall in visual angle.

Images obtained online and from a publicly available image database (Brady, Konkle, Alvarez, & Oliva, 2008) were turned into grayscale and resized to 80 × 80 pixels (2.50 visual degrees). Sixteen slots on an 8 × 5 imaginary grid were randomly selected to host the images on each trial. The images

were jittered with a maximum of 15 pixels left or right, 23 pixels up or down, and 30° clockwise or anticlockwise, to avoid appearing on a static grid.

Three labels were shown below the rectangular image window corresponding to the three response options on each trial: “binoculars,” “none,” and “lantern.” Positions of the three labels corresponded to the three arrow buttons, “left,” “down,” and “right,” that participants used for responses. Figure 5 shows an example of the search display.

Procedure

In each experimental session, the participant was first shown a page of instructions on the screen. Afterwards, a multiple-choice quiz was administered. Participants had to answer all the questions correctly to move on, otherwise he or she would be shown the instructions again. Participants were encouraged to prioritize accuracy over speed during the visual search task as response time was our main measure of interest.

The experiment started with nine trials in the practice block. There were three trials of each response type: binoculars, lantern, or target-absent. The order of the trials was randomized.

All trials started with a fixation cross for 200 ms. In a target-absent trial, four items from each of the four distractor sub-categories were shown on the screen (e.g., monkey, horse, elephant, and cow within the mammal category). Distractor

categories that a participant encountered depended on the condition that he or she was assigned to. One-third of all participants saw mammal images as distractors, another one-third had plants as distractors, and the remaining group had artifacts as distractors. These conditions are hereafter referred to as *mammal-training*, *plant training*, and *artifact-training conditions*. The same set of distractors was repeated in each trial for a given participant. On target-present trials, a pair of binoculars or a lantern image replaced one of the distractors, leaving 16 objects in total present. The images stayed on the screen until the participant responded. On each trial, participants had to indicate whether a target was present and what kind of target it was via a three-alternative forced-choice (3AFC) task, with the options “binoculars,” “lantern,” or “target-absent.” Once a response was detected, audio feedback was also delivered. The audio feedback indicated whether the participant made a correct or incorrect decision on the trial. If an incorrect response was made, a hollow circle showed the location of the target for 3 s. This feedback screen served as a delay in the task and indirectly encouraged high accuracy throughout the experiment. The next trial appeared after an 800-ms blank screen following feedback or immediately following the trial if there was no error. After the nine-trial practice block was over, the participant completed five additional training blocks. Each block contained 36 search trials. There was a 30-s break in between the blocks.

Once the participant completed the five training blocks, a new set of distractors was drawn from the image set. All the

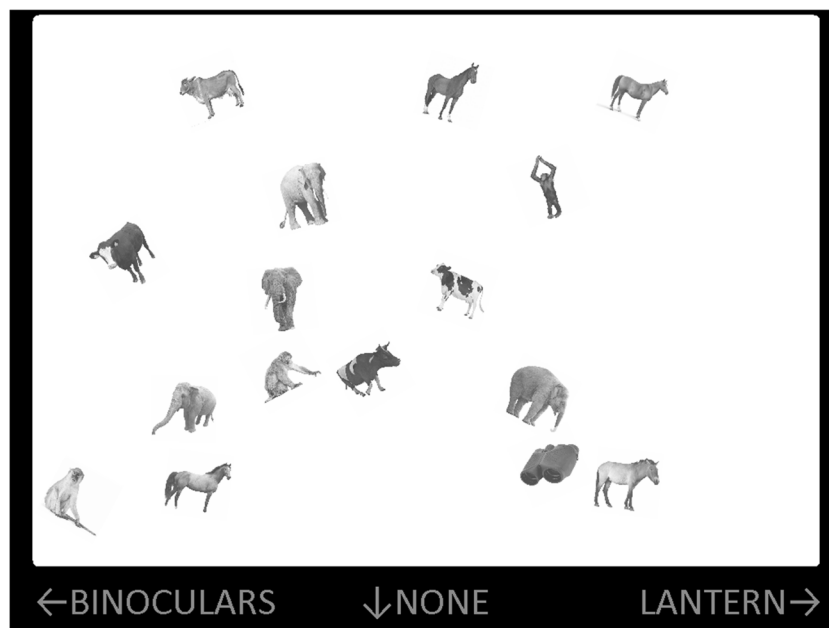


Fig. 5 A sample visual search screen with a pair of binoculars as the target. Participants searched for any pair of binoculars or any lantern throughout the whole experiment. In the first part of the experiment, participants searched among distractors that were either visually dissimilar to the target – mammals (pictured) or plants, or visually similar to the target – man-made artifacts. In target-absent trials, the distractor set consisted of four exemplars from four subordinate pre-defined categories

(i.e., four exemplars of monkeys, cows, horses, and elephants in the mammal-training condition). The same 16 exemplars showed up in each trial in the training phase. On target-present trials, a target replaced a distractor. This set of distractors was then replaced by a new set that was visually similar to the target artifacts in the second transfer-test phase of the study to assess the target template people had learned

distractors were replaced with 16 images of artifact distractors. Hence, participants in the artifact-training condition saw new exemplars within the same superordinate distractor category (artifact), while those in mammal-training and plant-training conditions were introduced to new distractor categories. Thus, the search task now featured distractors that were both visually and semantically similar to the targets for all conditions. The target definition remained the same, such that there were no changes in the task description (i.e., search for either a pair of binoculars or a lantern). There were a total of 238 pictures in the image bank, but each participant only saw a random set of 64 among these pictures. This was to avoid idiosyncratic features in any particular images harming the generalizability of the experiment.

A total of five transfer-test blocks, each with 36 trials, were administered. Afterwards, the participant was thanked, debriefed, and dismissed. Each experimental session took about 40 min.

Data-cleaning procedure

In the main text, only target-present trials are presented. The pattern for the target-absent trials is very similar, albeit slower, as expected (see [Appendix](#)).

Data from participants with an overall accuracy 3 standard deviations below the overall group mean were discarded. This resulted in the exclusion of three participants' data. We also excluded the response time of a trial if the preceding trial had an incorrect response. This is customary in visual search experiments as an incorrect response followed by feedback tends to lengthen the response of the following trial. This resulted in the exclusion of 5.4% of all trials. Trials with response times longer than 15 s or 3 standard deviations above individual participant means were recursively excluded, as they suggested a lapse of attention in the task. Those with response times shorter than 200 ms were also excluded, as they indicated likely errors in response recording. This resulted in the exclusion of an additional 0.4% of all trials. A total of 5.8% of all trials was excluded with the above-mentioned data-cleaning criteria.

Response time was deemed the main dependent variable in the experiment, and participants were asked to prioritize accuracy over speed during the session. Nevertheless, accuracy was not always maintained at ceiling. To simplify our analysis, we adopted a proposal to integrate response times with accuracy (Vandierendonck, 2017). The linear integrated speed-accuracy score (LISAS) adjusts response times upward with respect to the error rates.

$$LISAS = RT_{ij} + PE_{ij} \times \frac{s_r}{s_e}$$

where RT_{ij} is the mean response time of participant i in condition j , PE_{ij} is the proportion of error of participant i in condition j , s_r is standard deviation of the response time, and s_e is

the standard deviation of responses, coded as 0 and 1 for incorrect and correct responses, respectively. Conceptually, LISAS can be thought of as an adjusted response time, assuming a participant made no response errors. The adjustment did not change the general pattern or the conclusion of the results, when analyzed separately as (non-adjusted) response time and accuracy. It augments the effect size when there is a difference between conditions, compared to analyzing response time or accuracy alone, as LISAS combines the effects of the two measures, and the overall non-adjusted response times are positively correlated with error rates.

To aid the flow of the analysis, we only included LISAS-corrected response time analyses in the main text. Interested readers are encouraged to refer to separate response-time and error-rate analyses, and analyses on target-absent trials in the [Appendix](#).

Results

Predictions and outline of the analysis

We have three main areas of analysis: the training phase; the beginning of the transfer phase; and the entire transfer-test phase, with different predictions about each.

With reference to Schmidt and Zelinsky (2017), we expected the "artifact-training" group would spend the longest time in the training phase, as the artifact distractors were the most visually similar to the target objects. What was unclear was whether people would overcome the difficult search task through repeated searching. The first part of the analysis attempts to answer the question with an ANOVA. Any differences in the training phase must be attributed to differences caused by the distractors, since all three groups searched for the same targets but among different distractors.

However, our main analysis regarded performance at the beginning of the transfer phase, rather than during the training phase: In particular, we compared the performance of the first transfer-test block between conditions. Differences in the conditions denote the effect of search history, since the search task during the transfer phase was identical for all of the groups. Slow response times indicate that the training phase did not efficiently prepare the participants for this new search task.

The third part of the analysis looks at the entire transfer-test phase. Since all the groups had identical search displays, we can examine how long it took the participants to overcome the training effects. A persistent difference between groups at the end of the transfer test would indicate long-term effects of the different training tasks.

Training phase

The left panel of Fig. 6 shows participants' response times in the training phase. Participants spent more than twice as long

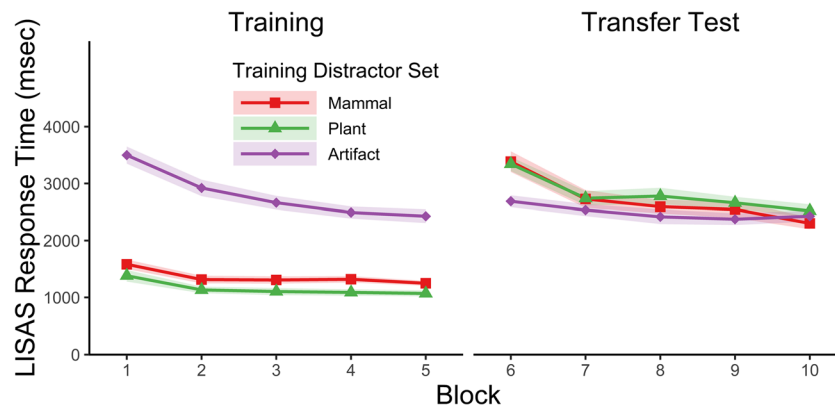


Fig. 6 Response times of Experiment 1. **Left:** All groups searched for a set of four pairs of binoculars and four lanterns. However, during training, the groups differed in the distractors they needed to distinguish these items from. The artifact group had a difficult set of distractors, as, like the binoculars and lanterns, their distractors were also man-made artifacts. The plant and mammal-training groups had considerably easier distractors to differentiate from the targets, reflected in their much faster response time. **Right:** During the transfer-test phase, beginning at block 6, all participants saw the same search displays, with binocular targets and

lantern targets now embedded amongst a new set of artifacts that was also hard to differentiate from the targets. The distractors changed even for the artifact-training group to new artifacts. There was a substantial switch cost for the groups that had been performing a search with easy-to-differentiate distractors when a new set of distractors was introduced that required a difficult search, but almost no cost for the artifact-training group which had already been performing this difficult search task. Error bars denote between-subject standard errors of the means

detecting the targets in the artifact-training condition (2,802 ms, SD = 503 ms), compared to either mammal-training (1,355 ms, SD = 236 ms) or plant training (1,156 ms, SD = 278 ms) conditions, $F(2,62) = 135.1$, $p < 0.001$, $\eta^2 = 0.69$. A post hoc Tukey Honest Significance Difference (HSD) test indicated that the plant training group had the shortest response times. The plant-training group was not significantly faster than the mammal-training group in detecting the targets ($p_{\text{adj}} = 0.20$), and both the plant-training and mammal-training groups were significantly faster than the artifact-training group ($p_{\text{adj}} < 0.001$).

In addition, visual search became more efficient over time, $F(4,248) = 42.9$, $p < 0.001$, $\eta^2 = 0.06$. The rate of improvement did differ across conditions, but the difference was small, $F(8,248) = 8.7$, $p < 0.001$, $\eta^2 = 0.02$.

First transfer test block

In the transfer-test phase, participants in all three conditions searched for binoculars and lantern targets among new artifact distractors (that is, the distractors changed for all groups, including the artifact-training group). Any differences between the groups indicate a learning effect from the experience of different distractors during training. As we are interested in how search history affects current search efficiency, the first transfer-test block is especially informative. The right panel of Fig. 6 captures the transfer test performance.

The response-time data were submitted to a one-way ANOVA, with condition as the factor. The ANOVA indicates a difference between conditions, $F(2,62) = 8.6$, $p < 0.001$, $\eta^2 = 0.22$. Specifically, the artifact-training group took the shortest time to locate the targets (2,689 ms, SD = 508 ms). The group

was significantly faster than both the plant-training (3,343 ms, SD = 567 ms, $p_{\text{adj}} < 0.01$) and the mammal-training (3,384 ms, SD = 812 ms, $p_{\text{adj}} < 0.01$) groups in a Tukey HSD test. Performance of the latter two groups did not differ significantly ($p_{\text{adj}} = 0.98$).

Transfer test phase

The rest of the analysis on the transfer-test phase focuses on how quickly the groups converge to the same level of performance. As all groups encountered the same stimuli, we expect the groups would eventually arrive at similar response times and hit rates. Response-time data were submitted to a 3×5 ANOVA, with condition and blocks as the factors.

The overall effect of condition was significant, $F(2,62) = 3.28$, $p < 0.05$, $\eta^2 = 0.05$. The overall response time for the artifact-training condition was the shortest (2,490 ms, SD = 365 ms), followed by the mammal-training group (2,712 ms, SD = 534 ms), and the plant-training group (2,810 ms, SD = 391 ms). As can be seen from Fig. 6, the difference was mainly driven by the shorter response time of the artifact-training group in the first transfer-test block. Performance between conditions quickly converged within the next block. This observation was supported by an improvement in performance over blocks, $F(4,248) = 25.8$, $p < 0.001$, $\eta^2 = 0.14$. The interaction between condition and block was also significant, $F(8,248) = 3.0$, $p < 0.01$, $\eta^2 = 0.03$.

Discussion

As expected, we found that the similarity between the targets and distractors drastically affected visual search performance

in the training phase. Participants were much faster when they searched for the same targets among plant and mammal distractors than among artifact distractors. Binoculars and lantern distractors should be easier to search for among mammal and plant distractors compared to artifacts, because the targets and the artifact distractors are all man-made, and thus share both conceptual features, and perhaps more importantly, more visual similarities than with the other two distractor categories (e.g., Long, Störmer, & Alvarez, 2017). In addition to this search difficulty manipulation, we also found that participants improved over the course of the training phase, in part because they learned about the visual features of targets and distractors (as demonstrated by the transfer phase).

The transfer-test phase showed the benefits of having a more difficult training task. Participants in the artifact-training conditions performed the worst in the training task based on both speed and accuracy measures, but they performed the best when a new set of artifact distractors was introduced in the transfer-test phase. This is consistent with the idea that the artifact training allows people to develop detailed target templates for the task, which in turn allows them to distinguish the targets even from a set of artifact distractors. By contrast, the mammal-training and plant-training groups had no need for detailed target templates and so might not have developed a detailed enough representation. Thus, they had a greater difficulty with the new and more difficult-to-discriminate distractors.

In addition, we showed that this detailed target template benefit did not last very long. Within a block of 36 transfer test trials, participants from the mammal-training and plant-training conditions caught up, showing that they could generate a detailed enough search template for the task. Their performance was no longer distinguishable from the artifact-training group by the second transfer-test block. Nevertheless, this switch cost is of practical importance. In a real-world setting, such as airport baggage checkpoints, observers do not usually have a block of trials to adjust to a new set of distractors. A change in the distractor sets would mean a delay in response, or a drop in search accuracy.

To what extent is the learning during the training phase semantic or verbal rather than visual in nature? In theory, participants may learn labels for the distractors in the first block, and reject them in subsequent searches based on the meaning of the images. When the set of distractors was replaced, the learned semantic labels of the distractors can no longer help participants reject them, perhaps resulting in lower performance in the transfer-test phase.

Attentional guidance through semantics is likely quite a bit weaker than guidance via visual features of the target, and probably cannot fully account for the effect (Bravo & Farid, 2009). For one thing, there was a difference in performance between the different conditions in the training phase, even though the semantic labels for all items in all conditions are

equally accessible, rejecting distractors based purely on semantics should be comparable. We also address this in Experiment 2, which examines whether there are semantic and/or general switch costs, or whether the level of detail required in the target template after the switch in distractors is the critical factor.

Experiment 2

In Experiment 2, participants searched for binoculars and lantern targets, as in Experiment 1. In one condition, participants searched for the targets among artifact distractors in the training phase; in the other two conditions, they searched for the targets among mammal and sea animal distractors, respectively. We changed the transfer test so that a detailed target template was not required for the task. In particular, during the transfer-test phase, all participants searched for the binoculars and lantern targets among mammal distractors. Searching for the targets among mammal distractors required only a coarse discrimination. Comparing the transfer tests between Experiments 1 and 2, the one in Experiment 2 was much easier. This allowed us to tease apart whether the transfer training benefits were general benefits that arise whenever participants performed a difficult task (e.g., Bjork, 1994) or were specific to the need for a detailed target template (which was learned only by the artifact-training group). In addition, this manipulation allows us to tell whether the detailed target template learned by participants in the artifact-training group was specific to the artifacts they had learned to discriminate the targets from (e.g., Geng, DiQuattro, & Helm, 2017; Navalpakkam & Itti, 2007; Scolari & Serences, 2009; Yu & Geng, 2019) or whether it was simply a detailed target representation. If the template was specific to the distractors, even switching to an “easier” search with different distractors would be expected to come at a cost to search performance; if it was simply a detailed target template, then no such cost would be expected.

Methods

Design

The overall design of Experiment 2 was identical to that of Experiment 1. Participants went through one practice block, five training blocks, and five transfer-test blocks. Participants searched for binoculars and lantern targets among distractors. We made two changes to the stimuli.

Most critically, in the transfer test, we utilized only mammal distractors, which are dissimilar from the targets and should create an easier search task. Thus, the transfer test was designed to require less detailed target templates than the transfer test in Experiment 1. In addition, we made a small change in that we replaced plant distractors with sea animals

during training. We did this in an attempt to create a wider range of performance across conditions, although it turned out to have no practical effect on the experiment.

Participants

Seventy participants (53 female) were recruited from the same population described above. The participants had a mean age of 20.7 years at the time of their participation. All participants gave informed consent prior to their participation and they participated for partial course credits.

We obtained a sample size that was comparable to Experiment 1 so that the results of the two experiments could be compared. Participants were randomly assigned into one of the three conditions, resulting in 23 participants in each of the mammal-training and sea animal-training conditions, and 24 participants in the artifact-training condition. Data from six participants were discarded using the criteria detailed in Experiment 1.

Apparatus and stimuli

Equipment used in Experiment 2 was identical to that of Experiment 1. Images used in Experiment 2 were identical to those of Experiment 1, except that the plant images were replaced by sea animals, in an attempt to create a more uniform stimulus set and create a wider range of performance. All images were obtained from the same source as stated in Experiment 1.

Procedure

Experiment 2 was administered in exactly the same way as in Experiment 1, except for the following changes.

The three training groups saw mammals, sea animals, and artifacts as distractors during the training phase. In the transfer test, all three groups searched for a lantern and a pair of binoculars as targets, among a new set of mammal distractors. Hence, the transfer test of Experiment 2 should be notably easier than that of Experiment 1.

Data-cleaning procedure

As in Experiment 1, when calculating response time for target-present trials, we removed trials with incorrect responses in the preceding trials. This led to 4.5% of all trials being removed. Trials with response times that were 3 standard deviations above a participant's mean response time, or those with response times shorter than 200 ms or longer than 15,000 ms were also removed. This constitutes another 0.5% of all trials. In total, 1,142 trials were removed from further analysis, accounting for 5.0% of all trials.

Results

As in Experiment 1, participants in all conditions searched for the same targets in the training phase; the conditions differed only in the distractor set. All participants had the same targets and encountered the same mammal distractor set during the transfer-test phase. Thus, differences in performance during the test phase can only be attributed to differences in target template or strategy developed during the training phase. In particular, performance in the first transfer-test block indicates the most direct influence of the training. Regardless of how the participants performed in the first transfer-test block, we expected performance of all groups to eventually converge at the end of the transfer-test phase.

We present only data of the LISAS-correct response time of target-present trials in the main text, making the analysis consistent with Experiment 1. Separate analyses for unadjusted response time, accuracy, and target-absent trials are deferred to the [Appendix](#).

Training phase

We submitted the response-time data to a 3×5 ANOVA, with condition as the between-subject factor and block as the within-subject factor. The left panel of Fig. 7 shows that features of distractor sets again affected performance during training. In general, participants in the artifact-training condition (2,588 ms, SD = 531 ms) spent around twice as long detecting the targets, compared to either mammal-training (1,334 ms, SD = 297 ms) or sea animal-training (1,221 ms, SD = 182 ms) conditions, $F(2,61) = 92.7$, $p < 0.001$, $\eta^2 = 0.59$. A post hoc Tukey HSD test indicates that the artifact-training group was significantly slower than the two other groups ($p_{\text{adj}} < 0.001$), but the difference between the sea animal-training and the mammal-training groups was not significant ($p_{\text{adj}} = 0.57$). This visual search task became faster over blocks, $F(4,244) = 43.5$, $p < 0.001$, $\eta^2 = 0.08$. Indicated by an interaction between the two factors, the rate of improvement was steeper for the artifact-training condition than the two other conditions, because of the relatively slower initial response times, $F(8,244) = 8.4$, $p < 0.001$, $\eta^2 = 0.03$.

First transfer-test block

In the transfer-test phase, participants in all three conditions searched for binoculars and lantern targets among new mammal distractors. Since the visual stimuli were identical across conditions, any differences in the conditions should be attributed to participants' training experience.

As in Experiment 1, the response-time data were submitted to a one-way ANOVA, with condition being the factor. The differences between groups were within 60 ms (response time (RT): 1,293–1,348 ms, SD = 240–305 ms). Unlike the results

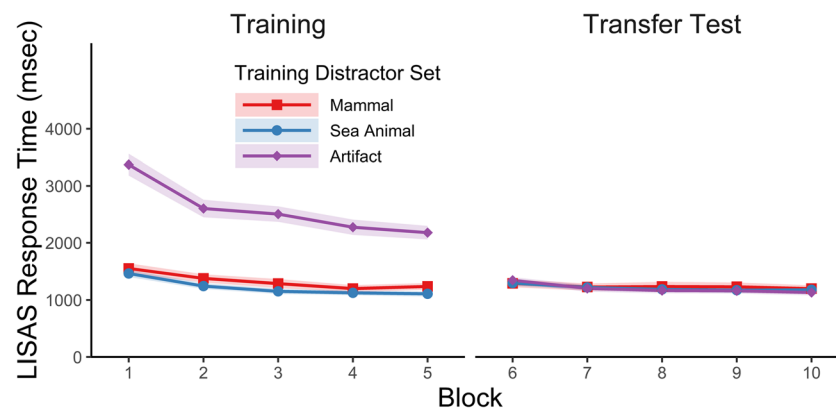


Fig. 7 Response times by condition and block in Experiment 2. **Left:** Participants were faster at searching with the same set of distractors being used. **Right:** The cost for switching was negligible when a new set of

distractors was introduced. Error bars in the graph denote between-subject standard error of the means

in Experiment 1, the ANOVA does not indicate any differences between conditions, $F(2,61) = 0.28$, $p = 0.76$, $\eta^2 < 0.001$, $BF_{10} = 0.16$.

Transfer-test phase

Response-time data for the whole transfer-test phase were submitted to a 3×5 ANOVA, as in Experiment 1, with condition and block being the factors. Response times across the three conditions were highly similar throughout the transfer-test phase, $F(2,61) = 0.14$, $p = 0.87$, $\eta^2 = 0.003$, $BF_{10} = 0.21$. A significant improvement in response times across blocks was seen, again demonstrating that performance was not at “ceiling” at the beginning of the transfer phase, although the improvement was small, $F(4,244) = 9.4$, $p < 0.001$, $\eta^2 = 0.04$, and the interaction between the two factors was not statistically significant, $F(8,244) = 0.82$, $p = 0.59$, $\eta^2 = 0.007$, $BF_{10} = 0.03$.

Discussion

Binoculars and lantern distractors were easier to search for among mammal and sea animal distractors during training, compared to artifact distractors. This is likely because artifact distractors are all man-made, and they share more visual similarities with the targets compared to the other two distractor categories (e.g., Long, Störmer & Alvarez, 2017). Ultimately, this difference gives rise to the differences in response times between the artifact-training condition and the two other conditions during training.

In the transfer-test phase, we found that while artifact distractors made the visual search task harder in the training phase, they did not have an immediate benefit in the transfer phase. Thus, there was not a general desirable difficulty effect where the harder participants worked on a task, the better they learned about the target (Bjork, 1994), and therefore the better they performed during the transfer test. In addition, participants in the artifact-training group did not appear to have

learned a template that was specific to the artifact distractors they had learned to discriminate the targets from (e.g., Geng, DiQuattro, & Helm, 2017; Navalpakkam and Itti, 2007; Scolari & Serences, 2009; Yu & Geng, 2019). If the template was specific to the distractors, even switching to an “easier” search with different distractors would be expected to come at a cost to search performance, which did not occur. This was true even though there was plenty of room for such participants to perform less well during transfer; for example, participants trained with mammal distractors improved by >250 ms during training, so there could potentially have been a 250-ms cost in the artifact-distractor group if their template was completely distractor-specific. Instead, given the results from Experiment 1, the current results are consistent with our hypothesis that the training groups learned different amounts of detail in their target templates, i.e., that the artifact-training group learned a more specific template (Geng & Witkowski, 2019). In the transfer-test phase of this experiment, even a less detailed template would support performance, and so there were no additional benefits of a more detailed template, whereas in Experiment 1, a more detailed template was required, and so already having learned this template supported improved performance.

In particular, the main difference between Experiments 1 and 2 was the distractor sets used in the transfer test. Experiment 2 utilized a set of mammal images as distractors, while Experiment 1 utilized a set of artifact images as distractors. In Experiment 1, we showed that when the distractors were switched, even when the targets stayed the same, people needed time to adjust to the new context. In Experiment 2, with a less detailed template needed during the transfer test, this did not occur. The artifact-training group generated detailed enough target templates that survive change in the context.

Importantly, Experiment 2 also provides evidence against the semantic/verbal label account that might be thought to account for the results of Experiment 1. Experiments 1 and 2

are symmetric from the perspective of semantic/verbal labels, yet they led to a large difference in transfer performance. This is consistent with our account of a more detailed versus less detailed target template but inconsistent with the idea of verbal labels being primarily responsible for transfer costs.

Experiment 3

Experiment 1 showed that when the transfer test requires fine discrimination between the targets and the distractors, a previous history of fine discrimination helps observers prepare for the task. With previous experience in only coarse target-distractor discrimination, observers become slower and/or less accurate when there is a change in task requirement to require more fine discrimination. Experiment 2 shows that this is not a general switch cost, but likely due to how precise a target template observers have in mind.

In Experiment 1, the artifact-training group showed very minimal cost when a new set of distractor exemplars was introduced. However, the set of new exemplars in the transfer test was not only still similar to the targets, but also belonged to the same distractor categories. This raises the question of whether the highly detailed target templates generated during this training phase were able to survive various kinds of context change. In other words, if the training and the transfer test both require fine target-distractor discrimination, but the distractors undergo a larger change, would the target templates still be useful? Or are they at least somewhat distractor specific, in addition to being detailed? To address this, Experiment 3 was designed to test how specific the target templates are with respect to the distractors in the training phase.

Methods

Design

The structure of Experiment 3 was similar to that of Experiments 1 and 2. Three groups of participants were recruited, all searching for the same targets. Two of the groups were identical to the mammal-training and artifact-training groups in Experiment 1. Participants were asked to search for binoculars and lantern targets among mammal or artifact distractors in the training phase. In the transfer-test phase, new exemplars from the same distractor categories were introduced. This provides a replication of the artifact-training group in Experiment 1. We called this artifact-training group “Artifact Training, Different Exemplar.”

The last condition was identical to the “Artifact Training, Different Exemplar” group, in that participants searched for the targets among artifact distractors (e.g., cellphone) in both the training and transfer-test phases. However, in the transfer-

test phase, a new set of distractors from a new set of artifacts (e.g., calculator) was introduced. For example, a participant might see cellphone, microscope, hourglass, and guitar as distractors in the training phase; and calculator, camera, goggle, and trumpet in the transfer-test phase. We named this condition “Artifact Training, Different Sub-category.” Comparing the performance in the latter two groups in the transfer-test phase gives us an idea of how specific the target templates are.

Participants

Participants were recruited from the same subject pool described in Experiments 1 and 2, but due to the COVID-19 pandemic, we modified the experiment such that participants could take part online. Ninety-two participants were recruited, with five of these participants discarded according to the data cleaning procedure outlined in Experiment 1. Participants received partial course credit for their participation.

Apparatus and stimuli

The stimuli used were the same as in Experiment 1. An additional four sets of artifacts (calculator, camera, goggle, and trumpet) were introduced as required by the new “Artifact Training, Different Sub-category” condition.

Participants were asked to take part in the experiment using their own devices. The instruction and stimulus displays were delivered through the web browser. As a result, the screen size, and hence the stimulus size, were not strictly controlled.

Procedure

The procedure of the experiment was identical to that in Experiments 1 and 2, except that the informed consent was delivered to the participants electronically (not in paper form as in Experiments 1 and 2). The procedure was approved by the University of California, San Diego’s Internal Review Board.

Data cleaning

Following the data-cleaning procedure outlined in Experiment 1, a total of 7.1% of all trials were removed prior to the calculation of response time. These included trials due to a preceding incorrect response and trials with response times that were too long or too short.

As in Experiments 1 and 2, only the LISAS-corrected response times of the target-present trials are presented in the main text. Interested readers can refer to the [Appendix](#) for additional analyses.

Results

Training phase

Response-time pattern in the training phase is similar to that in Experiment 1. Overall, there was an effect of block, $F(2,84) = 47.6$, $p < 0.001$, $\eta^2 = 0.11$, condition, $F(4,336) = 45.8$, $p < 0.001$, $\eta^2 = 0.36$, and an interaction between block and condition, $F(8,336) = 6.3$, $p < 0.001$, $\eta^2 = 0.03$.

The condition and interaction effects are driven by the fact that the mammal-training group (mean = 1,269 ms, SD = 217ms) responded faster than the two artifact-training groups ($p_{\text{adj}} < 0.001$). As would be expected since they had identical tasks and stimuli, participants in the two artifact-training groups took similar amounts of time finding the targets in the training phase (Different Sub-category: 2,594 ms, SD = 746 ms; Different Exemplars: 2,480 ms, SD = 523 ms). Performance between the two groups did not differ from each other in the Tukey HSD test ($p_{\text{adj}} = 0.72$).

First transfer-test block

In the transfer-test blocks, all three groups searched for the same targets among a new set of distractors. For the mammal-training group, the distractors switched from a set of mammal images to a set of artifact images. This group had the longest response time in the first transfer-test block (mean = 2,719 ms, SD = 593 ms). The “Artifact Training, Different Exemplars” group saw a new set of distractors from the same artifact categories, and they were fastest among the three groups (mean = 2,238 ms, SD = 698 ms). The “Artifact Training, Different Sub-category” group saw a new set of distractors from different artifact categories. The changes in the distractor set was in between the two other groups, and their response times were also in between the two other groups (mean = 2,524 ms, SD = 615 ms). A between-subject ANOVA indicates a significant difference between the groups, $F(2,84) = 4.2$, $p = 0.02$, $\eta^2 = 0.09$. This is driven by the difference between the mammal-training group and the “Artifact Training, Different Exemplars” group, as indicated by a Tukey HSD test ($p_{\text{adj}} = 0.01$). Comparisons between the “Artifact Training, Different Sub-category” group and the two other groups are not statistically significant ($p_{\text{adj}} > 0.2$), reflecting the intermediate position of this group in reaction time.

Transfer-test phase

As shown in the right panel of Fig. 8, performance of the three groups in the transfer test was largely similar, except for the first transfer-test block. There is little difference across conditions, $F(2,84) < 1$, $BF_{10} = 0.22$. There is a main effect of block, $F(4,336) = 29.2$, $p < 0.001$, $\eta^2 = 0.008$, and an interaction of

condition and block, $F(8,226) = 2.7$, $p < 0.01$, $\eta^2 = 0.02$. The interaction was driven by performance in the first transfer-test block, as explained in the previous paragraph.

Discussion

The two artifact-training groups differ in the distractors they encountered in the transfer-test phase. While there were minimal changes in the “Artifact Training, Different Exemplars” group, distractors in the “Artifact Training, Different Category” group changed visually and semantically. The element that we held constant between the two groups, across both transfer and testing blocks, was the level of detail required to perform the task – with the only difference being the extent to which the distractors stayed more versus less constant. Although not statistically significant, the intermediate performance of the “Artifact Training, Different Category” data was suggestive of the idea that generating a detailed target template was not 100% sufficient to survive the change in context, but that, in this context, as in much previous work (e.g., Avraham, Yeshurun, & Lindenbaum, 2008; Duncan & Humphreys, 1989; Pashler, 1987), target templates likely are shaped not only by information about the targets, but also the differences between the targets and the distractors.

General discussion

In the first two experiments, we had participants search for two categories of targets repeatedly. In both cases, the distractors were always consistent throughout an entire training phase. This allowed participants to learn about the visual features of the distractors, and create appropriate target templates. Although participants were neither explicitly told nor encouraged to learn about the distractors, they nevertheless learned about distractors during the visual search task, and response times became shorter when they searched for targets among the same set of distractors over time. Our critical manipulation was the introduction of a new set of distractors in the middle of the experiment. The results suggested that coarse target templates are generated during training when the targets are easily distinguishable from distractors (e.g., searching for binoculars among mammal distractors). The coarse target templates do not support performance well in a new context that requires detailed target templates (e.g., searching for binoculars among artifact distractors). On the other hand, if the training requires demanding target-distractor discriminations (e.g., searching for binoculars among artifact distractors), detailed target templates are likely to be generated, and these templates are sufficient for both hard and easy searches in transfer tasks. Importantly, changing the distractors per se did not impact search performance (e.g., the change from sea animals to mammals, or from artifacts to

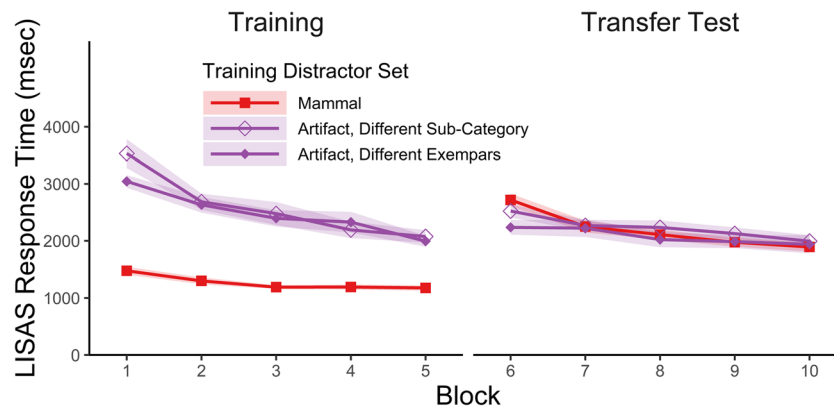


Fig. 8 Response times by condition and block in Experiment 4. **Left:** The general training phase pattern replicated that in Experiment 1. **Right:** The cost for switching was negligible when a new set of distractors was introduced. Error bars in the graph denote between-subject standard error of the means

mammals in Experiment 2). Instead, as long as the level of detail in the target template was sufficient to support the new search task, participants performed well during the transfer test. The main result was supported by Experiment 3. In Experiment 3, we showed that people are sensitive to the difference between targets and distractors in the training phase, and that a highly detailed template will better survive a transfer test with new distractors. However, even a highly detailed template does not seem to be sufficient to ensure 100% transfer: Once the context changes, with new distractors, the highly detailed target templates may need to be adjusted to take into account the new context. Overall, our results indicate the importance of having detailed search templates for transfer learning, but also a need to include relevant distractor information in the target templates.

Similarity

To ensure our stimuli did in fact require more versus less detailed target templates, we validated the visual and conceptual similarity of the stimuli. Using human similarity rankings, we found that categories clustered as expected: within-category stimuli form clusters in the MDS plots, suggesting high similarity within the categories we defined. In addition, artifacts were indeed much more similar to the search targets than were the other stimuli, suggesting the search tasks did require a more detailed template with artifacts as distractors.

This similarity analysis was supported by an additional image-similarity analysis using deep neural networks. Deep convolutional neural networks are useful models of human recognition and the human visual system (e.g., Yamins et al., 2014), and deep nets trained on categorization are sensitive to some extent to both visual and semantic features (e.g., Jozwik et al., 2017; Peterson et al., 2018). Such measures also reliably predict memory confusability (Brady & Störmer, 2020). Here, we found they captured the categorical structure of the individual images, and that they indicated that artifacts

would be more similar to the search targets than the other stimuli, providing further validation of our stimulus design.

Practical implications

Visual search has been an instrumental tool in understanding the human mind for decades. One reason for its popularity has to do with the potential practical applications. Revisiting the baggage-screening scenario we discussed earlier, airport baggage screeners generally search thousands of bags in a shift. The distractors are generally very limited and drawn from similar categories repeatedly. Our results support the idea that if passengers stuff their bags with unusual items, this could interfere with the screeners' search performance, with the amount of this cost depending on the degree to which the screeners have extremely detailed target templates versus coarser templates.

Our research design maps on this and other related real-world scenarios closely. Participants searched for a limited set of targets among a repeated set of distractors. Thus, in real-world tasks where new, but similar, distractors are constantly encountered, the task is likely to continue to be demanding (e.g., these distractors are similar to the target), we expect performance to match the results of our transfer test in Experiment 1. Our results shed light on the change in performance over time in such a situation, arguing that a change in the previously familiar distractors is likely to impose a cost to performance. The cost is likely to be minimal when it involves switching out exemplars within the same distractor categories that are semantically and visually similar, but when the change involves a new set of distractor categories, the cost in performance can be substantial (first transfer-test block in Experiment 1).

A practical solution to emolliate the effect in real-world scenarios would involve increasing the diversity of distractors (e.g., Hout & Goldinger, 2012). In the baggage-screening context, this could, for example, involve digitally inserting

distractors of different categories, in addition to the insertion of predefined targets that is currently in practice.

Connection to the visual search literature

Repeated visual search

The current study is closely related to a number of previous results in the literature. Here we highlight some differences between the current study and this previous work.

Hout and Goldinger (Experiments 1C and 2C, 2010; Experiment 2, 2012) studied how repeated distractors lead to improvements in search performance. Participants in the studies searched for a target among the same set of distractors within each block. The study utilized grayscale images of objects like we did in the current study. Unlike the current study, Hout and Goldinger's experiments specified a new target on each trial. The length of time that participants encountered the same set of distractors was also shorter than the current study. Participants saw the same set of distractors for no more than 40 trials. In addition, the association between the distractors was low. Differences between distractor sets were not systematically manipulated. The researchers did not find any costs when distractors switched from one block to the next.

Wolfe, Klemm, and Dahlen (2000) had participants search for a target letter among three or five letter distractors. In their repeated search condition, the same search display was visible to the participants throughout the experiment. A cue was presented to indicate the beginning of a trial and the target for the current trial. Participants became faster over the 700-trial session. Importantly, the researchers did not find any improvements in search efficiency when taking set size into account. They showed that participants spent around 40–60 ms per item in the search regardless of practice.

The current study asked a similar question to these researchers: does having repeated distractors help visual search. The major difference between the current study and the other studies mentioned here is that the target set in the search task was kept constant in the current experiment. The design allowed participants to not only learn about the distractors, but also the relationship between the target and the distractors. As Duncan and Humphreys (1989) and Pashler (1987) convincingly showed, one major determinant of visual search efficiency is the target-distractor discriminability. Our design provided an opportunity for participants to learn and improve this discrimination. As we are interested in a learning effect, we also had a longer training period for each set of distractors, compared to Hout and Goldinger's study. It appears to us that participants in certain conditions kept improving even after 180 trials with the same distractors, so a longer training is more ideal in studying learning effects of this sort.

Our results are in agreement with those discussed above, and showed that the learning effects can be enhanced with the use of real-world images, without varying the target set, and extending the training period that participants engage with the same set of distractors. They also show that memory is better utilized when the search task is sufficiently difficult (Solman & Smilek, 2012).

Veridicality of target templates

In the classic visual search literature, it has often been assumed that a participant's goal is to build an exact replica of the target. The closer the attentional template is to the actual target, the better the search performance. As an example previously noted in the *Introduction*, Bravo and Farid (2012) asked participants to search for a fish among naturalistic undersea distractors. The more details about the target given to a participant, the better the search was. More recently, however, researchers have challenged the view (see a recent review by Geng & Witkowski, 2019). These authors argue that participants try to build a target template that would allow them to maximally differentiate the target from the distractors. Support of the idea comes from work like that of Navalpakkam and Itti (2007), who showed that participants generated a distractor that was biased against the distractors, and has since been shown in a variety of other work (e.g., Geng, DiQuattro, & Helm, 2017; Scolari & Serences, 2009; Yu & Geng, 2019). Our experiments further support this argument, although in a slightly different way: The target templates did not appear to be strictly designed for distinguishing the target from an exact distractor set, as switching to an easier-to-distinguish set of distractors did not introduce a large or significant cost. We did find that given the same task (searching for the same targets), participants did not build the same target templates for all conditions. Instead, the group that was asked to search for targets among difficult-to-differentiate distractors built more detailed target templates for the task (see also Yu & Geng, 2019).

Possible mechanisms

We argue that people learn not only the characteristics about the targets and the distractors during visual search, but they also fine-tune their target templates that would maximally separate the feature space between the targets and distractors. As shown in Experiment 3, this target-distractor discriminability is likely to be task dependent and differ from condition to condition. When discrimination is easy, such as in the case of finding binoculars among mammal distractors, a coarse target template may be generated. The locus of this target template possibly lies in the mid-level features, such as curves, that are more likely to appear in animals than in artifacts (e.g., Long, Störmer, & Alvarez, 2017). When target-distractor

discrimination is difficult, such as the case of finding binoculars among other artifacts, a detailed target template is likely to be generated. Participants might also be motivated to generate a distractor template to efficiently reject task-irrelevant stimuli. Similar arguments had been made by Won and Geng (2018) with simpler stimuli. With the current design, we cannot partial out the effect of distractor template from target template. However, exploration of this concept is important to build a full understanding of visual-search performance. The locus of where the target template is generated is currently unclear. There may be very specific shapes that would differentiate binoculars and lanterns from our artifact distractors. Semantic labels of the objects may also be utilized if visual features alone are highly inefficient. Our data suggest that the effect reported here is mainly visual, but further studies may manipulate semantics of the stimuli to paint a full picture of how these factors interact.

While possible in some conditions, low-level priming does not seem to explain the learning effect in the current experiments (Kristjánsson & Driver, 2008). Priming happens when stimuli appear prior to the current ones, and bias people to process certain stimuli. The effect is usually short-lived. The continued improvements in visual search during the training phase may be partially accounted for by priming, but not in the transfer test. For one thing, if priming can account for most of the transfer effect in Experiment 2, we should expect a learning effect in the corresponding conditions in Experiments 1 and 3. The fact that there is an asymmetric training effect in the three experiments suggested that priming is unlikely to be the sole driving force in learning. The same logic applies to low-level visual adaptation such as frequency and feature adaptation due to prolonged exposure to certain visual features (Kohn, 2007).

Conclusion

In three experiments, we showed that people learned about the properties of distractors in a visual search task. Their prior experience affected the nature of their target templates. We found that people form target templates that include information about the distractors, and that when targets and distractors are more difficult to discriminate, people create target templates with more details, which are more likely to be transferable to a new context. In real-world search contexts, this suggests that a diverse set of distractors would be beneficial to the generalizability of visual search training. Creating scenarios with diverse distractors could thus avoid decreases in search performance when the search context and distractor set change abruptly.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13414-020-02213-w>.

Open Practices Statements Data associated with the experiments reported in the study are available at <https://osf.io/x6z53/>. None of the experiments was preregistered.

Acknowledgements The authors would like to thank Joy Geng, Michael Hout, and 2 anonymous reviewers for helpful comments on earlier versions of the manuscript. This research was generously supported by NSF CAREER Grant 1653457 and NSF Grant 1829434 to TFB.

References

- Avraham, T., Yeshurun, Y., & Lindenbaum, M. (2008). Predicting visual search performance by quantifying stimuli similarities. *Journal of Vision*, 8(4), 9.
- Bjork, R. A. (1994). Institutional impediments to effective training. *Learning, Remembering, Believing: Enhancing Human Performance*, 295–306.
- 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, USA*, 105(38), 14325–14329.
- Brady, T. F., & Störmer, V. S. (2020, June 17). Greater capacity for objects than colors in visual working memory: Comparing memory across stimulus spaces requires maximally dissimilar foils. <https://doi.org/10.31234/osf.io/25t76>.
- Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision*, 9(1), 34.
- Bravo, M. J., & Farid, H. (2012). Task demands determine the specificity of the search template. *Attention, Perception, & Psychophysics*, 74(1), 124–131.
- Chetverikov, A., Campana, G., & Kristjánsson, Á. (2016). Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition*, 153, 196–210.
- Chetverikov, A., Campana, G., & Kristjánsson, A. (2017). Rapid learning of visual ensembles. *Journal of Vision*, 17(2), 21.
- Chetverikov, A., Hansmann-Roth, S., Tanrikulu, Ö. D., & Kristjánsson, Á. (2019). Feature distribution learning (FDL): A new method for studying visual ensembles perception with priming of attention shifts. In *Spatial Learning and Attention Guidance* (pp. 37–57). New York, NY: Humana.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36(1), 28–71.
- Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, 10(4), 360–365.
- De Leeuw, J., & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, 31(3), 1–30.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433.
- Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, 11(5), 14.
- Geng, J. J., DiQuattro, N. E., & Helm, J. (2017). Distractor probability changes the shape of the attentional template. *Journal of*

- Experimental Psychology: Human Perception and Performance*, 43(12), 1993–2007.
- Geng, J. J., & Witkowski, P. (2019). Template-to-distractor distinctiveness regulates visual search efficiency. *Current opinion in psychology*.
- Goldstein, R. R., & Beck, M. R. (2018). Visual search with varying versus consistent attentional templates: Effects on target template establishment, comparison, and guidance. *Journal of Experimental Psychology: Human Perception and Performance*, 44(7), 1086.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778).
- Hout, M. C., Godwin, H. J., Fitzsimmons, G., Robbins, A., Menneer, T., & Goldinger, S. D. (2016). Using multidimensional scaling to quantify similarity in visual search and beyond. *Attention, Perception, & Psychophysics*, 78(1), 3–20.
- Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception, & Psychophysics*, 72(5), 1267–1282.
- Hout, M. C., & Goldinger, S. D. (2012). Incidental learning speeds visual search by lowering response thresholds, not by improving efficiency: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 38(1), 90.
- Hout, M. C., & Goldinger, S. D. (2015). Target templates: The precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics*, 77(1), 128–149.
- Hout, M. C., Papesch, M. H., & Goldinger, S. D. (2013). Multidimensional scaling. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(1), 93–103.
- Jozwik, K. M., Kriegeskorte, N., Storrs, K. R., & Mur, M. (2017). Deep convolutional neural networks outperform feature-based but not categorical models in explaining object similarity judgments. *Frontiers in Psychology*, 8, 1726.
- Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of Neurophysiology*, 97(5), 3155–3164.
- Kristjánsson, Á., & Driver, J. (2008). Priming in visual search: Separating the effects of target repetition, distractor repetition and role-reversal. *Vision Research*, 48(10), 1217–1232.
- Long, B., Störmer, V. S., & Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of Vision*, 17(6), 1–20.
- Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, 10(2), 4.
- Maxfield, J. T., Stalder, W. D., & Zelinsky, G. J. (2014). Effects of target typicality on categorical search. *Journal of Vision*, 14(12), 1.
- Navalpakkam, V., & Itti, L. (2007). Search goal tunes visual features optimally. *Neuron*, 53(4), 605–617.
- Pashler, H. (1987). Target-distractor discriminability in visual search. *Perception & Psychophysics*, 41(4), 285–292.
- Peterson, J. C., Abbott, J. T., & Griffiths, T. L. (2018). Evaluating (and improving) the correspondence between deep neural networks and human representations. *Cognitive Science*, 42(8), 2648–2669.
- Reeder, R. R., & Peelen, M. V. (2013). The contents of the search template for category-level search in natural scenes. *Journal of Vision*, 13(3), 13.
- Rosenholtz, R. (2001). Search asymmetries? What search asymmetries? *Perception & Psychophysics*, 63(3), 476–489.
- Schmidt, J., & Zelinsky, G. J. (2017). Adding details to the attentional template offsets search difficulty: Evidence from contralateral delay activity. *Journal of Experimental Psychology: Human Perception and Performance*, 43(3), 429.
- Schoonveld, W., Shimozaki, S. S., & Eckstein, M. P. (2007). Optimal observer model of single-fixation oddity search predicts a shallow set-size function. *Journal of Vision*, 7(10), 1.
- Scolari, M., & Serences, J. T. (2009). Adaptive allocation of attentional gain. *Journal of Neuroscience*, 29(38), 11933–11942.
- Solman, G. J., & Smilek, D. (2012). Memory benefits during visual search depend on difficulty. *Journal of Cognitive Psychology*, 24(6), 689–702.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136.
- Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, 49(2), 653–673.
- Vickery, T. J., King, L. W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, 5(1), 8.
- Williams, C. C., & Henderson, J. M. (2005). Incidental visual memory for targets and distractors in visual search. *Perception & Psychophysics*, 67(5), 816–827.
- 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.
- Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Cognitive psychology: rare items often missed in visual searches. *Nature*, 435(7041), 439.
- Wolfe, J. M., Klempe, N., & Dahlen, K. (2000). Postattentive vision. *Journal of Experimental Psychology: Human Perception and Performance*, 26(2), 693.
- Won, B. Y., & Geng, J. J. (2018). Learned suppression for multiple distractors in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 44(7), 1128.
- Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23), 8619–8624.
- Yang, H., Chen, X., & Zelinsky, G. J. (2009). A new look at novelty effects: Guiding search away from old distractors. *Attention, Perception, & Psychophysics*, 71(3), 554–564.
- Yu, X., & Geng, J. J. (2019). The attentional template is shifted and asymmetrically sharpened by distractor context. *Journal of Experimental Psychology: Human Perception and Performance*, 45(3), 336–353.
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115(4), 787.

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