# Effects of target typicality on categorical search

**3 authors**, including:

Justin Maxfield
Stony Brook University
**10** PUBLICATIONS   **34** CITATIONS

# Effects of target typicality on categorical search

**Justin T. Maxfield**

Department of Psychology, Stony Brook University, Stony Brook, NY ✉

**Westri D. Stalder**

Department of Psychology, Stony Brook University, Stony Brook, NY ✉

**Gregory J. Zelinsky**

Department of Psychology, Stony Brook University, Stony Brook, NY
Department of Computer Science, Stony Brook University, Stony Brook, NY 🏠 ✉

The role of target typicality in a categorical visual search task was investigated by cueing observers with a target name, followed by a five-item target present/absent search array in which the target images were rated in a pretest to be high, medium, or low in typicality with respect to the basic-level target cue. Contrary to previous work, we found that search guidance was better for high-typicality targets compared to low-typicality targets, as measured by both the proportion of immediate target fixations and the time to fixate the target. Consistent with previous work, we also found an effect of typicality on target verification times, the time between target fixation and the search judgment; as target typicality decreased, verification times increased. To model these typicality effects, we trained Support Vector Machine (SVM) classifiers on the target categories, and tested these on the corresponding specific targets used in the search task. This analysis revealed significant differences in classifier confidence between the high-, medium-, and low-typicality groups, paralleling the behavioral results. Collectively, these findings suggest that target typicality broadly affects both search guidance and verification, and that differences in typicality can be predicted by distance from an SVM classification boundary.

## Introduction

Most of our everyday searches for pens, cups, trash bins, and other common objects, are mediated by categorical representations of the search targets. These categorical search tasks, the search for an object from a target category, are ubiquitous in our day-to-day lives, yet still relatively poorly understood. The vast majority of studies in the search literature have used paradigms in which targets were cued using picture previews, or had targets repeat throughout a block—scenarios resulting in searchers knowing the specific appearance of a given target. Decades of research using these target-specific search paradigms have led to the discovery of many factors affecting the efficiency of search guidance to a target (Wolfe, 1994; Zelinsky, 2008), but it is unclear whether these findings might generalize to categorical search. To the extent that search guidance requires the extraction of specific features from a target cue and the maintenance of these features in visual working memory, guidance may not be possible in categorical search tasks where targets are designated by instruction or word cues and these specific target features are unavailable.

Early work on this topic concluded that categorical search is guided only weakly, or not at all (Vickery, King, & Jiang, 2005; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004). However, more recently Yang and Zelinsky (2009) quantified guidance using eye movement measures and reached a different conclusion: Although targets specified using a picture preview were first fixated more often than categorically designated targets, the proportion of immediate fixations on categorical targets was still well above chance. Other work elaborated upon this relationship between guidance and categorical search, showing that the degree of categorical guidance is proportional to one's knowledge of a target's specific appearance (Malcolm & Henderson, 2009; Schmidt & Zelinsky, 2009). Collectively, this work suggests that categorical search can be guided, and that this guidance improves with the availability of target-identifying features specified by the categorical cue.

Even more recently, Maxfield and Zelinsky (2012) investigated the effects of category hierarchy on categorical search. Targets were designated using either a word cue or a picture preview, and participants were asked to search through an array of objects consisting of a target or a categorical lure among non-targets from random superordinate categories. Among the categorical cue conditions, guidance, again measured as the proportion of trials in which the target was initially fixated, was strongest following a subordinate-level target cue (e.g., "taxi"). However, target verification, measured as the time from target fixation to the search judgment, was fastest for targets cued at the basic level (e.g., "car"), replicating the basic-level superiority effect (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; see also Murphy, 2002) in the context of a search task. This dissociation between search guidance and verification by hierarchical level draws attention to the importance of categories in modulating search behavior, and the need to better understand the factors affecting this modulation.

The present work explores another factor potentially affecting categorical search—how representative or *typical* a target is of its category. Typical exemplars of a category are verified faster and more accurately than atypical exemplars (Murphy & Brownell, 1985; Rosch et al., 1976). These benefits are attributed to typical objects being more likely to share features with the other members of its category, thereby making the task of category verification easier (Murphy & Brownell, 1985; Rips, Shoben, & Smith, 1973; Rosch, 1973, 1975; Tversky & Hemenway, 1984). To the extent that the features of typical objects are representative of those from its category, and given that search guidance improves with the availability of distinctive categorical features, it follows that guidance should be best for targets rated high in typicality.

In the only study to specifically address the role of target typicality in categorical search, Castelhano, Pollatsek, and Cave (2008) varied the typicality of the targets in search arrays relative to basic-level categorical precues. Consistent with the categorization literature (see Murphy, 2002), verification time was found to be shorter for typical targets than for atypical targets. However, the time between onset of the search display and the first fixation on the target did not differ between typical and atypical target conditions. To the extent that target latency is a valid measure of search guidance, this finding suggests that foreknowledge of likely target features is not used to guide categorical search. This finding also potentially points to an important limitation on categorical search—a case in which one search process (target verification) uses information about the target category, but another search process (guidance) does not.

## Experiment 1

Given the potential theoretical importance of their findings, we attempted to replicate the results from Castelhano et al. (2008). We hypothesized that their failure to find a relationship between target typicality and categorical guidance may have been due to their choice of distractors to include in the search displays. Specifically, distractors were computer-generated images of objects created to roughly match the visual features of the target that they accompanied in each display, but not to share non-visual semantic attributes with the target. It may be the case that this attempt to match distractors to targets increased target-distractor visual similarity (Alexander & Zelinsky, 2012), thereby creating lures that would work against finding effects of target typicality on search guidance. Like Castelhano et al. (2008), we used text-label categorical cues and images of common objects rated for typicality, but chose distractors randomly without regard to their visual feature similarity to the target categories. We also explored more stringent measures of search guidance, such as the proportion of immediate looks to the target. If Castelhano and colleagues were correct in asserting that target typicality does not modulate categorical guidance, we would expect to find no effects of typicality on immediate target fixations. However, if effects of target typicality on search guidance are found, this observation would suggest that typicality effects may have been masked by target-distractor similarity effects in the Castelhano et al. (2008) study, and that object typicality is yet another shared factor linking categorization and visual search.

## Methods

### Typicality rating task

Sixteen participants rated images from 42 basic-level categories on a 1 (not at all typical) to 7 (highly typical) scale, and raters were instructed to use consistently a single scale for all of these categories.[1] Raters were also given the option of selecting 0 if they were unsure whether an object was a member of the cued category. If a 0 rating was indicated for a given object by two or more raters, that object was excluded from use as a target in the search experiment. Each norming display depicted an image of an object accompanied by the rating scale shown below the object and a text-label shown above providing the object's basic-level category name. Using this procedure, 12 naturalistic objects from each category, or 504 images in total, were selected from the Hemera Photo-Objects™ image set and normed for typicality. Normed categories were eligible for inclusion in the search experiment if the

| apple[‡] | | | | clock[‡] | | | | pistol[‡] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| axe[‡] | | | | cup | | | | rifle[‡] | | |
| bag[‡] | | | | drill[‡] | | | | saw[‡] | | |
| basket | | | | drum[‡] | | | | scissor[‡] | | |
| bench[‡] | | | | flower | | | | stapler | | |
| bicycle[‡] | | | | grenade | | | | sword[‡] | | |
| bottle | | | | hammer[‡] | | | | table | | |
| cactus | | | | hat | | | | telephone | | |
| cake | | | | helmet | | | | tobacco pipe | | |
| camera[‡] | | | | iron[‡] | | | | trophy | | |
| candle[‡] | | | | lamp[‡] | | | | truck | | |
| chair[‡] | | | | necklace | | | | wrench[‡] | | |

Figure 1. Category names and objects used as targets in Experiment 1. Typicality is ordered from left to right, with the leftmost object from each three-object group being highly typical, followed by the medium- and low-typicality exemplars. All objects were presented in color. [‡]Denotes that this category was used in Experiment 2.

mean typicality of three category exemplars differed from each other by post-hoc $t$ test ($\alpha = 0.01$, uncorrected) following an ANOVA performed on the high-, medium-, and low-typicality groups. Thirty-seven categories met this inclusion criterion, and 36 of these were used in the experiment (Figure 1). The end result of this norming task was the creation of three 36-object groups of search targets varying in their typicality: high ($M = 6.20$, $S.E. = 0.48$), medium ($M = 4.30$, $S.E. = 0.60$), and low ($M = 2.51$. $S.E. = 0.68$), where each group contained one object from each of the 36 categories.

### Search task

*Participants*: Twenty Stony Brook University undergraduates participated for course credit, none of whom were participants in the rating task. All provided informed consent prior to participation (in accordance with the ethical standards stated in the 1964 Declaration of Helsinki), and had normal or corrected-to-normal vision by self-report. Of these, two participants were excluded from the study based on failure to comply with instructions as noted by the experimenter prior to analysis of their data.

*Apparatus and stimuli*: Three exemplars, high-, medium-, and low-typicality, from each of 36 categories
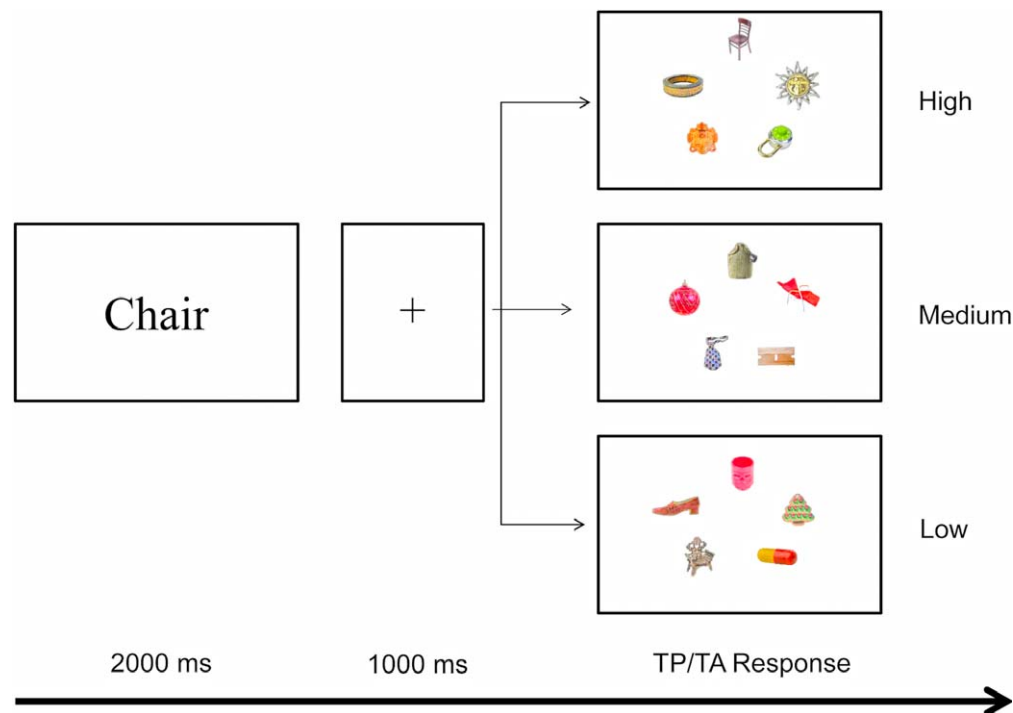
Figure 2. Procedure for the categorical search task used in Experiment 1. Objects are not drawn to scale.

were used as targets in the search experiment. Target cues were the same basic-level category names used during norming. Practice trials consisted of sixteen categories and eight target exemplars not used in the experimental trials. Distractors were 972 objects selected at random and without replacement from the Hemera Photo-Objects™ image set, with the constraint that no distractor was a basic-level exemplar of any of the 36 experimental or 16 practice target categories. No depictions of animals or people were included as targets or distractors.

Eye movements were recorded using an Eyelink 1000 (SR Research) eyetracker, and manual responses were registered using a gamepad controller interfaced through the computer's USB port. Saccades and fixations were defined using the tracker's default settings. Target cues and search arrays were presented on a Dell Trinitron UltraScan P991 monitor at a refresh rate of 100 Hz. Search arrays consisted of five objects, arranged in a circle around a central point corresponding to starting fixation. A target was present in half of the arrays in each 72-trial block. Arrays subtended 19° of visual angle in diameter, and each object was maximally 5° in height and width (average size was 2°–3°). Viewing distance was fixed at 70 cm from the monitor, using a chinrest and headrest, and objects were displayed in color against a white background.

*Design and procedure*: Written and verbal instructions were provided to each participant, followed by a nine-point calibration procedure needed to map eye position to screen coordinates. Calibrations were not accepted until the average and maximum tracking errors were less than 0.45° and 0.9°, respectively. Each trial began with the participant pressing a button on the game pad while fixating a black central dot presented on a white background. This button press caused a cue to appear, written in twenty-four point Times New Roman font, for two seconds, followed by another central fixation cross for one second, and then by presentation of a search array (Figure 2). Searchers indicated target presence or absence by pressing the left or right triggers of the game pad. "Correct" or "Incorrect" appeared at the center of the array following each response, with incorrect responses accompanied by a feedback tone.

There were 16 practice trials and 216 experimental trials. The experimental trials were evenly divided into target present/absent conditions and high-, medium-, and low-typicality conditions, leaving 36 trials per cell of the design. Trials were presented in three counter-balanced blocks, with each block containing half target-present trials that included a target from all 36 categories. Typicality and target presence were randomized over trials within each block, and searchers were allowed a short break following the first two blocks of 72 trials. Calibration was repeated after each break. No target or distractor was repeated throughout the experiment, which lasted approximately 50 minutes.
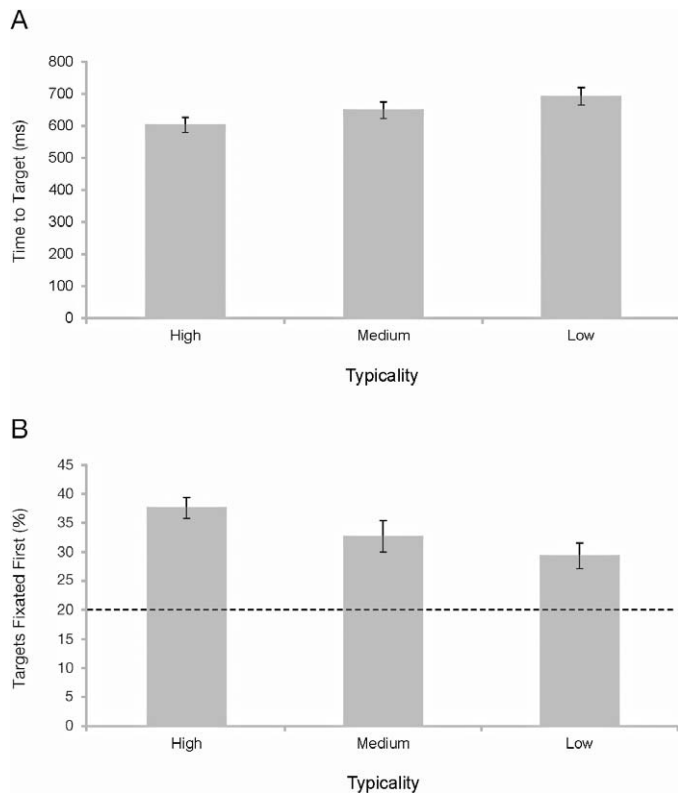
Figure 3. Categorical guidance measures for correct target-present trials from Experiment 1. (A) Mean time to target, and (B) percentage of trials in which the target was the first object fixated. Error bars show one standard error, and the dashed line in (B) indicates chance.

## Results and discussion

Significance is reported for contrasts having $p <$ 0.05, and Tukey's LSD correction was used for all post-hoc pair-wise contrasts following an overall ANOVA. Accuracy rates differed significantly between each typicality condition in target-present trials, $F(2, 34) =$ 128.08, $p < 0.001$, $\eta^2 = 0.883$. High-typicality trials ($M = 94.4\%$, $SD = 4.7$) resulted in more accurate responses than medium-typicality trials ($M = 81.2\%$, $SD = 8.4$), and both were more accurate than low-typicality trials ($M = 63.9\%$, $SD = 9.6$). Consistent with the categorization literature (Murphy & Brownell, 1985; Murphy, 2002), this result suggests that categorization difficulty increased with decreasing target typicality in our task. More interestingly, and given the fact that these targets were considered exemplars of the cued class (as indicated by the norming task), this finding suggests that searchers were using target features that were typical of the target categories—the mismatches between these features and the low-typicality targets were apparently so great that searchers often opted to make target-absent responses, producing false negative errors. Target-absent trials were relatively accurate ($M$

$= 94.3\%$, $SD = 3.4$), with this level of accuracy differing from the medium- and low-typicality target-present conditions, $t(17) \geq 6.93$, $p < 0.001$), but not for contrasts including the high-typicality condition. Subsequent analyses include data only from correct target-present trials.

Following previous work, we segregated our analysis of search performance into epochs of search guidance and target verification (Castelhano et al., 2008; Maxfield & Zelinsky, 2012; Schmidt & Zelinsky, 2009). Two measures of search guidance were used: the time from the onset of the search display until the participant fixated the target (time-to-target) and the percentage of trials in which the target was the first object fixated, a more conservative measure that captures early search guidance. Target verification was defined as the time from first fixation on the target until the button response indicating the target-present judgment. Supplementing this measure, we also calculated total dwell time on the target so as to exclude those rare cases when participants made their search decision after shifting gaze away from the target.

Contrary to Castelhano et al. (2008), who reported no reliable effects of target typicality on search guidance, we found significant differences in time-to-target between our three typicality conditions, $F(2, 34) = 16.86$, $p < 0.001$, $\eta^2 = 0.498$. As shown in Figure 3A, high-typicality targets were fixated sooner than medium-typicality targets ($p < 0.01$), which were fixated sooner than low-typicality targets ($p < 0.001$). A converging pattern is shown in Figure 3B for first-fixated objects, $F(2, 34) = 6.84$, $p < 0.005$, $\eta^2 = 0.287$; high-typicality targets were fixated first more often than medium- ($p < 0.05$) and low-typicality targets ($p < 0.001$), although the difference between the medium- and low-typicality conditions failed to reach significance ($p = 0.23$). Our failure to find a reliable difference in first-fixated targets between the medium- and low-typicality groups can likely be attributed to the high errors rates observed for low-typicality targets. Because highly atypical targets would be most likely to result in false negative errors, and because these error trials were excluded from the guidance analysis, the average typicality of the remaining low-typicality objects would be artificially inflated, shrinking the difference between the medium- and low-typicality conditions. However, consistent with previous work (Castelhano et al., 2008) we did find the expected significant effect of target typicality on verification times, $F(2, 34) = 51.09$, $p < 0.001$, $\eta^2 = 0.75$. As shown in Figure 4, high-typicality targets were verified faster than medium-typicality targets ($p < 0.001$), which were verified faster than low-typicality targets ($p < 0.001$). Target dwell times showed a similar pattern, $F(2, 34) = 112.43$, $p < 0.001$, $\eta^2 = 0.869$, suggesting that these differences reflect actual target verification effects, and not search
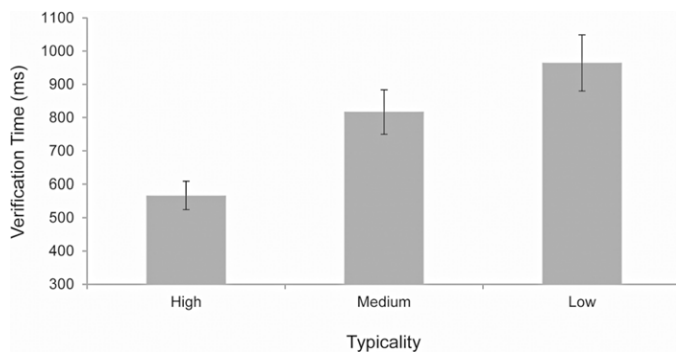
Figure 4. Mean target verification times for the basic-level categorical cues in Experiment 1. Error bars show one standard error.

decisions related to the distractors fixated after leaving the target.

In summary, we found that target typicality affects both guidance and verification during categorical visual search: The more typical a target is of the cued category, the faster this target will be found and verified. The observed relationship between target typicality and verification times dovetails nicely with the many studies in the categorization literature that used a category verification task and found a similar result (see Murphy, 2002), suggesting that these observations may share a common theoretical explanation. But of the two findings reported here, the effect of typicality on guidance is the more theoretically important and controversial. This finding is generally consistent with previous studies showing that search guidance is proportional to the specificity of the categorical cue (Malcolm & Henderson, 2009; Maxfield & Zelinsky, 2012; Schmidt & Zelinsky, 2009), but is inconsistent with the null result reported by Castelhano et al. (2008), who concluded that the effect of typicality on search is limited to target verification. We believe that this discrepancy can be explained by the fact that the distractors in the Castelhano et al. study were selected to share visual features with the targets, and that this introduced a high degree of target-distractor similarity in their task. Because target-distractor similarity relationships between real-world objects can profoundly affect guidance, both for target specific search (Alexander & Zelinsky, 2012) as well as categorical search (Alexander & Zelinsky, 2011), it may be that guidance to target-similar distractors in the Castelhano et al. study masked more subtle effects of target typicality. By using categorically distinct but otherwise random objects as distractors, these similarity relationships would have been far weaker in our study, enabling us to find the reported relationship between typicality and search guidance.

# Experiment 2

Experiment 1 helped to clarify the categorical search literature by showing a relationship between search guidance and target typicality that an earlier study had argued did not exist (Castelhano et al., 2008), but our work also produced a result raising a theoretically important question—how might typical target features, or more broadly, categorical features, be used to guide search? Historically, the search literature has not engaged this question, instead developing theories aimed at explaining search guidance when the exact features of a target are known, usually as a result of seeing a picture cue of the target prior to search (e.g., Zelinsky, 2008). However, as the evidence for categorical guidance mounts (Alexander & Zelinsky, 2011; Maxfield & Zelinsky, 2012; Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009), efforts are underway to extend search theory to include categorical targets. It has been suggested that categorical search can be modeled by adopting techniques from computer vision to find the visual features that best discriminate a target class from non-targets, and using these features as a target template much like the target-specific feature templates that are thought to be created and used to guide search following exposure to a target preview (Alexander & Zelinsky, 2011). Following this suggestion, in recent work target/non-target classifiers were trained, and the distance between an object and this classification boundary was used to estimate the categorical visual similarity between that object and the target class—with a larger distance indicating greater classifier confidence (Zelinsky, Peng, Berg, & Samaras, 2013; Zelinsky, Peng, & Samaras, 2013). Distance from a classification boundary has even been used to create a prioritized map of target evidence—a *categorical target map* (Zelinsky, Adeli, Peng, & Samaras, 2013), which was then used to model the guidance of eye movements to categorical targets (following Zelinsky, 2008).

Experiment 2 was conducted to determine whether this theoretical framework might also extend to target typicality effects—can the effects of target typicality observed in Experiment 1 be explained in terms of distance from a classification boundary? To answer this question we adopted techniques from computer vision, which were needed in order to accommodate the images of real-world objects used in the behavioral work. Classifiers trained on a subset of the target categories from Experiment 1 were used to obtain confidence estimates for our high-, medium-, and low-typicality target exemplars. To the extent that these confidence estimates are higher for high-typicality targets than for low-typicality targets, this difference would demonstrate that the feature overlap relationships presumed to underlie typicality effects can be made computa-

tionally explicit for the types of visually complex object categories that populate our everyday experience.

## Methods

Computer vision is a mature field, with many features and learning methods at its disposal (Everingham, Van Gool, Williams, Winn, & Zisserman, 2010). Because our goal was to assess the general feasibility of applying these tools to our question, and not to choose methods with the goal of tailoring results to best fit our behavioral data, for this experiment we simply chose the most widely used feature and learning method in this literature—the Scale Invariant Feature Transform (SIFT) and the linear-kernel Support Vector Machine (SVM). The SIFT feature represents the structure of orientation gradients in local image patches using 16 spatially-distributed histograms of scaled and normalized edge energy (Lowe, 1999, 2004). It is this histogram representation that endows SIFT features with rotation and scale invariance, factors that fueled the popularity of this feature. SVM is a powerful and elegant method for learning a classification boundary to separate exemplars of a target category (positive samples) from non-target exemplars (negative samples). In computer vision this learning is typically done using sets of training images that are different from those used for testing, with feature descriptors applied to these training images and then clustered and vector-quantized into a "bag of words" (Csurka, Dance, Fan, Willamowski, & Bray, 2004).

In the present experiment, training images were selected from the ImageNet database (http://www.image-net.org/), which contains hundreds of exemplars of thousands of object categories. Of the 36 basic-level target categories used in Experiment 1, 19 of these corresponded to nodes in the ImageNet hierarchy (Figure 1). Our investigation was therefore limited to this subset of 19 target categories. Additionally, 10 non-target categories were selected at random from ImageNet, with the images in these categories used as negative samples for training.[2] In total, there were 19,000 images of targets used as positive training samples (1,000/category) and 10,000 images of non-targets used as negative training samples (1,000/category). We used these positive and negative samples and a publically available version of SVM (Chang & Lin, 2001) to train 19 target/non-target classifiers, one for each target category. This training was done using the SIFT features and bag-of-words representation (Lazebnik, Schmid, & Ponce, 2006) made available by ImageNet, quantized using a code book of 1000 visual words obtained by performing k-means clustering on a random sampling of SIFT features extracted from images in the ImageNet database (for additional details, see http://www.image-net.org/download-features). Testing consisted of using the VLFeat implementation of dense SIFT (Vedaldi & Fulkerson, 2008) to extract features from the high-, medium-, and low-typicality exemplars for each of the selected 19 target categories from Experiment 1 (57 target images in total), using the same vocabulary of 1000 visual words from the training bag-of-words representation. Distances were then found between the 57 target images and the 19 corresponding SVM classification boundaries, where larger distances again indicate greater confidence in the classification. These distances were finally converted to probabilities using a probability estimation method (Platt, 2000), with these probabilities being used as the classifier confidence estimates reported in the results.

## Results and discussion

Our goal was to assess the plausibility of using methods from computer vision to capture the target typicality ratings from our behavioral participants. Object typicality is commonly believed to decrease with increasing distance between the features of an object exemplar and the features that define that object's category (Murphy & Brownell, 1985; Rips et al., 1973; Rosch, 1973, 1975; Tversky & Hemenway, 1984); the larger this distance, the less overlap between these features and the more atypical the exemplar. To the extent that this relationship is also true for purely visual features, and in particular the visual features for the common object categories used in Experiment 1, it may be possible to use an exemplar's distance from an SVM classification boundary to predict its typicality (see also Zelinsky, Peng, Berg, et al., 2013; Zelinsky, Peng, & Samaras, 2013). We conducted this analysis and found significant differences between the mean classifier confidence values for target exemplars in the three typicality conditions, $F(2, 54) = 6.02$, $p < 0.005$, $\eta^2 = 0.182$); the more typical a target was rated of its category, the more confident its classification as a target ($M_L = 0.25$, $SD = 0.20$, $M_M = 0.37$, $SD = 0.30$, and $M_H = 0.53$, $SD = 0.25$ in the low-, medium-, and high-typicality conditions, respectively). This finding suggests that classifier confidence derived from purely visual features (SIFT) and straightforward methods from computer vision (SVM with bag-of-words) may be a reasonable predictor of behavioral typicality judgments.

## General discussion

This study explored the relationship between object typicality and categorical search from a behavioral and

computational perspective. In Experiment 1 we replicated the effect of target typicality on verification times previously reported by Castelhano et al. (2008), but failed to replicate their null result showing no effect of typicality on search guidance. We found that increasing target typicality not only resulted in the faster verification of search targets, it also resulted in their preferential fixation. In Experiment 2 we trained computer vision classifiers on a subset of these target categories and showed that classifier confidence predicted the low-, medium-, and high-typicality judgments from behavioral raters. Both of these experiments have important implications for search theory.

The behavioral finding that target typicality affects search guidance adds to the rapidly accumulating body of knowledge about categorical search. Whereas categorical search was once believed to be unguided (Vickery et al., 2005; Wolfe et al., 2004), we now know that categorical guidance *does* exist (Yang & Zelinsky, 2009). We also know that this guidance is proportional to the degree of target-specifying information in the categorical cue (Schmidt & Zelinsky, 2009), that guidance depends on the categorical level at which the target is specified (Maxfield & Zelinsky, 2012), and, with the present work, that categorical guidance is modulated by target typicality. Guidance even exists to objects that are just visually similar to the categorical target (Alexander & Zelinsky, 2011: Zelinsky, Peng, & Samaras, 2013).

Methods from computer vision are valuable in specifying the information used to mediate these many factors affecting categorical search guidance. Recent work borrowing techniques from computer vision has shown that it is possible to model categorical guidance using purely visual features (Zelinsky, Adeli et al., 2013; Zelinsky, Peng, Berg et al., 2013). The logic is straightforward: To the extent that a classifier learned from computer vision features is able to predict the effect of some factor on categorical search, we know that the influence of that factor can be purely visual because visual information is all that could be captured by these features. This is not to say that guidance from the semantic properties of objects does not exist (Hwang, Wang, & Pomplun, 2011), only that these hypothetical higher-level factors are unnecessary to explain categorical guidance. Extending this reasoning to the current study, an effect of target typicality on search guidance and verification can be mediated by purely visual information extracted from pixels in images. More generally, this modeling work suggests that it is possible to learn the visual features that discriminate a target class from non-targets, and to use these features as a guiding template in much the same way as features actually extracted from a picture preview of the target. This suggestion is important as a reliance on a picture preview is overly restrictive and unrealistic—picture previews simply do not exist in the vast majority of our everyday searches. With the present quantitative demonstration of purely visual categorical guidance, there is now a means to move beyond the picture preview and to start addressing the types of categorical search tasks that occupy our day-to-day lives.

As knowledge about categorical search grows, it is inevitable that connections will be made to the categorization literature, and this is a second contribution of the present work. Previously it was found that target verification in a categorical search task shows a standard basic-level advantage (faster verification following a basic-level categorical cue; Rosch et al., 1976), but that search guidance was strongest for cues specified at the subordinate level (Maxfield & Zelinsky, 2012). The current work strengthens this connection between search and categorization theory by showing that object typicality—a core concept in the categorization literature—also affects how efficiently gaze is guided to a search target. Typicality effects are widely believed to reflect the discriminative features learned from a set of category exemplars, with typicality decreasing with increasing distance from the mean of these features (e.g., Reed, 1972; Rosch, 1975) or the features of the most frequently occurring exemplar (e.g., Medin & Schaffer, 1978; Nosofsky, 1986). We believe that a similar framework might extend to a categorical search task; the features of an atypical target may overlap with those of one or more distractors nearly as much as the features learned for that target category, resulting in the observed decrements in search performance.

The approach adopted in Experiment 2 adds to this intuitive framework a simple premise, that the feature distance between an exemplar and a category can be approximated by the distance from an SVM classification boundary. Previous work has attached behavioral (Zelinsky, Adeli et al., 2013; Zelinsky, Peng, Berg et al., 2013; Zelinsky, Peng, & Samaras, 2013) and neurophysiological (Carlson, Ritchie, Kriegeskorte, Durvasula, & Ma, 2014) significance to these distances (see also Jäkel, Schölkopf, & Wichmann, 2008, for a more general discussion), although such interpretations should be made cautiously; SVM forms boundaries to maximize the classification accuracy of positive and negative training samples, not to capture metrical similarity relationships between those objects. Identical classification boundaries might therefore form despite underlying differences in visual similarity. Nevertheless, we found that greater positive (towards the target category) distances from the classification boundary predict greater target typicality. Conversely, atypical targets, because they are closer to the boundary of the target category, would likely generate weaker guidance

signals and might even be mistakenly categorized as distractors as their distance becomes negative with respect to the target classification boundary.

More generally, we believe this focus on classification provides an opportunity to bridge the behavioral categorization and computer vision literatures. Classification is the assignment of stimuli, or their features, to categories based on learning derived through feedback, and this problem has been approached from very different perspectives. The behavioral categorization literature is rich with formal models of this process: The Generalized Context Model (Nosofsky, 1986), AL-COVE (Kruschke, 1992), and SUSTAIN (Love, Medin, & Gureckis, 2004) to name but a few (for a review, see Kruschke, 2008). Some of these models even form explicit decision bounds to separate the features of Category A from those of Category B (Ashby & Maddox, 1994; Maddox & Ashby, 1996). Classification has also been extensively studied in the computer vision literature (Everingham et al., 2010), with the problem of object detection being the most relevant to the present study.

These differing perspectives each have their strengths and weaknesses. Models in the categorization literature are essentially models of category learning, and great importance is attached to how exposure to each new instance of an exemplar changes slightly the probability that subsequent exemplars will be assigned to a category. In computer vision the more common practice is to learn classifiers from training sets—the emphasis here is on how well a classifier works (i.e., makes correct classifications) rather than on how well the learning of a classifier maps onto human behavior. Studies of perceptual classification also rely overwhelmingly on very simple stimuli having a straightforward dimensional structure (see Ashby & Maddox, 2005, for a review). Whether they are lines of different lengths or oriented Gabors, the psychological dimensions of interest are clear, and this enables a clearer focus on the underlying learning process that these models are designed to explain. Computer vision models, however, are designed for applications in the real world and therefore do not have this luxury—they must be able to work with images or videos of visually complex objects and scenes having unknown dimensional structure or else they are useless.

We showed that a model having questionable psychological validity, one based on a linear SVM classifier learned from SIFT features and labeled training data, can nevertheless predict the typicality ratings of search targets from visually complex real-world categories. Given that this relatively generic model lacked the assumptions of incremental learning central to models from the categorization literature, the implication of this demonstration is that these assumptions are not essential to predicting typicality—

more important is probably the actual classification boundaries that are learned, and not the learning process per se. Following previous work (Zelinsky, Adeli, Peng, & Samaras, 2013), future work will attempt to build from classifier confidence values a map of prioritized target evidence and use it to predict effects of target typicality on the guidance and verification processes expressed in the eye movements made during categorical search.

*Keywords: typicality, visual search, eye movements, categorization classification*

## Footnotes

[1]Of course there is no way of knowing whether this instruction was actually followed, but to the extent that raters adopted different scales for different categories this eventuality would introduce error variance and work against us finding the reported effects of typicality in our guidance and verification measures.

[2]Note that there are multiple ways of selecting negative samples for training a classifier, and how best to do this is still an open question in computer vision (Perronnin, Akata, Harchaoui, & Schmid, 2012; Zhu, Vondrick, Ramanan, & Fowlkes, 2012). It is also true that different compositions of training sets would produce slightly different classifiers, and therefore slightly different results. Two broad options were available to us: to have every negative sample come from a different non-target category (i.e., only one exemplar per category), or to randomly choose a subset of non-target categories and to use as negative samples multiple exemplars of each. We adopted the latter

method so as to make the non-target categories more comparable to the target categories (each category consisted of 1000 exemplars). Importantly, no effort was made to choose negative samples so as to fit our behavioral data.

# References

Alexander, R. G., & Zelinsky, G. J. (2011). Visual similarity effects in categorical search. *Journal of Vision, 11*(8):9, 1–15, http://www.journalofvision. org/content/11/8/9, doi:10.1167/11.8.9. [PubMed] [Article]

Alexander, R. G., & Zelinsky, G. J. (2012). Effects of part-based similarity on visual search: The Frankenbear experiment. *Vision Research, 54,* 20–30.

Ashby, F. G., & Maddox, W. T. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology, 38,* 423–466.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review Of Psychology, 56,* 149–178.

Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2014). Reaction time for object categorization is predicted by representational distance. *Journal of Cognitive Neuroscience, 26*(1), 132–142.

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

Chang, C., & Lin, C. (2001). LIBSVM: Alibrary for SVMs. Software available at http://www.csie.ntu. edu.tw/~cjlin/libsvm

Csurka, G., Dance, C., Fan, L., Williamowsky, J., & Bray, C. (2004). Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, European Conference on Computer Vision (ECCV),* pp. 1–22.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The PASCAL visual objects classes (VOC) challenge. *International Journal of Computer Vision, 88*(2), 303–338.

Hwang, A. D., Wang, H.C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research, 51,* 1192–1205.

Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2008). Generalization and similarity in exemplar models

of categorization: Insights from machine learning. *Psychonomic Bulletin & Review, 15*(2), 256–271.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*(1), 22–44.

Kruschke, J. K. (2008). Models of categorization. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 267–301). New York: Cambridge University Press.

Lazebnik, S., Schmid, C., & Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2,* 2169–2178.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*(2), 309–332.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Proceedings of the International Conference on Computer Vision, 2,* 1150–1157.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91–110.

Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 795–817.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision, 9*(11):8, 1–13, http://www. journalofvision.org/content/9/11/8, doi:10.1167/9. 11.8. [PubMed] [Article]

Maxfield, J. M., & Zelinsky, G. J. (2012). Searching through the hierarchy: How level of target categorization affects visual search. *Visual Cognition, 20*(10), 1153–1163.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

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

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 70–84.

Nosofsky, R. M. (1986). Attention, similarity, and the

identification categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Perronnin, F., Akata, Z., Harchaoui, Z., & Schmid, C. (2012). Towards good practice in large-scale learning for image classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 1–8.

Platt, J. C. (2000). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans (Eds.), *Advances in large margin classifiers*. MIT Press, Cambridge, MA.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3,* 382–407.

Rips, L., Shoben, E., & Smith, E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12,* 1–20.

Rosch, E. (1973). Natural categories. *Cognitive Psychology, 4,* 328–350.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192–233.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *The Quarterly Journal of Experimental Psychology, 62,* 1904–1914.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113,* 169–193.

Vedaldi, A., & Fulkerson, B. (2008). VLFeat: An open and portable library of computer vision algorithms. Retrieved from http://www.vlfeat.org

Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision, 5*(1):8, 81–92, http://www.journalofvision.org/content/5/1/8, doi:10.1167/5.1.8. [PubMed] [Article]

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

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,* 1411–1426.

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research, 49,* 2095–2103.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115,* 787–835.

Zelinsky, G. J., Adeli, H., Peng, C., & Samaras, D. (2013). Modeling eye movements in a categorical search task. *Philosophical Transactions of the Royal Society B, 368*(1628), 1–12.

Zelinsky, G. J., Peng, Y., Berg, A. C., & Samaras, D. (2013). Modeling guidance and recognition in categorical search: Bridging human and computer object detection. *Journal of Vision, 13*(3):30, 1–20, http://www.journalofvision.org/content/13/3/30, doi:10.1167/13.3.30. [PubMed] [Article]

Zelinsky, G. J., Peng, Y., & Samaras, D. (2013). Eye can read your mind: Decoding gaze fixations to reveal categorical search targets. *Journal of Vision, 13*(14):10, 1–13, http://www.journalofvision.org/content/13/14/10, doi:10.1167/13.14.10. [PubMed] [Article]

Zhu, X., Vondrick, C., Ramanan, D., & Fowlkes, C. (2012). Do we need more training data or better models for object detection? *British Machine Vision Conference (BMVC),* 1–11.