Published in final edited form as:
J Vis. ; 11(8): . doi:10.1167/11.8.9.

Visual Similarity Effects in Categorical Search

Robert G. Alexander[1], **Gregory J. Zelinsky**[1,2]

[1]Department of Psychology, Stony Brook University, USA

[2]Department of Computer Science, Stony Brook University, USA

Abstract

We asked how visual similarity relationships affect search guidance to categorically-defined targets (no visual preview). Experiment 1 used a web-based task to collect visual similarity rankings between two target categories, teddy bears and butterflies, and random-category objects, from which we created search displays in Experiment 2 having either high-similarity distractors, low-similarity distractors, or "mixed" displays with high, medium, and low-similarity distractors. Analysis of target-absent trials revealed faster manual responses and fewer fixated distractors on low-similarity displays compared to high. On mixed displays, first fixations were more frequent on high-similarity distractors (bear=49%; butterfly=58%) than on low-similarity distractors (bear=9%; butterfly=12%). Experiment 3 used the same high/low/mixed conditions, but now these conditions were created using similarity estimates from a computer vision model that ranked objects in terms of color, texture, and shape similarity. The same patterns were found, suggesting that categorical search can indeed be guided by purely visual similarity. Experiment 4 compared cases where the model and human rankings differed and when they agreed. We found that similarity effects were best predicted by cases where the two sets of rankings agreed, suggesting that both human visual similarity rankings and the computer vision model captured features important for guiding search to categorical targets.

Keywords

Visual search; eye movements; categorical guidance; visual similarity; computer vision

You have probably had the experience of searching through a crowded parking lot and locating several other vehicles of the same color or model before finally finding your car. This is an example of visual similarity affecting search; the presence of these target-similar distractors made it harder to find the actual thing that you were looking for.

Such visual similarity effects have been extensively studied in the context of search, with the main finding from this effort being that search is slower when distractors are similar to the target (e.g., Duncan & Humphreys, 1989; Treisman, 1991). Models of search have also relied extensively on these visual similarity relationships (e.g., Hwang, Higgins, & Pomplun, 2009; Treisman & Sato, 1990; Wolfe, 1994; Zelinsky, 2008). Despite their many differences,

Address correspondence to: Gregory J. Zelinsky, Psychology B-240, Stony Brook University, Stony Brook, NY 11794-2500, Phone: 631-689-7087, Gregory.Zelinsky@stonybrook.edu.

all of these models posit a very similar process for how similarity relationships are computed and used; the target and scene are represented by visual features (color, orientation, etc.), which are compared to generate a signal used to guide search to the target and to target-like distractors in the scene. In general, the more similar an object is to the target, the more likely that object will be fixated (see also Eckstein, Beautter, Pham, Shimozaki, & Stone, 2007; Findlay, 1997; Tavassoli, van der Linde, Bovik, & Cormack, 2009; Zelinsky, 2008).

All of these models, however, assume knowledge of the target's specific appearance in the creation of this guidance signal. This assumption is problematic, as it is often violated in the real world. Descriptions of search targets are often incomplete and lacking in visual detail; exact knowledge of a target's appearance is an artificial situation that typically exists only in the laboratory. Particularly interesting are cases in which a target is defined categorically, as from a text label or an instruction (i.e., no picture preview of the target). Given the high degree of variability inherent in most categories of common objects, search under these conditions might have limited visual information about a target that could be confidently compared to a scene to generate a guidance signal. Indeed, a debate exists over whether categorical search is guided at all, with some labs finding that it is (Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009) and others suggesting that it is not (e.g., Castelhano, Pollatsek, & Cave, 2008; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004; see also Vickory, King, & Jiang, 2005).

In the present study we ask not only whether categorical search is guided, but also whether categorical guidance to realistic targets is affected by target-distractor visual similarity. Guidance from a pictorial preview is known to decrease with increasing visual similarity between a target and distractors; does this same relationship hold for categorically-defined targets? It may be the case that categorical target descriptions are dominated by non-visual features, such as semantic or functional properties of the target category. [1] There is an ongoing debate in the literature as to whether eye movements can be guided by semantic information, with some researchers reporting guidance for even very early eye movements (Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Becker, Pashler, & Lubin, 2007; Underwood, Templeman, Lamming, & Foulsham, 2008; Bonitz & Gordon, 2008; Rayner, Castelhano, & Yang, 2009), and others showing that early eye movements are not guided by semantic information (De Graaf, Christiaens, & d'Ydewalle, 1990; Henderson, Weeks, & Hollingworth, 1999; Võ & Henderson, 2009). If semantic factors either cannot affect early eye movements, or can do so only weakly, and categorical search relies on these factors, then guidance to these targets may be weak or even nonexistent, potentially explaining why some researchers have found evidence for categorical guidance and others have not. To the extent that categorical search does use non-visual features, effects of target-distractor visual similarity would not be expected. However, if target categories are represented visually, one might expect the same visual target-distractor similarity relationships demonstrated for target-specific search to extend to categorical search (see Duncan, 1983, for a similar question applied to simple stimuli).

[1] Although the semantic properties of a search object must ultimately be accessed via visual features, we distinguish between visual and non-visual features to acknowledge the possibility that the type of information used to guide search might be either visual or semantic.

It is unclear how best to manipulate visual similarity in the context of categorical search. Traditional methods of manipulating target-distractor similarity by varying only a single target feature are clearly suboptimal, as realistic objects are composed of many features and it is impossible to know *a priori* which are the most important. This problem is compounded by the categorical nature of the task; the relevance of a particular target feature would almost certainly depend on the specific category of distractor to which it is compared. It is not even known how best to derive specific target features for such a comparison; should features be extracted from a particular exemplar that is representative of the target category or should an average be obtained from many target exemplars (see Levin, Takarae, Miner, & Keil, 2001, and Yang & Zelinsky, 2009)? And even if the relevant feature dimensions were known, the similarity metric used within a feature dimension may itself be categorical (Wolfe, Friedman-Hill, Stewart, & O'Connell, 1992), and therefore largely unknown.

In light of the difficulties associated with directly manipulating the specific features underlying visual similarity, we opted in Experiment 1 for a more pragmatic and holistic approach—to use ratings of visual similarity collected from subjects. [2] Using these estimates of visual similarity, Experiment 2 asked whether the visual similarity relationships known to affect search for specific targets also extends to categorical search. Previous arguments for the use of visual features to guide categorical search appealed to evidence showing the preferential direction of initial saccades to categorical targets (Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009). However, although such an early expression of guidance makes an influence of semantic target-distractor similarity less likely, such non-visual contributions to this behavior cannot be ruled out completely. More compelling evidence for the visual direction of categorical search would be the demonstration of an effect of target-distractor visual similarity on categorical guidance; providing this evidence was the primary goal of Experiment 2. Experiment 3 replicated Experiment 2 using search displays assembled from similarity estimates obtained from a computer vision model (rather than from behavioral ratings). [3] We did this in order to guarantee the use of purely visual features in any observed relationship between target-distractor similarity and categorical search guidance. Finally, in Experiment 4 we explored cases in which the behavioral similarity estimates and the computer vision similarity estimates agreed or disagreed. We did this in hopes of learning whether these different similarity measures use different features to guide gaze in a categorical search task.

Experiment 1: Web-based Similarity Rankings

The goal of Experiment 1 was to obtain visual similarity estimates between random real-world objects and the "teddy bear" and "butterfly" categories, for the purpose of using these estimates to select distractors in further search experiments. A web-based task was used to collect these visual similarity estimates, due to the relatively large number of subjects that we anticipated needing to obtain stable similarity estimates between random objects and these target categories.

[2] Aspects of Experiment 1 were presented at the 2008 meeting of the Cognitive Science Society (Zhang, Samaras, and Zelinsky, 2008).
[3] Aspects of Experiments 2 and 3 were presented at the 2010 meeting of the Cognitive Science Society (Alexander, Zhang, & Zelinsky, 2010).

Method

Participants.—One hundred and forty two students from Stony Brook University participated in exchange for course credit.

Stimuli.—The two target categories were teddy bears, images obtained from Cockrill (2001), and butterflies (including some moths), images obtained from the Hemera collection (Hemera® Photo-objects). Similarity ratings were collected for 2000 non-target objects representing a broad range of categories. These images were also selected from the Hemera collection.

Procedure.—Upon following a link to the experiment, subjects were randomly assigned to either a butterfly target category or a teddy bear target category, and then participated in a training phase and a ranking phase of the experiment. During training, subjects were shown 200 example images from their assigned target category (either teddy bear or butterfly). This was done to familiarize subjects with the types of objects that constituted the target category, and with the feature variability among these objects. In the ranking phase, subjects were shown groups of five non-target objects randomly selected from the 2000 object set, and asked to rank order these five objects from most visually similar (5) to least visually similar (1) relative to the target category. Figure 1 shows a screenshot of the ranking phase for one representative teddy bear trial. Each subject completed 100 ranking trials. Given our use of random objects, the task of rank ordering the objects was preferable to the task of assigning an independent similarity score to each object, as the latter task would likely have resulted in a large number of "very dissimilar" responses. Importantly, subjects were instructed to use only visual similarity and to disregard categorical or associative relationships between the objects and the target category when making their judgments.

Results and Discussion

Subjects produced a total of 71,000 butterfly and teddy bear similarity estimates for 2,000 different objects. The rankings for each object varied substantially between subjects (see Figure 2). Rankings for the highest level of similarity (rank 5) were the most consistent for both teddy bears and butterflies, followed by the rankings for the least target-like objects (rank 1). Subjects were much more likely to agree on extreme similarity rankings than on intermediate ones. In addition, subject rankings were more consistent for teddy bears than for butterflies. This difference in consistency between the two target categories might be due to teddy bears having a more prototypical color than butterflies (many were brown), or to the butterfly object class being in general more variable. Figure 2 also shows that the adoption of stricter criteria for subject agreement resulted in fewer consistently ranked objects. To ensure that sufficient stimuli would be available to assemble trials in our search tasks, we adopted an inter-subject consistency of 60% or more when selecting high-similarity and low-similarity objects for use in Experiment 2.

Experiment 2: Searching Through Human-Ranked Distractors

In Experiment 2 we selected ranked objects from Experiment 1 (referred to as "distractors" throughout the rest of the paper) and placed these into search displays in order to test whether the visual similarity relationships known to affect previewed search also affect categorical search. While previous work has shown that subjects preferentially fixate the targets in categorical search (Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009), a demonstration that subjects also preferentially fixate target-similar distractors would indicate that subjects preattentively guide their eye movements to the features of the target category. If subjects' eye movements are preferentially directed to visually target-similar items, this would provide evidence that this categorical guidance is due to visual factors, rather than semantic factors or other non-visual information.

We were also interested in determining whether explicit visual similarity judgments are predictive of effects of target-distractor visual similarity on categorical search. Search guidance is a largely implicit process, and can be expressed in even the first search saccade (e.g., Chen & Zelinsky, 2006); the task of assigning rankings to objects in a web experiment is comparatively slow and far more explicit. Do these two tasks use fundamentally different sources of information, or can visual similarity estimates obtained from explicit judgments be useful in describing guidance during search? Answering this question was a secondary goal of this experiment.

If categorical search is guided by target-distractor visual similarity, and if this relationship can be captured by explicit similarity judgments, we would expect a relatively high proportion of initial saccades to high-similarity distractors, and relatively few initial saccades to low-similarity distractors. However, if categorical guidance is mediated by non-visual factors, or if the visual similarity estimates obtained from an explicit task cannot be extended to search, we would expect no effect of our similarity manipulations on overt search guidance or manual search efficiency.

Method

Participants.—Twenty-four students from Stony Brook University participated in exchange for course credit, none of whom participated in Experiment 1. All subjects reported normal or corrected to normal vision.

Stimuli and apparatus.—Gaze position was recorded using an SR Research EyeLink® II eye tracking system. This eye tracker is video-based and has a sampling rate of 500 Hz and a spatial resolution of ~0.2º. Target present/absent search decisions were made using a GamePad controller connected to a USB port. Head position and viewing distance were fixed at 72 cm from the screen with a chin rest. Search arrays were displayed on a flat-screen CRT monitor at a resolution of 1024 × 768 pixels (subtending 28º × 21º) using a refresh rate of 85 Hz.

The two target categories were again teddy bears and butterflies, the same images shown to subjects during the training phase in Experiment 1. The distractors were also selected from the pool of objects ranked in Experiment 1 based on their visual similarity estimates to the two target categories. The 1–5 ranking was used to assign distractors to three different similarity levels per target category (teddy bear or butterfly). Distractors with a consistent ranking of "1" were considered "low-similarity", and distractors with a consistent ranking of "5" were considered "high-similarity". See Figure 3 for representative examples of teddy bear and butterfly targets, as well as objects rated as being low and high in visual similarity to these two target categories. Consistency for both low-similarity and high-similarity distractors was based on a 60% level of inter-subject agreement. Due to the relatively low level of inter-subject agreement for objects given rankings 2–4 (see Figure 2), we refer to these objects and objects having less than 60% inter-subject ranking agreement as “medium similarity” throughout the remainder of the paper. Objects were normalized for size, with the mean object size subtending ~2.8º of visual angle.

Procedure.—Half of the subjects searched for a teddy bear target, the other half searched for a butterfly target. This search was categorical; subjects were not shown a specific bear or butterfly target preview prior to each search trial. Rather, subjects were told the target category at the start of the experiment. They were also shown examples of the target category, none of which were used as actual targets in the experimental trials.

Each trial began with the subject fixating a central dot and pressing a button on the controller to initiate the search display. The search display consisted of six evenly-spaced objects arranged on an imaginary circle with a radius of 300 pixels (8.4º) relative to the center of the screen. On target present trials (50%), one object was either a bear or a butterfly, depending on the condition, and the other five objects were randomly selected distractors. On target absent trials (50%), distractors were selected based on the similarity rankings from Experiment 1. Each object was repeated only once throughout the experiment, and was never repeated in the identical context (i.e., a repeated target appeared with different distractors).

There were three target absent conditions: 8 high-similarity trials (all distractors were high-similarity items, with respect to the target category), 8 low-similarity trials (all distractors were low-similarity items, with respect to the target category), and 24 “mixed” trials, where two distracters were selected from the high-similarity category, two from the low-similarity category, and two from the medium similarity category. The high and low similarity conditions were included to determine whether visual similarity affects search accuracy and manual reaction times (RTs). The mixed condition allowed us to directly examine whether overt search was guided differentially to distractors depending on their similarity to the target category.

Target presence/absence and similarity condition were within-subject variables, and both were randomly interleaved throughout the experiment. Subjects were asked to make their present/absent judgments as quickly as possible while maintaining accuracy. Accuracy feedback was provided following each response.

Results and Discussion

As the similarity manipulation was limited to the target absent trials, analyses were restricted to these data. Errors were less than 6% in all conditions, and were excluded from all subsequent analyses. This low false positive rate means that subjects did not confuse the high-similarity distractors for targets (e.g., a stuffed bunny distractor was not mistakenly recognized as a teddy bear).

RTs were longest in the high-similarity condition and shortest in the low-similarity condition, with the mixed condition yielding intermediate RTs (Table 1). These effects of similarity were significant for both butterfly targets (*F*(2,22) = 46.87, *p* < .001) and for bear targets (*F*(2,22) = 53.85, *p* < .001). Post-hoc t-tests with Bonferroni correction showed slower RTs in the high-similarity condition relative to the mixed condition (*p* < .01 for both teddy bears and butterflies) and faster RTs in the low-similarity condition relative to the mixed condition (*p* < .001 for both teddy bears and butterflies).

The number of distractors fixated during target absent search also differed between the similarity conditions, and this again occurred for both butterfly (*F*(2,22) = 30.41, *p* < .001) and bear targets (*F*(2,22) = 59.55, *p* < .001). More distractors were fixated on high-similarity trials (3.16 ± 0.23 for bears, 2.50 ± 0.36 for butterflies) compared to either mixed trials (2.53 ± 0.24 for bears, *p* < .01; 1.83 ± 0.31 for butterflies, *p* < .001) or low-similarity trials (1.51 ± 0.23 for bears, *p* < .001; 1.29 ± 0.26 for butterflies, *p* < .001), and more distractors were fixated on mixed trials than on low-similarity trials (*p* < .001 for bears and *p* < .01 for butterflies). As distractor similarity to the target increased, so did the number of fixations on these distractors. All of these patterns are consistent with the suggestion that visual similarity rankings are predictive of search efficiency.

One of the most conservative measures of search guidance is the first fixated object—the object looked at first following search display onset. Analysis of first object fixations on the mixed condition trials revealed significant effects of our similarity manipulation for both the teddy bear (*F*(2,22) = 30.15, *p* < .001) and butterfly (*F*(2,22) = 10.13, *p* < .01) search tasks. Consistent with the RT analyses we found that distractor similarity to the target category determined the type of distractor that was first fixated (Figure 4A). High-similarity distractors were more often fixated first compared to medium-similarity distractors (*p* < .01 for bears and *p* = .05 for butterflies), which were more often fixated first compared to low-similarity distractors (*p* < .01 for bears and butterflies). Moreover, first fixations on high-similarity distractors were well above chance (*t*(11) = 4.70, *p* < .01 for bears; *t*(11) = 7.04, *p* < .001 for butterflies), and first fixations on low-similarity distractors were well below chance (*t*(11) = 18.89, *p*< .001 for bears; *t*(11) = 11.90, *p*<.001 for butterflies), indicating that initial saccades were guided towards target-similar distractors and away from target-dissimilar distractors. We also analyzed the latencies of these initial saccades to see whether these patterns could be attributed to speed-accuracy tradeoffs, but none were found; initial saccade latencies did not reliably differ between the similarity conditions for either butterfly (*F*(2,22) = 1.51, *p* = 0.24) or bear targets (*F*(2,22) = 0.41, *p* = 0.65). The observed effects of visual similarity reflect actual changes in search guidance.

Two conclusions follow from these data. First, categorical search guidance is affected by target-distractor visual similarity. As the visual similarity between a distractor and a target category increased, search efficiency decreased. This decreased efficiency is due to distractors becoming more distracting, as evidenced by an increase in the number of first fixations on the high similarity distractors. More generally, this finding adds to the growing body of evidence suggesting that categorical search is indeed guided (Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009), a question that had been the topic of debate (Castelhano et al., 2008; Wolfe et al., 2004). Not only is categorical search guided, it is guided by matching visual features of the search objects to a visual representation of the target category.

The second conclusion following from these data is that explicit visual similarity rankings obtained from a web task are highly predictive of categorical search. Given the dramatic differences between these tasks, this finding is surprising. Judgments in the web task were highly deliberative. In piloting, a subject was observed agonizing over whether a wooden box or a backpack was visually more similar to a teddy bear. These highly explicit similarity judgments can be contrasted with the largely implicit visual similarity computations driving search guidance. Whereas the web-based judgments could be measured in seconds, effects of similarity on search guidance appeared almost immediately, at least within the first 199 msec following search display onset (the mean latency of initial saccades in this experiment). Our data suggest a common thread between these two decisions. Regardless of whether a visual similarity relationship had to be completed in time for an initial eye movement, or the opportunity existed to deliberate on this relationship for an extended period, the same features seem to have been represented and compared.

Experiment 3: Searching Through Model-Ranked Distractors

Were subjects from Experiment 1 confining their similarity judgments to purely visual dimensions? The fact that this was the instructed task does not guarantee that non-visual factors were not creeping into the similarity judgments, raising the possibility that these factors, and not visual similarity, were responsible for the categorical guidance observed in Experiment 2. Experiment 3 addressed this possibility.

It is unclear how best to separate visual from non-visual factors in estimates of similarity (Medin, Goldstone, & Gentner, 1993). Even when search stimuli are oriented bars with no compelling semantic properties, semantic distinctions might still influence perceptual decisions (Wolfe et al., 1992). The task of separating these factors using purely behavioral methods is even more daunting in the present study, as our stimuli are realistic objects having an untold number of visual and semantic dimensions. Previous research manipulated semantic factors while matching objects on visual dimensions (e.g., Dahan & Tanenhaus, 2005; Bonitz & Gordon, 2008), but this matching was primarily limited to size and/or shape and relied heavily on the subjective decisions of the experimenters as to whether objects were matched or not. Still other research determined that effects of semantic manipulations were not likely due to visual factors, such as bottom-up salience (Becker, Pashler, & Lubin, 2007; Võ & Henderson, 2009). However, these studies did not tease apart semantic from visual factors with regard to the similarity relationships guiding search.

In Experiment 3 we take a different approach to this problem—turning to the computer vision literature to obtain target-distractor similarity estimates. Recent years have seen considerable success in the development of automated methods for the detection of object categories in realistic scenes (see Everingham, Van Gool, Williams, Winn, & Zisserman, 2009), a task with obvious relevance to categorical visual search. At the core of these methods is the computation of visual similarity relationships between images of scenes and features extracted from training exemplars of a target category. These similarity relationships are potentially useful for our current purpose, as they provide estimates of purely visual similarity between distractors and a categorically-defined target, free from any contamination by semantic properties. Whereas the similarity estimates collected in Experiment 1 may have been based on some mix of visual and non-visual information, the similarity estimates obtained from a computer vision method are undeniably exclusively visual.

To obtain these purely visual similarity estimates we used the computer vision method described in Zhang, Samaras, and Zelinsky (2008). This model works by having multiple visual features contribute flexibly and independently to target classification (see also Zhang, Yu, Zelinsky, & Samaras, 2005), and has already been successfully applied to the identical target and distractor objects used in the present study (Zhang et al., 2008). Specifically, it was used to successfully classify the high-similarity and low-similarity objects from Experiment 1 with respect to both the teddy bear and butterfly object classes. This makes it an obvious choice for our goal of relating computer-vision-based similarity estimates to search guidance; not only was this method able to learn classifiers to discriminate our target categories from random objects, these classifiers were also shown to be successful in capturing human visual similarity relationships between these random objects and the bear and butterfly target categories.[4]

To the extent that the Zhang et al. (2008) model is successful in capturing human visual similarity relationships, and to the extent that these similarity estimates extend to a search task, then displays constructed from high-similarity or low-similarity distractors, as rated by the model, should produce the same patterns of guidance found in Experiment 2. Initial saccades should be preferentially guided to high-similarity distractors, and preferentially guided away from low-similarity distractors, with guidance to medium similarity distractors falling between these two levels. Replicating these patterns in the context of new search displays, assembled using the purely visual similarity estimates from a computer vision model, would offer converging evidence for our claim that visual similarity affects categorical search. Of course failing to replicate these patterns would weaken this claim, and would raise concerns that the evidence for guidance reported in Experiment 2 might have been due to semantic, associative, or other non-visual sources of information.

Computational Methods

The computational model used here combines color histogram features (Swain & Ballard, 1991), texture features (the Scale Invariant Feature Transform, or SIFT; Lowe, 2004), and global shape context features (Belongie et al., 2002) with a well-studied machine learning technique (AdaBoost; Freund & Schapire, 1997) to create teddy bear and butterfly classifiers.

[4] Note that this agreement to human behavior does not mean that the features and learning method used by this model accurately describe how humans arrive at their visual similarity estimates. Making this correspondence is a goal to which we aspire, but one that we believe is still out of reach. However, this modest level of agreement does suggest that this model has the potential to generate visual similarity estimates having behavioral significance, making it relatively unique with respect to other purely automated computational approaches.

A histogram of hues was used to describe a global color feature of an object, similar to the approach used by (Swain & Ballard, 1991). Each sample image was first transformed into the HSV color space; background (white) and achromatic pixels internal to an object were excluded from the histogram by setting a threshold on the saturation channel ($S < 0.15$). The hue channel was evenly divided into 11 bins, and each pixel's hue value was assigned to these bins using binary interpolation. The final color histogram was normalized to be a unit vector. The similarity between a given pair of color histogram features, *CH*1 and *CH*2, was measured using the χ^2 statistic:

$$x^2(CH1, CH2) = SUM\left(\frac{\left([CH1(i) - CH2(i)]^2\right)}{CH1(i) + CH2(i)}\right) \tag{1}$$

where *CH*(*i*) is the value of the *i*th dimension.

The texture feature consisted of a set of local SIFT (Scale Invariant Feature Transform) descriptors applied at image coordinates indicated by an interest point detector. Following Lowe (2004), interest points were selected by finding local extremes on Difference-of-Gaussian (DoG) maps. A SIFT feature localized at each point encoded gradient information (orientation and magnitude) for all pixels within a 16×16 image patch surrounding a given interest point. Each patch was further divided into smaller regions, with each subregion represented by an orientation histogram. The SIFT descriptor has been shown to be robust to rotation, translation and occlusion (Lowe, 2004). To estimate the similarity between a SIFT feature, *P*, and a sample object, *S*, we found $\min D(P, Q_i)$, where $\{Q_i\}$ refers to the set of SIFT features from sample *S*, and *D*(.) computes the Euclidean distance between a pair of SIFT features.

Shape was represented using the global shape context feature descriptor (Belongie et al., 2002). For each image, a fixed number of edge points evenly distributed along the object's contour were sampled. The distribution of these points was described by a coarse histogram feature consisting of uniform bins in log-polar space. The origin of the space was set to the center of the image. By counting the number of edge points grouped by discrete log-distances and orientations, each histogram captured the global shape properties for a given object. The similarity between shape context features was measured by χ^2 distance, similar to the metric used for the color histogram feature (Eq. 1).

Each color histogram, SIFT, and shape context feature obtained from positive training samples was used as a candidate feature that could be selected and used to classify target from non-target objects. To select the most discriminative features for classification from this training set, a popular machine learning technique was used, AdaBoost (Freund & Schapire, 1997). The application of AdaBoost, or boosting, refers to the general method of producing a very accurate prediction rule by combining relatively inaccurate rules-of-thumb (Viola & Jones, 2001). In this study, AdaBoost with heterogeneous features was used, as described in Zhang et al. (2005). This method is similar to AdaBoost, except that the different features are processed independently. This means that separate similarity scores are computed between each sample and each feature type, resulting in separate feature-specific classifiers. Two classifiers were learned and used in this study, one discriminating teddy bears from non-bears and the other discriminating butterflies from non-butterflies. The original sources should be consulted for additional details regarding the AdaBoost method.

Distractors were ordered based on how well they fit the classifier. This resulted in the creation of two rank ordered lists, one indicating distractor visual similarity to teddy bears and the other to butterflies. To create target-distractor similarity conditions analogous to those used in Experiment 2, we divided these rank ordered lists into thirds. The top third of the distractors were considered to be highly similar to the target category, the middle third medium-similarity, and the bottom third low-similarity.

Behavioral Methods

Participants.—Twenty-four Stony Brook University students participated in exchange for course credit, none of whom participated in Experiments 1 or 2. All subjects reported normal or corrected to normal vision. Half searched for a teddy bear target, the other half searched for a butterfly target.

Stimuli and apparatus.—Experiment 3 was conducted using the same equipment as in Experiment 2. The stimuli were also objects selected from the same set of images, although the new selection criteria (described below) required the potential placement of these objects into different conditions. The search displays were therefore different, but were assembled from the same set of objects.

Procedure.—Experiments 2 and 3 had the same conditions and followed the same procedure, with the only difference being the distractor composition of target absent trials; distractors were now selected based on visual similarity estimates obtained from the computer vision model rather than from the behavioral similarity rankings obtained from the Experiment 1 web task. High-similarity trials for each target category were constructed from distractors ranked in the top third of each rank-ordered list, and low-similarity trials were constructed from distractors ranked in the bottom third. Mixed trials consisted of high-similarity distractors from the top third, low-similarity distractors from the bottom third and medium-similarity distractors from the middle third. All other methodological details were identical to those described for Experiment 2.

Results and Discussion

Errors were less than 3% in all conditions and were again excluded from subsequent analyses. These infrequent errors were likely just motor confusions rather than cases of confusing teddy bears or butterflies with random objects.

If categorical search is affected by the visual similarity between our target categories and random distractors, and if the computer vision method is able to capture these relationships, then manual RTs should be slowest on high-similarity trials, faster on mixed trials, and fastest on low-similarity trials. These predictions were confirmed (Table 1). Search times varied with target-distractor visual similarity for both teddy bears ($F(2,22) = 35.84$, p< .001) and butterflies (F(2,22) = 60.95, $p < .001$); post-hoc t-tests with Bonferroni correction showed slower RTs in the high-similarity condition relative to the mixed condition ($p < .01$

for both teddy bears and butterflies) and faster RTs in the low-similarity condition relative to the mixed condition ($p < .01$ for both teddy bears and butterflies).

Analysis of the number of distractors fixated during search revealed the same patterns. Fixated distractors varied with visual similarity for both butterfly targets ($F(2,22) = 74.55$, $p < .001$) and bear targets ($F(2,22) = 93.55$, $p < .001$). More distractors were once again fixated on high-similarity trials (2.42 ± 0.20 for bears, 3.66 ± 0.24 for butterflies) compared to either mixed trials (2.10 ± 0.17 for bears, $p < .01$; 2.88 ± 0.23 for butterflies, $p < .001$) or low-similarity trials (1.01 ± 0.19 for bears, $p < .001$; 1.94 ± 0.24 for butterflies, $p < .001$), with more distractors also fixated on mixed trials than on low-similarity trials ($p < .001$ for bears and butterflies). As similarity between the target categories and the distractors increased, more distractors were fixated.

The availability of high-, medium-, and low-similarity distractors in the mixed condition displays again enabled us to look for direct oculomotor evidence for categorical search guidance (Figure 4B). Analyses of these trials showed a relationship between visual similarity and the probability of first fixation on an object ($F(2,22) = 19.42$, $p < .001$ for butterflies; $F(2,22) = 36.60$, $p < .001$ for bears). As in Experiment 2, high-similarity distractors were more often fixated first compared to medium-similarity distractors ($p < .05$ for bears and $p < .01$ for butterflies). Medium-similarity distractors were more often fixated first compared to low-similarity distractors for bears ($p < .001$) but did not reliably differ from low-similarity distractors for butterflies ($p = .35$). First fixations on high-similarity distractors were well above chance ($t(11) = 5.89$, $p < .01$ for bears; $t(11) = 10.01$, $p < .01$ for butterflies), and first fixations on low-similarity distractors were well below chance ($t(11) = 25.47$, $p < .01$ for bears; $t(11) = 8.32$, $p <.01$ for butterflies), indicating that initial saccades were once again guided towards target-similar distractors and away from target-dissimilar distractors. As before, analysis of initial saccade latencies revealed no reliable differences between the similarity conditions for either butterfly ($F(2,22) = 1.29$, $p = 0.30$) or bear targets ($F(2,22) = 0.76$, $p = 0.48$), arguing against a speed-accuracy interpretation of these guidance patterns.

The conclusion from this experiment is clear; while the results of Experiment 2 could have been confounded by the unintentional inclusion of non-visual features in the behavioral similarity rankings, the same cannot be said for the similarity estimates used in Experiment 3. Even when estimates reflected purely visual features, target-distractor similarity still predicted categorical search performance. This strongly suggests that categorical guidance not only exists, but that it may operate in much the same way as search guidance from a pictorial target preview. The visual features used to represent a categorical target may be different and come from a different source (learned and recalled from memory rather than extracted from a target preview), but the underlying process of comparing these visual features to the search scene and using this signal to guide search may be the same.

Experiment 4: Combining Human- and Model-Ranked Distractors

Sometimes the target-distractor similarity estimates from subjects and the computer vision model agreed, and sometimes they did not. Of the objects that were given consistent rankings by subjects, 39.9% of the objects in the teddy bear condition and 36.7% of the objects in the butterfly condition received the same ranking (high similarity, medium similarity, or low similarity) by both subjects and the model. If there is overlap between the features used by the model to capture visual similarity and the features used by subjects, it is likely to be found in these cases. Potentially even more interesting are cases of disagreement between the model and human similarity estimates. Of the consistently ranked objects, 58.5% of those in the teddy bear condition and 61.4% of those in the butterfly condition received either a high or low ranking by one (subjects or model) but an intermediate ranking by the other (subjects or model). Only rarely did subjects and the model contradict each other completely (1.6% for teddy bears, 1.9% for butterflies); meaning that one measure gave a most-similar (or least-similar) estimate while the other gave a least-similar (or most-similar) estimate. Disagreements between subjects and the model might arise for any number of reasons: perhaps subjects based their estimates in part on semantic features, whereas the model obviously did not, perhaps they both used exclusively visual features, but that these features were different or differently weighted, or perhaps subjects simply used more features than the few that were enlisted by the model. Regardless of the source of the disagreement, how would guidance to these objects compare to those in which the model and subjects agreed? Exploring the effects of these agreements and disagreements on search guidance was the goal of Experiment 4.

Experiments 2 and 3 produced remarkably similar effects of target-distractor similarity on search guidance (Figure 4), but were the judgments from our subjects and the estimates from our model tapping into different aspects of search? To begin addressing this question we pit high-similarity (and low-similarity) objects against each other, where similarity was estimated by subjects, the model, or both. More specifically, target absent trials depicted four objects: a “medium” distractor, which was an object given a medium similarity ranking by both subjects and the model, a “human-only” distractor, which was an object given either a most-similar or a least-similar ranking by subjects, but not by the model, a “model-only” distractor, which was an object ranked as either most-similar or least-similar by the model, but not by subjects, and a “human+model” distractor, which was an object for which subjects and the model agreed on its similarity ranking. We also had an equal number of high-similarity and low-similarity trials, which refers to whether distractors were target-similar or target-dissimilar. For example, on a high-similarity trial the human-only, model-only, and human+model distractors would all be ranked as target-similar by subjects, the model, or both, respectively. Likewise, on a low-similarity trial all three types of distractors would be ranked as target-dissimilar. This was done to evaluate overt search guidance both *towards* a high-similarity distractor, as well as *away from* a low-similarity distractor.

These conditions allow us to test several predictions about the relative usefulness of the behavioral and model similarity estimates in describing search behavior. If distractors in the human+model condition (subjects and model in agreement) are fixated most frequently on high-similarity trials (or fixated least frequently on low-similarity trials), this would suggest that the features underlying the subject and model similarity estimates are both useful in guiding search. If the human-only distractors are fixated as frequently as the human+model distractors, and both are fixated more (assuming high-similarity trials) than the model-only distractors, this would suggest that subjects just use features from the behavioral rankings to guide their search, even when features from the model similarity estimates are also available. This pattern would also suggest that behavioral similarity rankings are more useful in predicting search performance than those from the model, perhaps because these estimates reflect the use of more (or more powerful) features than the relatively small set of features used by the model. Alternatively, if the model-only distractors are fixated as frequently as the human+model distractors, and both are fixated more (again assuming high-similarity trials) than the human-only distractors, this would suggest that the features used by the model are preferable to the ones underlying the behavioral similarity estimates, and are used preferentially to guide search. This somewhat counterintuitive result might be obtained if non-visual features crept into the behavioral similarity rankings, but only basic visual features are used to guide search; objective similarity estimates based on color, texture, and shape features might therefore be better predictors of search guidance than similarity estimates from actual human raters. Finally, the degree of guidance toward target-similar distractors, and away from target-dissimilar distractors, will be assessed for all three conditions (human-only, model-only, and human+model) by comparing these levels to the level of guidance observed to the medium distractor, which serves as a similarity baseline present on each search trial.

Method

Participants.—Twenty-four Stony Brook University students participated in exchange for course credit, none of whom participated in Experiments 1, 2, or 3. All subjects reported normal or corrected to normal vision. Half searched for a teddy bear target, the other half searched for a butterfly target.

Stimuli and apparatus.—Experiment 4 was conducted using the same equipment as in Experiments 2 and 3. The stimuli were also objects selected from the same image set, although the new selection criteria (described below) again required the placement of these objects into different conditions.

Design and Procedure.—Experiment 4 followed the same procedure as Experiments 2 and 3, the only difference being the composition of the search displays. Unlike the previous experiments each display depicted only four objects. On target present trials (104 trials), the displays consisted of an object from the target category (bear or butterfly) and three random, unranked distractors. Target absent trials were divided into four interleaved conditions (26 trials per condition): high-similarity teddy bear, low-similarity teddy bear, high-similarity butterfly, low-similarity butterfly. Distractors for the low- and high-similarity teddy bear and butterfly trials were chosen based on similarity estimates obtained relative to the teddy bear and butterfly target categories, respectively. All subjects, regardless of whether they were searching for a teddy bear or a butterfly, saw the identical target absent trials; the only difference between the two groups of subjects was the designated target category. Target present trials were also identical across groups, except that the target from one category was replaced with a target from the other (teddy bear replaced with a butterfly, or vice versa). By having all subjects search through the same target absent displays, we control for all visual factors unrelated to the similarity between the distractors and the designated target category.

On all four types of target absent trials (high- and low-similarity for teddy bear and butterfly targets) there were four types of distractors, one distractor of each type per display. The human-only distractor was ranked by subjects in Experiment 1 to be consistent with the target absent condition (e.g., if the trial was from the high-similarity butterfly condition, this object would have been consistently ranked by subjects as being most like a butterfly), but this object was given either an opposite or intermediate similarity estimate by the model. The model-only distractor was chosen to be consistent with the target absent condition based on the model similarity estimates, but was given the opposite similarity ranking (i.e., low similarity for high-similarity trials) or an intermediate ranking by subjects in Experiment 1. The human+model distractor was chosen to agree with the target absent condition by both the human and model similarity estimates; if the trial was from the low-similarity teddy bear condition, this distractor would be ranked as dissimilar to a teddy bear by both subjects and the model. The medium distractor was selected to have a medium similarity ranking by the model and by subjects, and served as a trial-by-trial baseline against which overt search guidance could be assessed.

Results and Discussion

To determine the similarity measure (human-only, model-only, or human+model) that is most predictive of search behavior, we analyzed each target absent trial to find the type of distractor that was fixated first. We then grouped these data by target absent condition (high- and low-similarity for teddy bears and butterflies) and plotted their relative frequencies in Figure 5.

On high-similarity trials (Figure 5A) we expected the most frequently fixated first distractor to indicate the object considered by the search process to be most similar to the target category. For the butterfly search, a highly significant difference was found across distractor type ($F(3,33) = 15.97$, $p < .001$). Post hoc LSD tests confirmed that the human+model distractors were fixated first most frequently ($p < .05$ for all comparisons). Gaze was directed first to these objects on about 40% of the trials, far more frequently that what would be expected by chance (25%). Fixated next most frequently were the human-only distractors, which were fixated first more often than either the model-only ($p < .01$) or the medium ($p < .01$) distractors. These latter two types of distractors did not differ in their first fixation frequency ($p = .70$). A qualitatively different data pattern was found for the teddy bear search. Although distractor types again differed in their first fixation frequency ($F(3,33) = 5.01$, $p < .01$), no significant differences were found between the human-only, model-only, and human+model objects ($p \geq .37$ for all comparisons). All three, however, were fixated more frequently than the medium distractors ($p < .05$ for all comparisons), which were fixated first well below chance.

On low-similarity trials (Figure 5B) we expected the distractors ranked by the most predictive similarity measure to be first fixated *least* frequently, indicating guidance to other, more similar objects. For the butterfly search, we again found a significant difference across distractor type ($F(3,33) = 5.48$, $p < .05$). Note the nearly symmetrical reversal of pattern relative to the corresponding high-similarity data (Figure 5A). Whereas the human+model and human-only distractors were first fixated above chance in the high-similarity conditions, in the low-similarity conditions these distractors were both fixated well below chance. Post hoc LSD tests confirmed that first fixations on human+model distractors were less frequent than those on the three other distractor types ($p < .05$ for all comparisons), and that first fixations on human-only distractors were less frequent than those on model-only and medium distractors ($p < .05$ for both comparisons). First fixations on model-only and medium distractors did not reliably differ ($p = .50$), and both were fixated well above chance. Similar patterns were found for the teddy bear search. Distractor types again differed in their first fixation frequency ($F(3,33) = 8.13$, $p < .01$), with human+model distractors first fixated less than model-only and medium distractors ($p < .01$ for both comparisons) but not human-only distractors ($p = .14$). Human-only distractors were first fixated less frequently than model-only or medium distractors ($p < .05$ for both comparisons), with the difference between model-only and medium distractors not reaching significance ($p = .12$). First fixations on medium distractors were well above chance.

In Experiment 2 we found that similarity estimates obtained from subjects were good predictors of search performance, and in Experiment 3 we found that the same was true for similarity estimates obtained from a computer vision model. In Experiment 4 we examined cases in which the two estimates of similarity agreed or disagreed. Taken together, the human+model distractors were generally better predictors of search guidance than the human-only distractors, and the human-only distractors were generally better predictors of search guidance than the model-only distractors. Although this pattern was most consistent for the butterfly search task, it does suggest that effects of similarity on search are best captured by objects ranked by both subjects and our model, and that this is true regardless of whether these objects were ranked as being most-similar or least-similar to the target category. The fact that human+model distractors best predicted search guidance further suggests that both similarity estimates captured features that are useful in guiding search. However, there is an alternative explanation that must be considered. It may be that the benefit found for the human+model distractors is due to subjects and the model basing their respective rankings in part on factors that are *not* useful for guiding search. For example, subjects might have included in their similarity rankings semantic information that either cannot guide search (De Graaf, Christiaens, & d'Ydewalle, 1990; Henderson, Weeks, & Hollingworth, 1999; Võ & Henderson, 2009) or is irrelevant to the search task ("I had one of these as a child"). Given that the model would not represent such information, by requiring the behavioral and model rankings to agree we may have inadvertently constrained the human+model distractors to those objects in which such factors did not play a role. A related argument might apply to the model's features. Some of these computer vision features may be useful in predicting search guidance by subjects, and others may not. By requiring that the human+model distractors agree in their respective similarity rankings, objects rich in visual features that are less "human-like" may have been excluded. The guidance benefit for human+model distractors may therefore be due to the selective exclusion of problematic objects from this set, ones in which subjects relied on non-visual features in their rankings and ones in which the model used visual information not used by human raters. Better identifying the specific features instrumental in producing similarity effects on search guidance will be an important direction for future work.

In this experiment we also looked at cases in which the behavioral and model similarity estimates disagreed, and found that search guidance was generally better predicted by the human-ranked objects when model-only and human-only distractors appeared in the same display. More first fixations were on distractors ranked as high-similarity by subjects (28% −30%) than on high-similarity distractors ranked by the model (14%−25%), and fewer first fixations were on distractors ranked as low-similarity by subjects (16–19%) than on low-similarity distractors ranked by the model (28–38%). In fact, for the teddy bear category the model contributed very little to guidance beyond what was already captured by the behavioral rankings, as evidenced by the non-significant differences in first fixations between the human-only and the human+model distractors. These patterns suggest that the features used in behavioral similarity judgments are more useful for guiding search than the features used by the present model, a finding that is perhaps unsurprising given that the simple color, texture, and shape features used by this model were never intended to be an accurate or complete characterization of the visual information used by subjects to guide their search. Note also that this does not mean that search was guided based on a semantic analysis of the search display. While it is true that the model used information from only visual features and that the behavioral similarity judgments were not restricted in this way, it is also true that the behavioral rankings may have included other varieties of purely visual information not considered by the model. It is quite likely that the relationship between categorical search guidance and target-distractor similarity uses more than just the three visual features considered in this study, and that this explains the difference in predictability between the behavioral and model similarity estimates, even though these three features were successful in predicting guidance when other, perhaps more preferred features were unavailable (as shown in Experiment 3).

Conclusion

Search guidance from a pictorial preview is known to decrease with increasing visual similarity between a target and distractors; in the present study we extend this well established relationship to categorically-defined targets. Previous research had suggested that search is unguided to categorical targets (e.g., Castelhano et al., 2008; Wolfe et al., 2004). In light of the present findings, as well as other recent evidence, this suggestion should be revisited. Multiple studies have now shown guidance in the very first saccades made to categorical targets (Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009). Our work extends this finding to non-target objects that are visually similar to the target category. Specifically, in the absence of a target our subjects preferentially directed their initial saccades *to* distractors that were target-similar, and *away* from distractors that were target-dissimilar (Figures 4 and 5). These patterns, when combined with the patterns of manual search efficiency found in the high-similarity and low-similarity distractor conditions (Table 1), provide strong converging evidence for categorical search guidance in our tasks. The fact that these results were obtained despite the highly non-obvious similarity relationships between random objects and teddy bears / butterflies, makes the clear expression of guidance reported here all the more striking.

We can also conclude that these effects of similarity on categorical search can be well described by objective visual similarity estimates, regardless of whether these estimates were based on explicit visual similarity rankings (Experiments 1 and 2), derived from a computer vision model of object category detection (Experiment 3), or both (Experiment 4). This too is a striking finding. The lengthy deliberations that accompanied the behavioral similarity rankings, and certainly the simplistic visual features underlying the model's estimates, might have easily resulted in no success whatsoever in predicting categorical search behavior. The fact that these radically different methods both successfully predicted patterns of search guidance is informative, suggesting that the computation of visual similarity is not only a core cognitive operation, but one that is relatively stable across estimation method. We speculate that visual similarity is computed early and automatically during perception, and once derived is used to mediate a variety of perceptual (e.g., search guidance) and cognitive (similarity judgments) behaviors. To the extent that this is true, it bodes well for the diversity of researchers in cognitive psychology, human-computer interaction, and vision science, all attempting to better understand human visual similarity relationships.

Acknowledgments

We thank Nancy Franklin and Antonio Freitas for helpful comments and feedback in the preparation of this manuscript. We also thank Ryan Moore, Jonathan Ryan, and Arunesh Mittal for their help with data collection, and all the members of the Eye Cog lab for many invaluable discussions. This work was supported by NIH grant R01-MH063748 to GJZ.

References

Alexander RG, Zhang W, & Zelinsky GJ (2010). Visual similarity effects in categorical search. In Ohlsson S. & Catrambone R. (Eds.), Proceedings of the 32nd Annual Conference of the Cognitive Science Society (pp. 1222–1227). Austin, TX: Cognitive Science Society.

Becker MW, Pashler H, & Lubin J. (2007). Object-intrinsic oddities draw early saccades. Journal of Experimental Psychology: Human Perception and Performance, 33, 20–30. [PubMed: 17311476]

Belongie S, Malik J, & Puzicha J. (2002). Shape matching and object recognition using shape contexts. Pattern Analysis and Machine Intelligence, 24(4), 509–522.

Bonitz VS, & Gordon RD (2008). Attention to smoking-related and incongruous objects during scene viewing. Acta Psychologica, 129, 255–263. [PubMed: 18804752]

Castelhano MS, Pollatsek A, & Cave KR (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. Psychonomic Bulletin & Review, 15(4), 795–801. [PubMed: 18792506]

Chen X, & Zelinsky GJ (2006). Real-world visual search is dominated by top-down guidance. Vision Research, 46, 4118–4133. [PubMed: 17005231]

Cockrill P. (2001). The teddy bear encyclopedia. New York: DK Publishing Inc.

Dahan D, & Tanenhaus MK (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. Psychonomic Bulletin & Review, 12(3), 453–459. [PubMed: 16235628]

De Graef P, Christiaens D, & d'Ydewalle G. (1990). Perceptual effects of scene context on object identification. Psychological Research, 52, 317–329. [PubMed: 2287695]

Duncan J. (1983). Category effects in visual search: A failure to replicate the 'oh-zero' phenomenon. Perception & Psychophysics, 34(3), 221–232. [PubMed: 6646963]

Duncan J, & Humphreys G. (1989). Visual search and stimulus similarity. Psychological Review, 96 (3), 433–458. [PubMed: 2756067]

Eckstein MP, Beutter BR, Pham BT, Shimozaki SS, & Stone LS (2007). Similar neural representations of the target for saccades and perception during search. Neuron, 27, 1266–1270.

Everingham M, Van Gool L, Williams CKI, Winn J, & Zisserman A. (2009). The PASCAL Visual Object Classes Challenge 2009 (VOC2009). Retrieved from http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2009/index.html

Findlay JM (1997). Saccade target selection during visual search. Vision Research, 37, 617–631. [PubMed: 9156206]

Freund Y, & Schapire R. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119–139.

Henderson JM, Weeks PA, & Hollingworth A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. Journal of Experimental Psychology: Human Perception and Performance, 25, 210–228.

Hwang AD, Higgins EC & Pomplun M. (2009). A model of top-down attentional control during visual search in complex scenes. Journal of Vision, 9 (5):25, 1–18.

Levin DT, Takarae Y, Miner AG, & Keil F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. Perception and Psychophysics, 63(4), 676–697. [PubMed: 11436737]

Loftus GR, & Mackworth NH (1978). Cognitive determinants of fixation location during picture viewing. Journal of Experimental Psychology: Human Perception and Performance, 4, 565–572. [PubMed: 722248]

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

Medin DL, Goldstone RL, & Gentner D. (1993). Respects for similarity. Psychological Review, 100(2), 254–278.

Rayner K, Castelhano M, & Yang J. (2009). Eye movements when looking at unusual/weird scenes: Are there cultural differences? Journal of Experimental Psychology: Learning, Memory, and Cognition, 35(1), 254–259.

Swain M, & Ballard D. (1991, November). Color indexing. International Journal of Computer Vision, 7(1), 11–32.

Tavassoli A, van der Linde I, Bovik AC, & Cormack LK (2009). Eye movements selective for spatial frequency and orientation during active visual search. Vision Research, 49, 173–181. [PubMed: 18992270]

Treisman AM (1991). Search, similarity, and integration of features between and within dimensions. Journal of Experimental Psychology: Human Perception and Performance. 17 (3), 652–676. [PubMed: 1834783]

Treisman AM, & Sato S. (1990). Conjunction search revisited. Journal of Experimental Psychology: Human Perception and Performance, 16, 459–478. [PubMed: 2144564]

Underwood G, & Foulsham T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. Quarterly Journal of Experimental Psychology, 59, 1931–1949.

Underwood G, Templeman E, Lamming L, & Foulsham T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. Consciousness and Cognition, 17, 159–170. [PubMed: 17222564]

Vickory TJ, King L-W, & Jiang Y. (2005). Setting up the target template in visual search. Journal of Vision, 5, 81–92. [PubMed: 15831069]

Võ ML-H, & Henderson JM (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. Journal of Vision, 9(3), 1–15.

Wolfe JM (1994). Guided search 2.0: A revised model of visual search. Psychonomic Bulletin & Review 1(2), 202–238. [PubMed: 24203471]

Wolfe JM, Friedman-Hill S, Stewart M, & O’Connell K. (1992). The role of categorization in visual search for orientation. Journal of Experimental Psychology: Human Perception and Performance, 18, 34–49. [PubMed: 1532193]

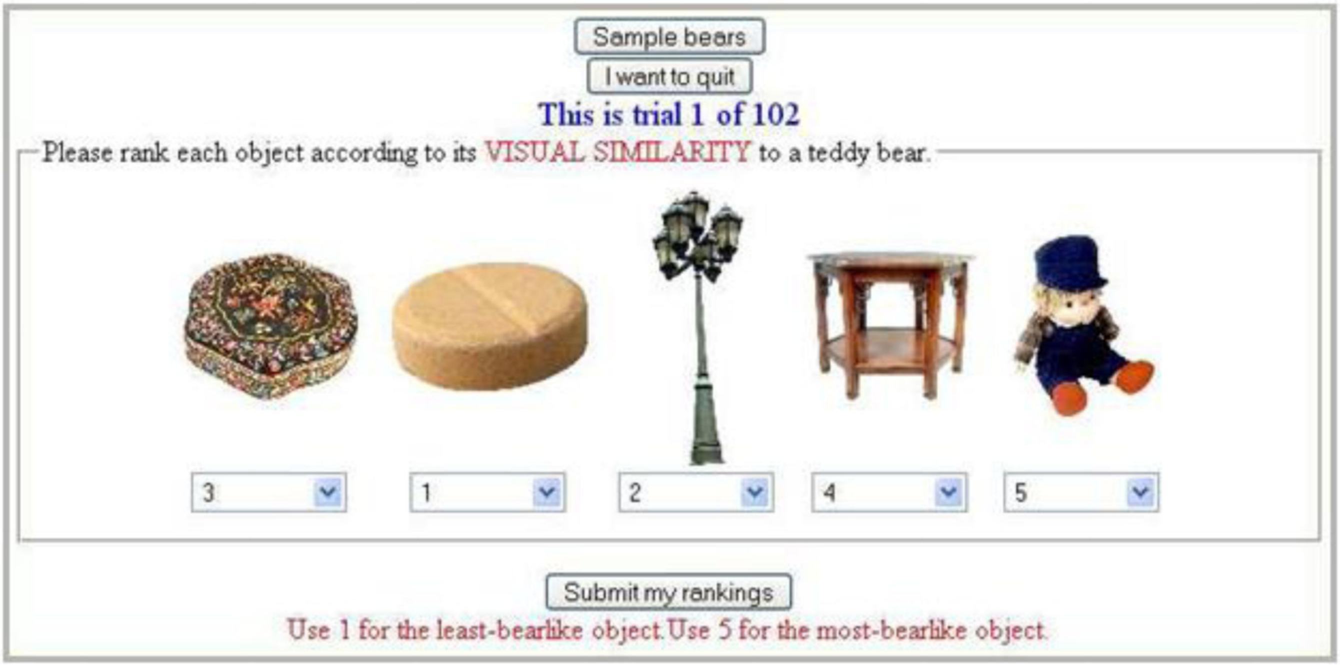
Wolfe JM, Horowitz TS, Kenner N, Hyle M, & Vasan N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. Vision Research, 44, 1411–1426. [PubMed: 15066400]

Yang H, & Zelinsky GJ (2009). Visual search is guided to categorically-defined targets. Vision Research, 49, 2095–2103. [PubMed: 19500615]

Zelinsky GJ (2008). A theory of eye movements during target acquisition. Psychological Review 115(4), 787–835. [PubMed: 18954205]

Zhang W, Samaras D, & Zelinsky GJ (2008). Classifying objects based on their visual similarity to target categories. Proceedings of the 30th Annual Conference of the Cognitive Science Society (pp. 1856–1861).

Zhang W, Yu B, Zelinsky GJ, & Samaras D. (2005). Object class recognition using multiple layer boosting with heterogeneous features. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2, 323–330.



This is trial 1 of 102

Please rank each object according to its VISUAL SIMILARITY to a teddy bear.

3 1 2 4 5

Use 1 for the least-bearlike object.Use 5 for the most-bearlike object.

Figure 1.

Screen-shot of a teddy bear trial from Experiment 1.

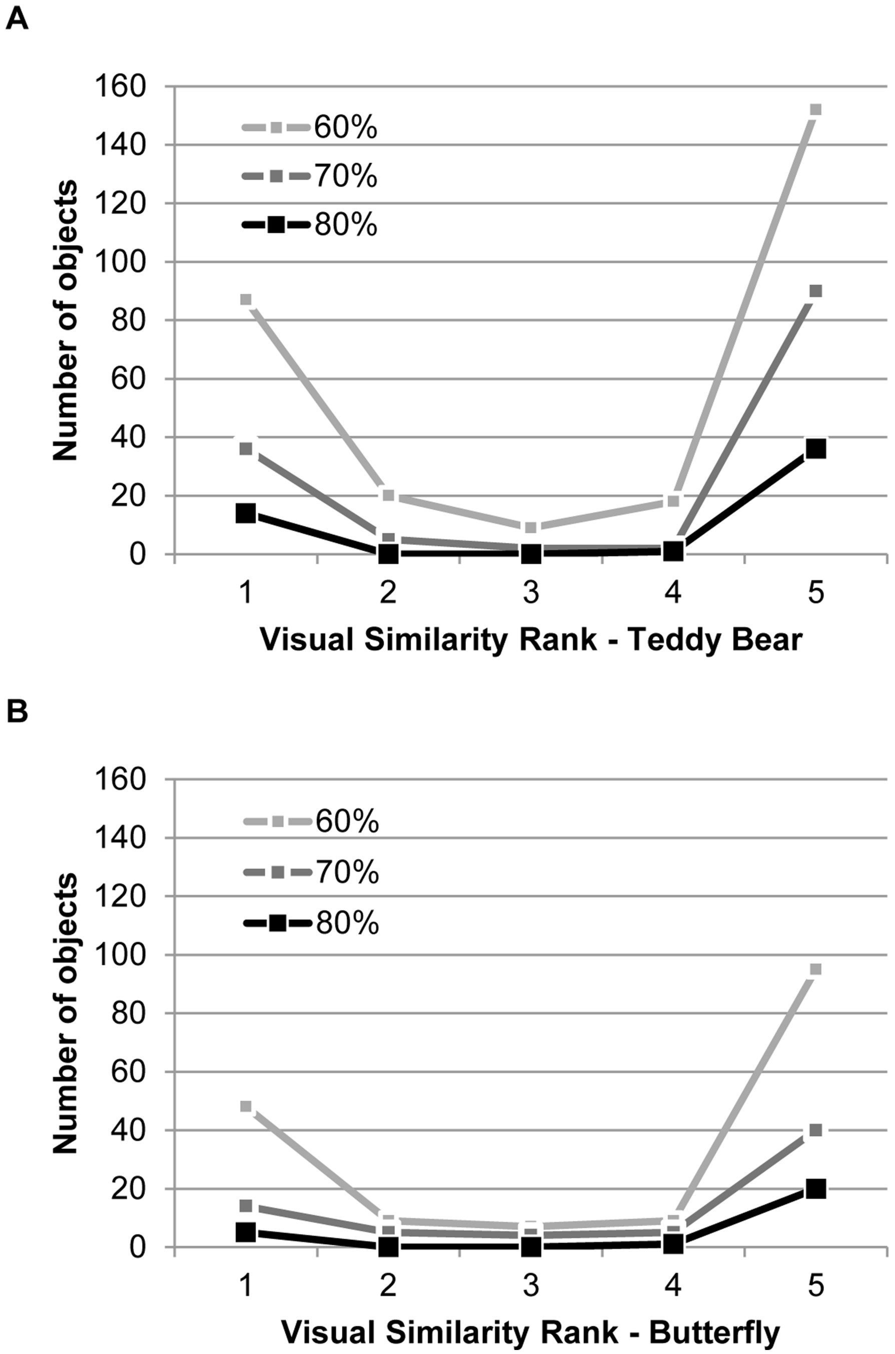


Figure 2.

The number of objects corresponding to 60%, 70%, and 80% levels of inter-subejct agreement for each of the five visual similarity rankings (1 = least similar; 5 = most similar). (A) Teddy bears. (B) Butterflies.

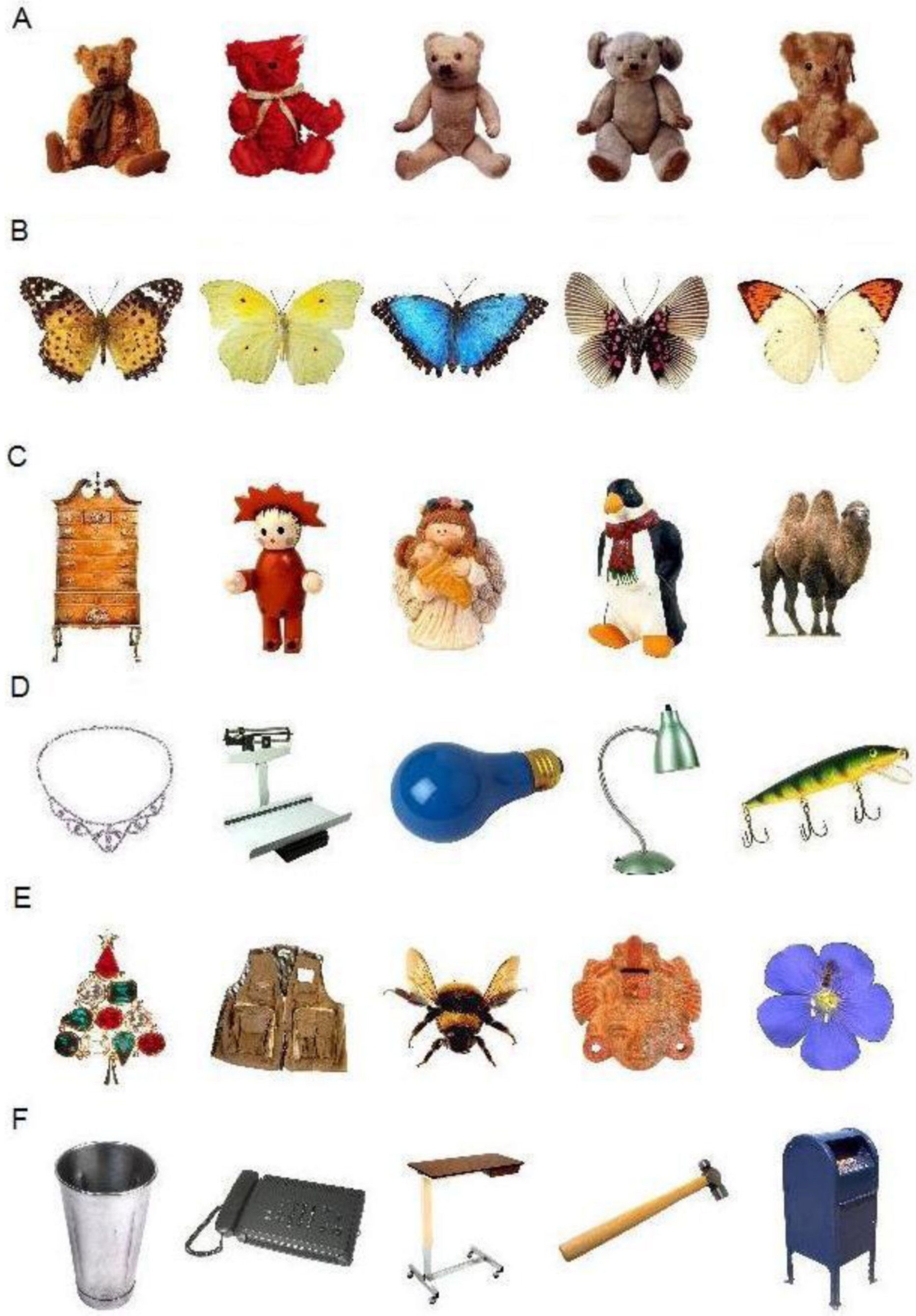


Figure 3.

Representative target and non-target objects. (A) Teddy bears, (B) butterflies, (C) high similarity to teddy bears, (D) low similarity to teddy bears, (E) high similarity to butterflies, (F) low similarity to butterflies.

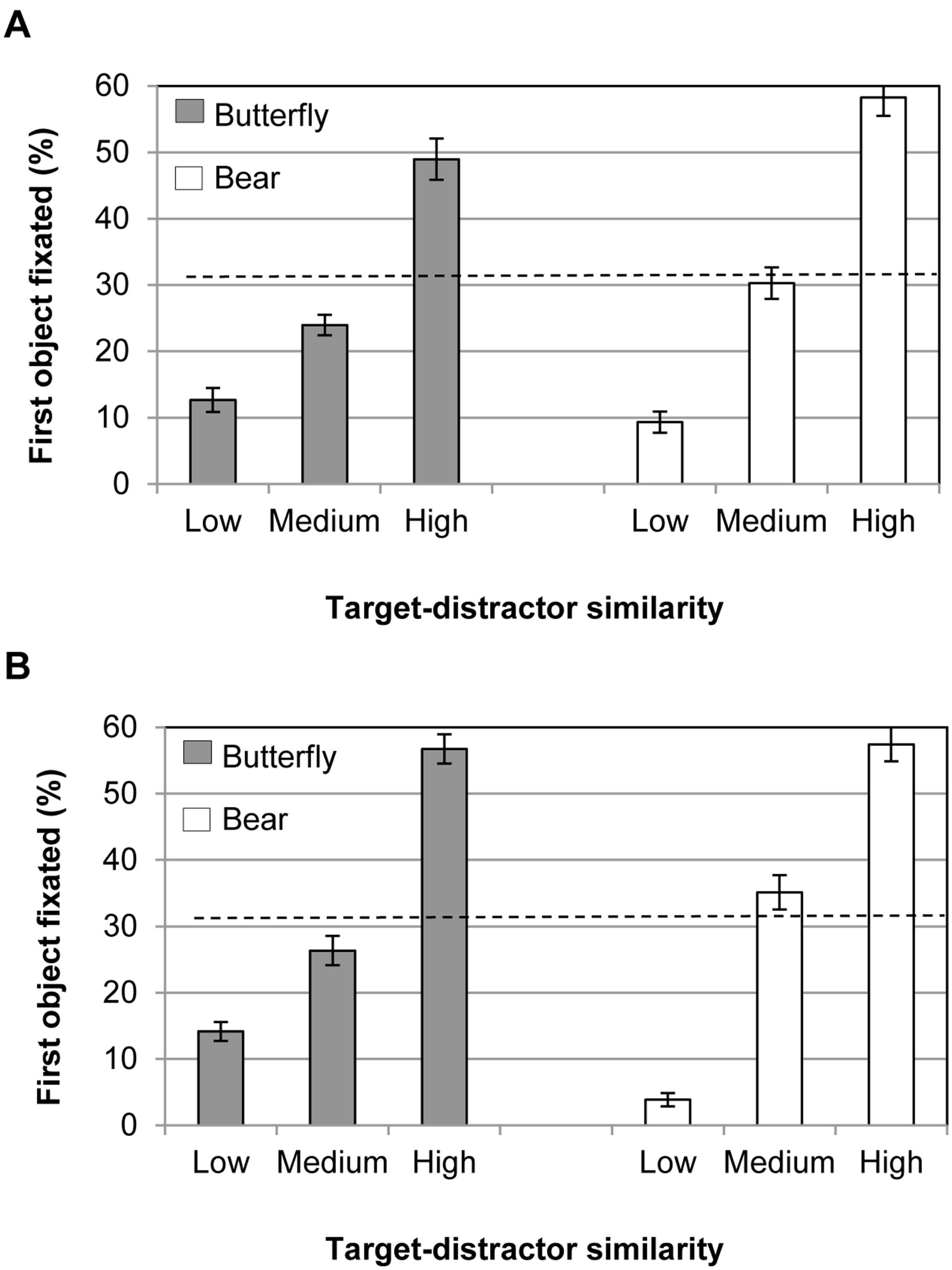


Figure 4.

Percentage of mixed condition trials in which the first object fixated was ranked as having a low, medium, or high target-distractor similarity for (A) Experiment 2 and (B) Experiment 3. Error bars show one standard error. Dashed lines indicate chance.

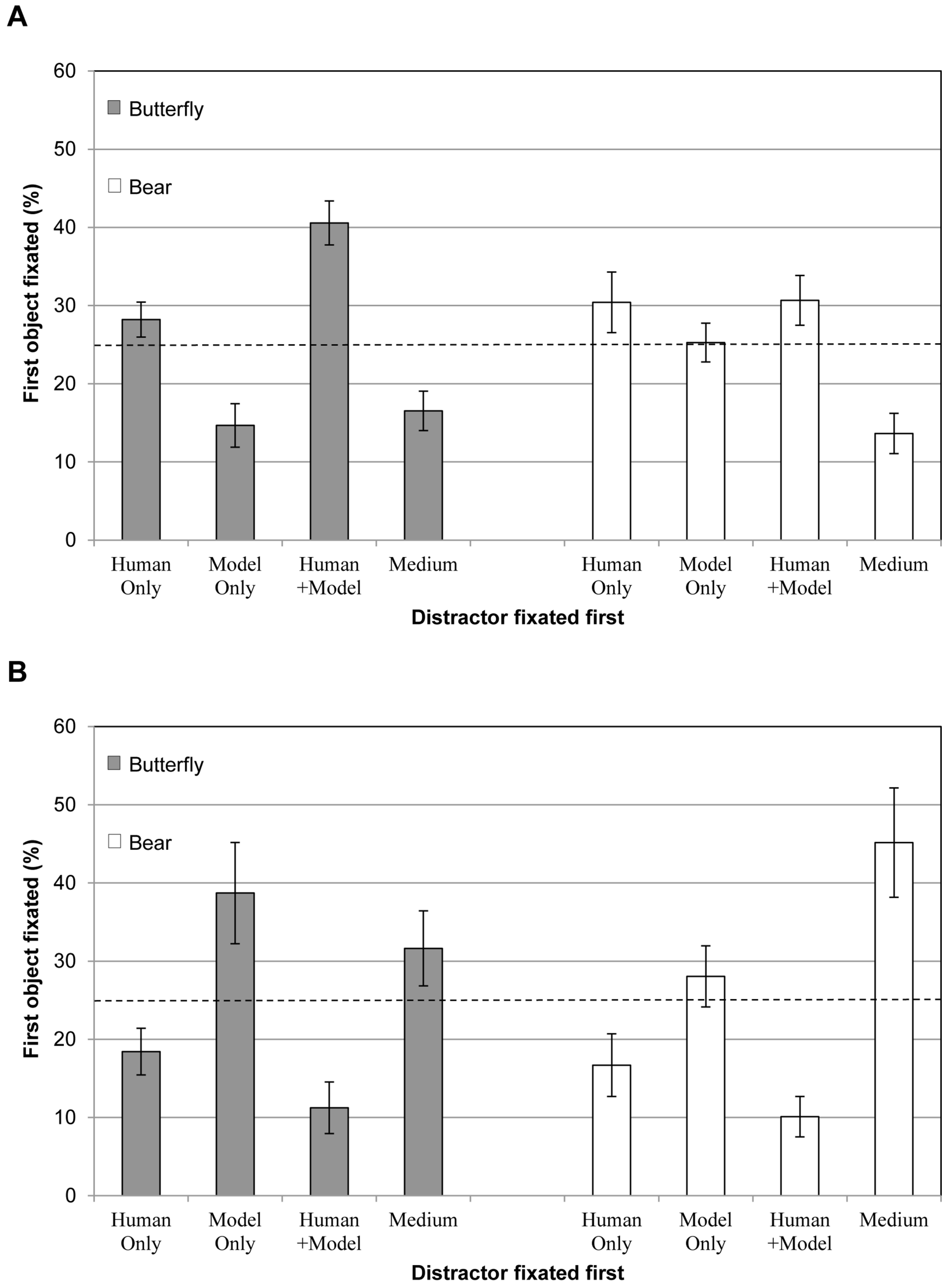


Figure 5.

Percentage of high-similarity (A) and low-similarity (B) trials in which the first object fixated was a human-only, model-only, human+model, or medium distractor in Experiment 4. Error bars show one standard error. Dashed lines indicate chance.

Table 1

Manual RTs (seconds) by similarity condition and target category in Experiments 2 and 3

	Experiment 2		**Experiment 3**	
	Butterfly	**Bear**	**Butterfly**	**Bear**
High	1.17 (.06)	1.48 (.14)	1.59 (.13)	1.24 (.15)
Mixed	0.97 (.06)	1.15 (.11)	1.25 (.10)	1.07 (.15)
Low	0.82 (.05)	0.84 (.08)	0.92 (.09)	0.74 (.09)

Note. Values in parentheses indicate one standard error.