Title
Visual Similarity Effects in Categorical Search

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)

ISSN
1069-7977

Authors
Alexander, Robert
Zhang, Wei
Zelinsky, Gregory

Publication Date
2010

Peer reviewed

Visual Similarity Effects in Categorical Search

Robert G. Alexander[1] (rgalexander@notes.cc.sunysb.edu), Wei Zhang (weiz@microsoft.com)[2,3]
Gregory J. Zelinsky[1,2] (Gregory.Zelinsky@stonybrook.edu)
[1]Department of Psychology, Stony Brook University
[2]Department of Computer Science, Stony Brook University
[3]Microsoft Corporation

Abstract

The factors affecting search guidance to categorical targets are largely unknown. We asked how visual similarity relationships between random-category distractors and two target classes, teddy bears and butterflies, affects search guidance. Experiment 1 used a web-based task to collect visual similarity rankings between these target classes and random objects, from which we created search displays having either high-similarity distractors, low-similarity distractors, or "mixed" displays with high, medium, and low-similarity distractors. Subjects made faster manual responses and fixated fewer distractors on low-similarity displays compared to high. On mixed trials, first fixations were more frequent on high-similarity distractors (bear=49%; butterfly=58%) than low-similarity distractors (bear=9%; butterfly=12%). Experiment 2 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 is indeed guided by visual similarity.

Keywords: Visual search; eye movements; categorical guidance; visual similarity; object class detection

Introduction

You have probably had the experience of searching for your car in a parking lot and finding 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 target of your search.

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., Pomplun, 2006; Treisman & Sato, 1990; Wolfe, 1994; Zelinsky, 2008). Despite their many differences, 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 a display. In general, the more similar an object is to the target, the more likely that object will be fixated.

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 would have few visual features of the 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 et al., 2008; Wolfe et al., 2004).

The present study enters this debate on the existence of categorical guidance, focusing it on the relationship between target-distractor visual similarity and guidance to categorically-defined realistic targets. 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? Given that the representation of categorical targets is largely unknown, it may be the case that target descriptions are dominated by non-visual features, such as semantic or functional properties of the target category. If this is the case, guidance to the target may be weak or even nonexistent, potentially explaining the discrepant findings. To the extent that categorical search does use non-visual features, effects of target-distractor visual similarity would therefore not be expected. However, if target categories are represented visually, one might expect the same target-distractor similarity relationships demonstrated for target-specific search to extend to categorical search.

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 an average be obtained from many target exemplars or should features be extracted from a particular exemplar that is representative of the target class?

In light of the difficulties associated with directly manipulating the specific features underlying visual similarity, we opted for a more holistic approach—to use ratings of visual similarity obtained from subjects. Specifically, we obtained ratings from Zhang et al. (2008), who used a web experiment to collect visual similarity estimates between random objects and categorical targets for the purpose of comparing these estimates to the behavior of a computational model of object class detection. Subjects were randomly assigned to either a butterfly target class or a teddy bear target class, and their task was to rate real-world objects (from the Hemera collection) to these target categories. They did this by rank-ordering groups of five objects; each trial showed five random objects, and the subjects' task was to give each a 1-5 ranking, where "1" indicated low target similarity and 5 indicated high target similarity (objects given the 2-4 rankings and objects with low inter-subject ranking agreement were considered medium similarity). There were 142 subjects, yielding a total of 71,000 butterfly and teddy bear similarity estimates for 2,000 different objects. 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. Consult Zhang et al. (2008) for additional details regarding this web-based collection of visual similarity estimates.

Using these estimates of visual similarity, Experiment 1 asked whether the visual similarity relationships known to affect search for specific targets also extends to categorical search. Previous arguments for the existence of categorical search guidance relied on evidence showing the preferential direction of initial saccades to targets (Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009). Although there is good reason to believe that these initial saccades are dominated by visual features, and occur too early in search to be influenced by semantic relationships between targets and distractors, it is still possible that the preferential fixation of categorical targets might have been influenced by non-visual factors. More compelling would be a demonstrated relationship between categorical guidance and a manipulation of target-distractor visual similarity; providing this evidence was the primary goal of this experiment.

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 as discussed can be expressed in even the first search saccade (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 guidance or manual search efficiency.

Experiment 1

Method

Participants Twenty-four students from Stony Brook University participated in exchange for course credit. All subjects reported normal or corrected to normal vision.

Stimuli and Apparatus Targets and distractors were selected from the objects used by Zhang et al. (2008). The target categories were teddy bears, obtained from Cockrill (2001), and butterflies, obtained from the Hemera collection. The distractors were also Hemera objects. Each object was sized to subtend ~2.8º of visual angle.

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. Trials were displayed on a flat-screen monitor at a resolution of 1024 × 768 pixels (subtending 28º × 21º) and a refresh rate of 85 Hz.

Design and 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 the Zhang et al. (2008) web task.

There were three target absent conditions: high-similarity trials (all distractors were similar to the target category), low-similarity trials (all distractors were dissimilar to the target category), and "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 (see Figure 1). 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 which distracters were preferentially fixated (i.e., search guidance) as a function of target-distractor similarity.

Target presence/absence and similarity condition were within-subjects 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 only the target absent trials contained the similarity manipulation, 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 were not confusing the high-similarity distractors for targets (e.g., a stuffed bunny distractor was not 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 differences were significant for both butterfly targets ($F(2,22) = 46.87$, $p < .001$) and for bear targets ($F(2,22) = 53.85$, $p < .001$). The number of distractors fixated during search also differed between the similarity conditions, and this again occurred for both butterfly ($F(2,22) = 30.41$, $p < .001$) and bear targets ($F(2,22) = 59.55$, $p < .001$). Distractors were fixated most frequently on the high-similarity trials (3.16±0.23 for bears; 2.50±0.36 for butterflies), followed by the medium-similarity trials (2.53±0.24 for bears; 1.83±0.31 for butterflies), and finally the low-similarity trials (1.51±0.23 for bears; 1.29±0.26 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. Consistent with the RT analyses we found that distractor similarity to the target determined which objects were fixated first on mixed condition trials (Figure 2A). High-similarity distractors were more often fixated first compared to medium-similarity distractors, which were more often fixated first compared to low-similarity distractors, and this pattern was found for both butterflies ($F(2,22) = 10.13$, $p < .01$) and for bears ($F(2,22) = 30.15$, $p < .001$).

Two conclusions follow from our data. First, categorical search guidance is affected by target-distractor visual similarity. As the visual similarity between a distractor and a target category increases, search efficiency decreases. 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

A **B** **C**

Figure 1: Objects from a typical mixed trial. (A) low-similarity, (B) medium-similarity, and (C) high-similarity distractors, as ranked to the teddy bear target category.

had been the topic of debate (Castelhano et al., 2008, and Wolfe et al., 2004). Not only is categorical search guided, it is guided by matching visual features to a visual representation of the target category.

The second conclusion following from our data is that explicit visual similarity rankings 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 that drove 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 ms following search display onset (the average latency of initial saccades in this experiment). Our data suggest a common thread between these two processes. Regardless of whether a visual similarity relationship has to be completed in time for an initial eye movement, or the opportunity exists to deliberate on this relationship for an extended period, the same features seem to be represented and compared.

Table 1: Manual RTs by similarity condition, in seconds

	Experiment 1		Experiment 2	
	Butterfly	Bear	Butterfly	Bear
High	1.17 (.06)	1.48 (.14)	1.59 (.13)	1.24 (.15)
Medium	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.

Experiment 2

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 observed categorical guidance. Experiment 2 addressed this possibility.

It is unclear how best to separate visual from non-visual factors in estimates of similarity. Even when stimuli are oriented bars with no compelling semantic properties, semantic features might still influence perceptual decisions (Wolfe et al., 1992). The task of separating these factors

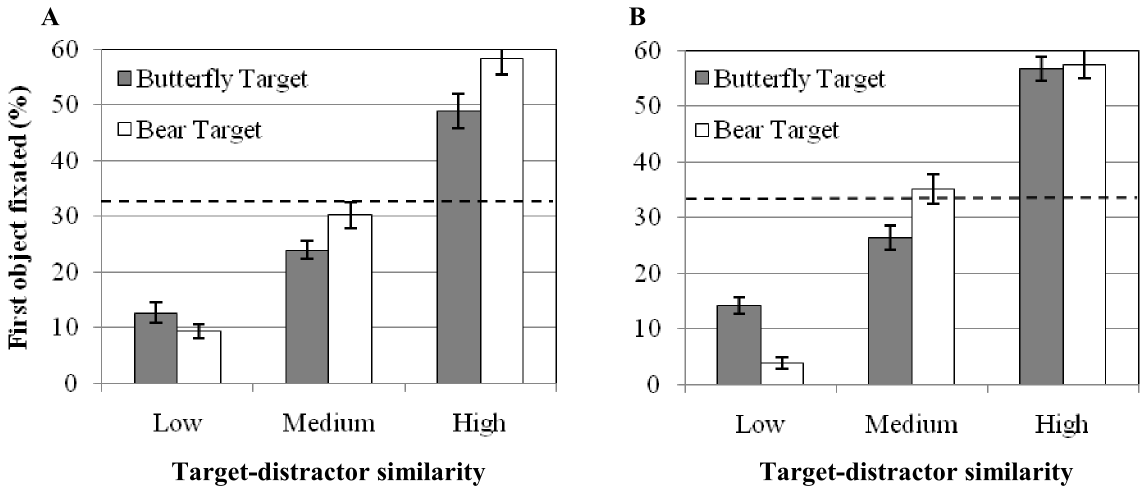

Figure 2: Percentage of mixed condition trials in which the first object fixated had a low, medium, or high target-distractor similarity for (A) Experiment 1 and (B) Experiment 2. Error bars show one standard error. Dashed lines indicate chance.

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.

In Experiment 2 we take a different approach to this problem—turning to the computer vision literature to obtain similarity estimates. Recent years have seen considerable success in the development of automated methods for the detection of object categories in realistic scenes, a task with obvious relevance to categorical visual search. At the core of these methods is the computation of visual similarity relationships between visual images and features extracted from a target class. 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 used 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 incontrovertibly exclusively visual.

To obtain these purely visual similarity estimates we used the computer vision method described in Zhang et al. (2008). We chose this method for two reasons. First, it works by having multiple visual features contribute flexibly to target classification (see also Zhang et al., 2005). Specifically, it combines state-of-the-art 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 classifiers having features tailored for the detection of specific target categories. The advantage of this method over other automated object classification techniques is that similarity estimates can be based on the contribution of multiple features, not just one.

Our second reason for choosing the Zhang et al. (2008) model is that it has already been successfully applied to the identical target and distractor objects used in the present study. Specifically, it successfully classified the high-similarity and low-similarity objects from the above-described web task, regardless of whether the target category was a teddy bear or a butterfly. This makes the Zhang et al. model an obvious choice for our goal of collecting computer-vision-based similarity estimates; not only was this model able to learn classifiers to discriminate our target categories from random objects, these classifiers were also shown to be partially successful in capturing human visual similarity relationships between these random objects and the bear and butterfly target classes.[1]

To the extent that the Zhang et al. model is successful in capturing human visual similarity relationships, and to the extent that these similarity estimates extend to a search task (as we found in the previous experiment), then displays constructed of high-similarity or low-similarity distractors, as rated by the model, should produce the same patterns of guidance found in Experiment 1. 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

[1] Note that this agreement to human behavior does not mean that the features and learning method used by this model accurately describes 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 the model's multi-feature approach has the potential to generate visual similarity estimates having behavioral significance, which makes it relatively unique with respect to purely automated computational approaches.

Experiment 1 might have been due to semantic, associative, or other non-visual sources of information.

Method

Participants Twenty-four Stony Brook University students participated in exchange for course credit, none of whom participated in Experiment 1. 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 2 was conducted using the same equipment as in Experiment 1. The stimuli were also objects selected from the same image set, 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.

Design and procedure Experiments 1 and 2 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 Zhang et al. (2008) model rather than from similarity rankings from the web task. To derive these similarity estimates we again trained an AdaBoost-based classifier for each target class using color, shape, and texture features, then evaluated these same features for the distractors to compute target-distractor similarity. This resulted in the creation of two rank ordered lists, one indicating visual similarity to teddy bears and the other to butterflies. High-similarity trials for each target category were then constructed from distractors ranked in the top third of each 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.

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 Zhang et al. (2008) model is able to capture these relationships, then RTs should be the slowest on high-similarity trials, faster on mixed trials, and the fastest on low-similarity trials. These predictions were confirmed (Table 1). Search efficiency 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 ($t(11) = 5.77$, $p < .01$ for teddy bears and $t(11) = 6.50$, $p < .01$ for butterflies) and faster RTs in the low-similarity condition relative to the mixed condition ($t(11) = 5.15$, $p < .01$ for teddy bears and $t(11) = 6.22$, $p < .01$ for 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 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; 2.88±0.23 for butterflies) or low-similarity trials (1.01±0.19 for bears; 1.94±0.24 for butterflies).

The availability of high-, medium-, and low-similarity distractors in mixed condition displays again enabled us to look for direct oculomotor evidence for categorical search guidance. 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 – see Figure2B). Moreover, 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; $t(11) = 8.32$ for butterflies), indicating that initial saccades were guided towards target-similar distractors and away from target-dissimilar distractors.

We also analyzed initial saccade latencies 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 ($F(2,22) = 1.29$, $p = 0.30$) or bear targets ($F(2,22) = 0.76$, $p = 0.48$). The observed effects of visual similarity reflect actual changes in search guidance.

The conclusion from this experiment is clear. While the results of Experiment 1 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 2. 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. A goal of future work will be to determine what these categorical features are for a variety of real-world target classes. The present work constrains this goal by requiring that these features capture target-distractor visual similarity relationships.

Conclusions

Previous research had suggested that search is unguided to categorical targets (e.g., Castelhano et al., 2008; Wolfe et al., 2004). In light of 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). The present work extends this finding to non-target objects from categories that are visually similar to the target class. 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 (mixed condition; Figure 2). This pattern, when combined with the patterns of manual search efficiency found in the high-similarity and low-similarity distractor conditions (Table 1), provides strong converging evidence for categorical search guidance in our task. 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 guidance are visual, and can be well described by explicit similarity estimates regardless of whether these estimates were obtained from behavioral rankings using a web task (Experiment 1) or generated by a computer vision model of object category detection (Experiment 2). This too is a striking finding. The lengthy deliberations that accompanied the behavioral judgments, 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 remarkably stable across 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 Ryan Moore, Jonathan Ryan, and Arunesh Mittal for their help with data collection. This work was supported by NIH grant R01MH063748-07 to GJZ.

References

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

Castelhano, M. S., Pollatsek, A., & Cave, K.R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review, 15*(4), 795-801.

Chen, X., & Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research, 46*, 4118-4133.

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

Duncan, J., & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review, 96* (3), 433-458.

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.

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

Pomplun, M. (2006). Saccadic selectivity in complex visual search displays. *Vision Research*, *46*, 1886-1900.

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology*, *62 (10)*, 1904-1914.

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

Treisman, A. M. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance. 17* (3), 652-676.

Treisman, A. M., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 459-478.

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

Wolfe, J. M., 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.

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*(4), 787-835.

Zhang, W., Samaras, D., & Zelinsky, G. J. (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, G. J., & 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.