Eye can read your mind: Decoding gaze fixations to reveal categorical search targets

Gregory J. Zelinsky

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

Yifan Peng

Department of Computer Science, Stony Brook University, Stony Brook, NY

Dimitris Samaras

Department of Computer Science, Stony Brook University, Stony Brook, NY

Is it possible to infer a person's goal by decoding their fixations on objects? Two groups of participants categorically searched for either a teddy bear or butterfly among random category distractors, each rated as high, medium, or low in similarity to the target classes. Target-similar objects were preferentially fixated in both search tasks, demonstrating information about target category in looking behavior. Different participants then viewed the searchers' scanpaths, superimposed over the target-absent displays, and attempted to decode the target category (bear/butterfly). Bear searchers were classified perfectly; butterfly searchers were classified at 77%. Bear and butterfly Support Vector Machine (SVM) classifiers were also used to decode the same preferentially fixated objects and found to yield highly comparable classification rates. We conclude that information about a person's search goal exists in fixation behavior, and that this information can be behaviorally decoded to reveal a search target—essentially reading a person's mind by analyzing their fixations.

Introduction

Is it possible to infer a person's goal or intention by analyzing their eye fixations? There have been tremendous advances in techniques for the discovery of a person's perceptions, thoughts, or goals—a topic referred to in the popular press as "mind reading" (e.g., Grandoni, 2012). Most mind reading research has used a technique called *neural decoding* that infers a person's thoughts or percepts based on only their neural activity (e.g., Kay, Naselaris, Prenger, & Gallant, 2008; Tong, & Pratte, 2012). We introduce the technique of *behavioral decoding*, the use of fine-grained behavioral measures to similarly infer a person's thoughts, goals, or mental states.

We attempted behavioral decoding in the context of a visual search task, with the decoding goal being the category of a person's search target. The behaviors that we decoded were the objects fixated during search. The many fixations on non-target objects, or *distractors*, made during search are not random—the more similar these objects are to the target category, the more likely they are to be fixated first (Eckstein, Beutter, Pham, Shimozaki, & Stone, 2007) or fixated longer (Becker, 2011) compared to less target-similar objects. In this sense, categorical search obeys the same rules known to govern target-specific search (Alexander & Zelinsky, 2012), with one important caveat—the expression of target-distractor similarity relationships in fixation behavior is proportional to the specificity of the target cue (Malcolm & Henderson, 2009; Schmidt & Zelinsky, 2009). Specific cues, such as the picture previews used in most search studies (Wolfe, 1998), often lead to an efficient direction of gaze to the target (Zelinsky, 2008), but are highly unrealistic—outside of the laboratory one does not often know exactly how a target will appear in a search context. The vast majority of real-world searches are categorical, with the information used to cue a target being far less reliable and specific. It is not known whether these less specific categorical cues generate sufficient information in fixation behavior to decode the category of a search target.

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

doi: 10.1167/13.14.10 Received November 23, 2012; published December 12, 2013 ISSN 1534-7362

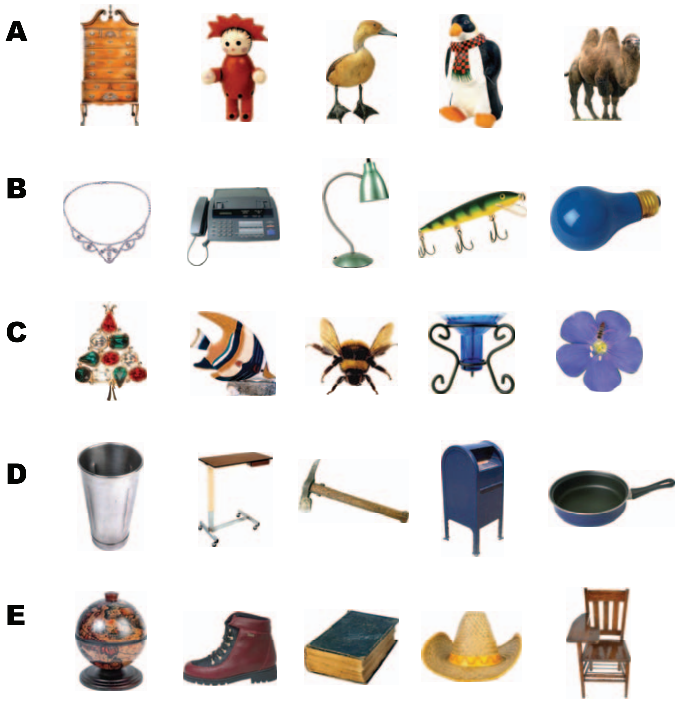

Figure 1. Representative examples of similarity-rated objects used as distractors. (A) Bear-similar objects. (B) Bear-dissimilar objects. (C) Butterfly-similar objects. (D) Butterfly-dissimilar objects. (E) Medium-similar objects.

This study answers four questions. First, does the cueing of different categorical targets produce different patterns of fixations during search? If such behavioral differences do not exist, behaviorally decoding the target category will not be possible. Second, are these behavioral differences sufficient for decoding—can one person read another person's mind, inferring their search target by decoding their fixations on random category distractors? Third, what are the limits of behavioral decoding in this task—is it possible to decode a target from the objects fixated on a single trial, and does decoding success vary with target-distractor similarity? Finally, how does a person's mind reading ability compare to that of computer vision classifiers? Machine decoders are valuable in that they make explicit the information from fixations used to decode search targets. To the extent that human and machine decoders agree, this agreement may suggest that human decoders use similar information. This latter goal makes our approach conceptually related to other quantitative techniques for capturing and revealing information embedded in human behavior (e.g., Baddeley & Tatler, 2006; Caspi, Beutter, & Eckstein, 2004; Eckstein & Ahumada, 2002; Gosselin & Schyns, 2001; Tavassoli, van der Linde, Bovik, & Cormack, (2009), with a difference being that our approach uses computer vision methods to decode from this information the identity of real-world object categories.

Experiment 1

To determine whether the features of distractors can attract and hold gaze during categorical search, two groups of participants searched for either a teddy bear or a butterfly target among random category distractors. Finding differences in the distractors preferentially fixated during these two search tasks would indicate that information about target category is embedded in a searcher's fixation behavior, and therefore available for decoding.

Method

Participants were sixteen students from Stony Brook University, half of whom searched for a teddy bear target and the other half a butterfly/moth target. Target category was designated by instruction; observers were not shown a specific target preview prior to each search display. Bear targets were adapted from Cockrill (2001, as described in Yang & Zelinsky, 2009); butterflies were selected from the Hemera object collection. Distractors were objects from random categories (also Hemera), and were selected based on their visual similarity to the bear and butterfly target categories using similarity estimates obtained from a previous study (Alexander & Zelinsky, 2011). Specifically, distractors were divided into five groups: bear-similar, bear-dissimilar, butterfly-similar, butterfly-dissimilar, and medium-similarity objects, with the last group rated as having intermediate visual similarity to both target classes. These objects will be referred to as *high*, *medium*, and *low* similarity distractors, where similarity is relative to their respective target category (Figure 1).

From these groups of similarity-rated objects we constructed four types of visual search displays. *Target-present* displays (TP, 44 trials) depicted either a bear or a butterfly target with three medium-similarity distractors. There were also three target-absent (TA) conditions. *High-medium* displays (TA-HM, 40 trials) depicted one high-similarity distractor and three medium-similarity distractors, and *high-medium-low* (TA-HML, 44 trials) displays depicted one low-similarity distractor, one high-similarity distractor, and two medium-similarity distractors. There were also *random* displays (TA-MED; 44 trials) in which all four distractors had medium visual similarity to the target. Importantly, search displays were crafted so as to have these similarity relationships apply to both target classes. For example, a TA-HM display would have one bear-similar and one butterfly-similar object, with each object rated as medium-similarity with respect to the non-target category. The same logic applied to the TA-HML condition. Except for the identity of the

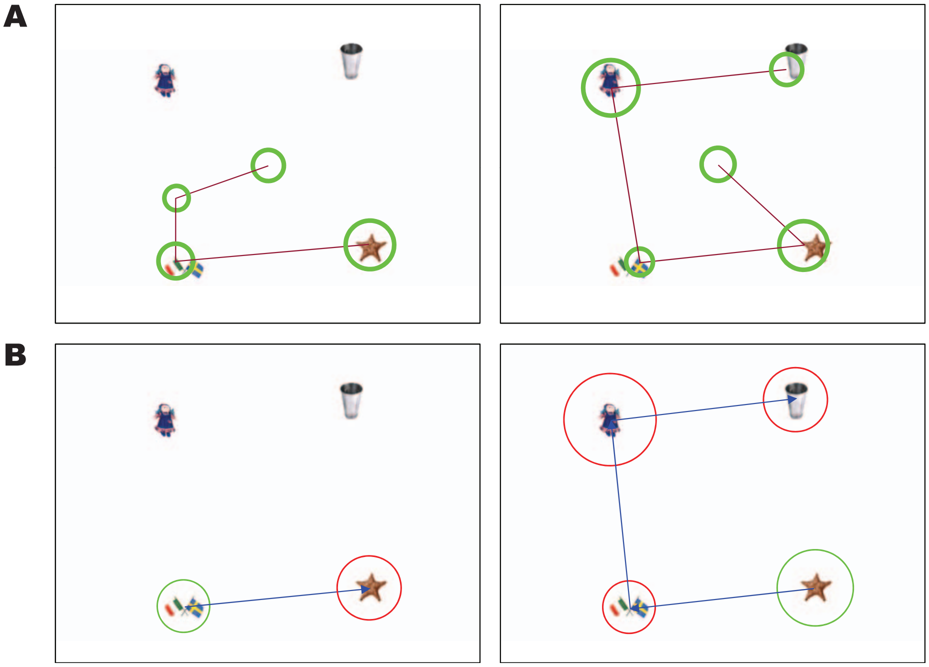

Figure 2. (A) Representative target-absent display presented to searchers in Experiment 1, with superimposed fixation behavior (not visible to participants). The locations of fixations are indicated by the green circles, with the circle diameters plotting their relative durations. Fixation order is indicated by the red lines, relative to a starting central position. The left display shows fixations from a butterfly searcher; the right display shows fixations from a bear searcher. Both trials are from the TA-HML condition and contain a high-similarity and low-similarity distractor with respect to each target category. (B) Corresponding displays presented to decoders in Experiment 2, with the fixation behavior from the searchers in (A) abstracted and visualized on the actual displays shown to participants. The first-fixated object was circled in green. Four fixed levels of circle size were used to indicate the time spent looking at an object. The longest-fixated object was enclosed by the largest circle, with up to three evenly-spaced smaller circles drawn around the other distractors that were fixated for a shorter time. When fewer than four distractors were fixated, circles were drawn starting with the smallest diameter and working upwards for as many levels as there were fixated objects.

target object on target-present trials, participants in the bear and butterfly search tasks therefore viewed the same search displays—the same distractors appearing in the same display locations. The use of identical target-absent displays in these two search tasks is critical to the decoding goals of this study, as differences in bear and butterfly classification rates cannot be attributed to differences in the composition of the search displays.

Upon fixating a central dot and pressing a button, a search display appeared and the task was to make a speeded presence/absent target judgment. The four objects in each search display were arranged on an imaginary circle (8.9° radius) surrounding central fixation (Figure 2). No object was repeated throughout the experiment, and each was normalized to subtend ~ 2.8° of visual angle. Eye position was sampled at 500 Hz using an EyeLink II eye tracker with default saccade detection settings. Participants gave their informed consent prior to the experiment, which was run in accordance with the ethical standards stated in the 1964 Declaration of Helsinki.

Results and discussion

Table 1 summarizes search performance, grouped by display type. Manual responses were accurate and

	Display type			
	TP	TA-HML	TA-HM	TA-MED
Errors (%)	4.0 (0.8)	1.7 (0.6)	3.0 (0.6)	0.9 (0.5)
Reaction time (ms)	697 (29.6)	863 (59.7)	897 (60.6)	829 (47.8)
Trials with no fixated object (%)	0.3 (0.2)	3.2 (1.5)	2.4 (1.3)	8.4 (3.5)

Table 1. Summary search performance by display type. *Note*: Values in parentheses indicate one standard error of the mean.

efficient, with faster responses in the target-present data compared to each target-absent condition, $t(7) \geq 4.33$, $p < 0.001$, paired group. No objects were fixated on 0.3–8.4% of the trials, depending on display type. These no-fixation trials and error trials were excluded from all subsequent analyses.

Does information about target category exist in the objects that people fixate as they search? Two fixation behaviors were explored, the first object fixated during search and the object that was fixated the longest. These measures capture different aspects of target-distractor confusion: first-fixated objects (FFO) indicate preferentially selected distractors to which search was guided (Malcolm & Henderson, 2009; Schmidt & Zelinsky, 2009); longest-fixated objects (LFO) indicate distractors that were difficult to reject due to their similarity to the target (Shen, Reingold, Pomplun, & Williams, 2003; Becker, 2011).[1] For trials having only one fixated object, the first fixated object would also be the longest fixated object. This eventuality occurred on 73% of the target-present trials and 32% of the target-absent trials. We will refer to the objects fixated either first or longest as *preferentially fixated objects*, and we assume that it is these objects that contain the most information available for decoding search targets.

Figure 3 shows the percentages of trials in which the preferentially fixated object was either the target or a target-similar distractor. Both the FFO and LFO measures showed a pronounced preference for the target on target-present trials—when a target was present in the display it was fixated first and longest far more than the 25% rate predicted by chance. Both results were expected. In the case of FFO, these data further support the existence of strong search guidance to categorically defined targets (Alexander & Zelinsky, 2011; Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009). In the case of LFO, longer looking times on the target reflect a searcher's tendency to hold gaze on that object while making their manual judgment. More interesting are the target-absent data, where distractors rated as visually similar to teddy bears (3A) and butterflies (3B) were preferentially fixated far above chance, and in some cases nearly as often as the actual targets. Bear-similar distractors in the TA-HML condition were also more likely to be fixated first, $t(7) = 11.25$, $p < 0.001$, and fixated longest, $t(7) = 3.85$, $p = 0.006$, compared to bear-similar distractors in the TA-HM condition. A similar pattern was found for butterfly search in the case of LFO, $t(7) = 4.20$, $p = 0.004$, but not FFO, $t(7) = 0.75$, $p = 0.48$. These patterns suggest that a target-dissimilar distractor in the TA-HML displays resulted in better guidance to the bear-similar distractor (FFO) and an easier rejection of this object after its fixation (LFO).

These data clearly indicate that fixations on non-target objects carry information about the category of a search target. This demonstration of information was particularly true for the LFO measure—people overwhelmingly looked longer at distractors rated as visually similar to the target category. Target-similar distractors were also fixated first, but searchers were less consistent in this behavior. On average, selection of the target or target-similar object was 26% more likely using LFO compared to FFO, a significant difference for all conditions, $t(7) \geq 4.01$, $p \leq 0.005$.

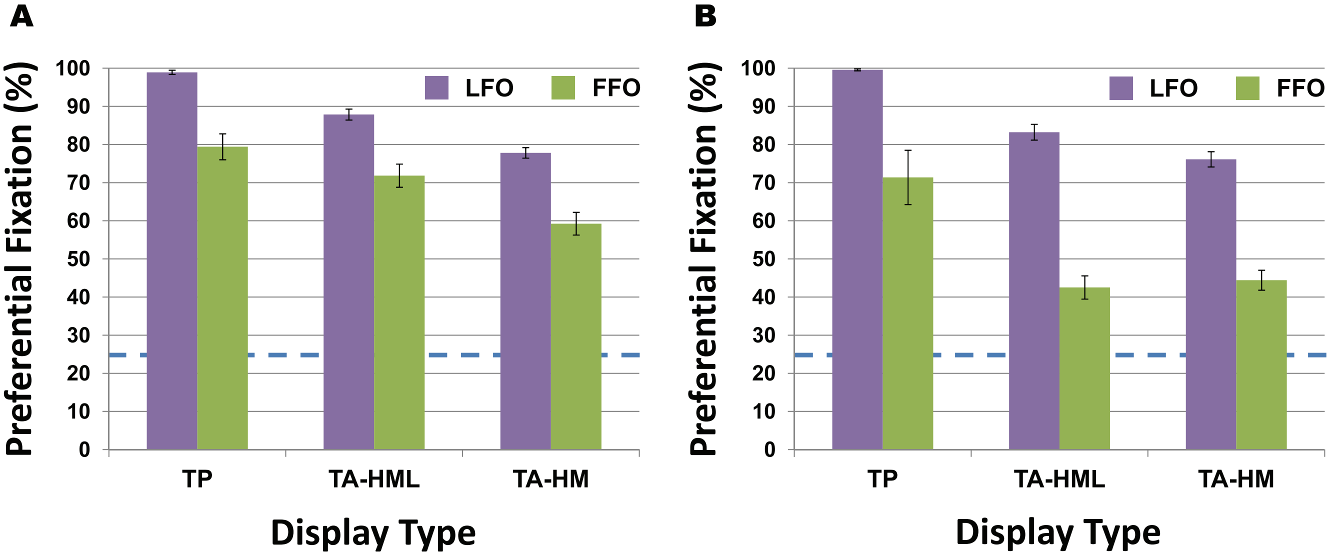

Figure 3. Percentages of correct trials in which the preferentially fixated object was a target (TP condition only) or a target-similar distractor (TA-HML or TA-HM conditions only). Data cannot be shown for the TA-MED display type because there were no target-similar distractors in that condition. Purple bars indicate longest-fixated objects (LFO); green bars indicate first-fixated objects (FFO). Dashed lines indicate chance. Error bars indicate one standard error of the mean. (A) Bear search task. (B) Butterfly search task.

Experiment 2

Experiment 1 showed that decoding is theoretically possible—that information about a target category is available from a searcher's fixation behavior. This demonstration does not mean, however, that people can extract this information and use it to infer a search target. To explore the potential for behavioral decoding, we asked whether one person can infer another person's search target based solely on their distractor fixations.

Method

Six new participants viewed the target-absent search displays from all 16 of the searchers from Experiment 1. Superimposed over each search display was a representation of a given searcher's fixation behavior on that trial (Figure 2B). Highlighted in each representation was the first-fixated distractor, around which a green ring was drawn, and the relative duration that each distractor was fixated in its initial viewing, indicated by the diameter of the circle around each object. The longest-fixated distractor was therefore indicated by the largest circle.[2] Each decoder judged, on a trial-by-trial basis, whether the searcher's viewing behavior on a given scanpath display indicated search for a bear or a butterfly target. They did this by making a 6-level confidence judgment for every trial, with a −3 indicating high confidence for one target category and a +3 indicating high confidence for the other. The assignment of positive/negative to bears/butterflies was counterbalanced over participants, and 0 ratings were not allowed—decoders had to choose a target category on every trial.

Decoders were familiarized with the bear and butterfly target categories by viewing the 88 target-present trials from the search experiment, half of which depicted a bear and the other half a butterfly. They were then trained to use the fixation information depicted in the scanpath displays by viewing the 24 practice trials from Experiment 1, with fixation behavior provided by a random bear/butterfly searcher. Decoders were instructed to attend to the first-fixated and longest-fixated distractor(s), but to use whatever information they deemed best when making their target judgments. Following training, each decoder viewed the 128 target-absent scanpath displays from each of the 16 searchers, yielding approximately 2,048 judgments per participant.[3] Importantly, the order of trials was determined by randomly sampling (without replacement) target-absent displays from the 16 searchers, meaning that decoders did not know from trial-to-trial whether they were viewing the behavior of a bear or butterfly searcher. The entire experiment lasted ∼ 3.5 hours, not including three enforced breaks. Participants gave their informed consent prior to the experiment, which was run in accordance with the ethical standards stated in the 1964 Declaration of Helsinki.

Results and discussion

Can one person decode another person's search target? To answer this question we first determined for every decoder whether each trial was correctly classified as a bear or butterfly search, ignoring the magnitude of the confidence judgments. We refer to this as the *by-trial* classification rate. Individual trial confidence ratings (−3 to +3) for each decoder were also grouped by searcher and averaged. If this average value was negative for a given searcher, and assuming a decoder was instructed to give negative ratings to bear judgments (depending on counterbalancing), that searcher would be classified as looking for a bear target. Doing this for every decoder and searcher gave us a *by-searcher* classification rate.

Figure 4A shows strikingly high by-searcher classification rates, even in the TA-MED condition where distractors were not visually similar/dissimilar to the search target. The target category for the eight bear searchers was classified with near perfect accuracy. The butterfly searchers were classified less well, but still far above chance (50%; $t(5) \geq 2.22$, $p \leq 0.04$, paired one-tailed) in all except the TA-HM condition. This advantage for bear over butterfly classification was significant for all conditions, $t(10) \geq 2.38$, $p \leq 0.04$, and is consistent with the greater availability of information about bear targets reported in Experiment 1.

By comparison, by-trial classification rates were much lower (Figure 4B). Although bear targets were classified above chance, $t(5) \geq 4.97$, $p \leq 0.002$ (paired one-tailed), rates were only in the 60–68% range. Butterfly classification rates were not significantly above chance, $t(5) \leq 1.71$, $p \geq 0.074$ (paired one-tailed). Bear classification success also varied with display type, $F(2, 15) = 4.41$, $p = 0.03$; rates were highest in the TA-HML condition, lower in the TA-HM condition, and lowest in the TA-MED condition, all $p \leq .05$. A similar but nonsignificant trend was found for butterfly search, $F(2, 15) = 1.75$, $p = 0.21$. These patterns of bear and butterfly classification parallel those shown in Figure 3, suggesting that decoders extracted and used the target-distractor similarity relationships reported in Experiment 1 when making their trial-by-trial target classification judgments. As for by-trial classification rates being lower than by-subject classification rates, this result was expected: by-trial decoding was limited to information from the 1-4

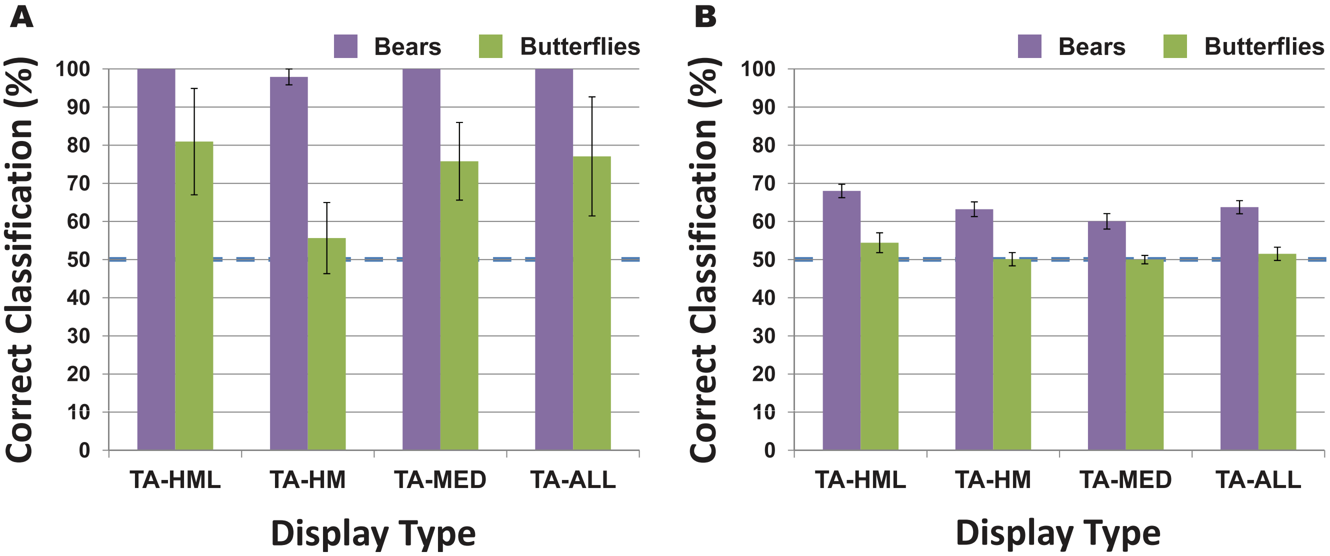

Figure 4. Percent accuracy of decoders (Experiment 2) classifying the target category of searchers (Experiment 1) grouped by display type and collapsed over display type (TA-ALL). Purple bars indicate the classification rate for bear searchers; green bars indicate the classification rate for butterfly searchers. The dashed line indicates chance. Error bars indicate one standard error of the mean. (A) By-searcher classification rates, decoding whether a given person was searching for a teddy bear or a butterfly. (B) By-trial classification rates, decoding whether the target on a given trial was a teddy bear or a butterfly.

fixations depicted in a given scanpath display; by-searcher decoding accumulated this information over dozens of trials to classify a person as a bear or butterfly searcher.

Experiment 3

Experiment 2 showed that human decoders were able to use information in a person's fixations to classify, with surprising accuracy, the category of their search target. In Experiment 3 we compared these human decoder classification rates to those of machine decoders; bear and butterfly discriminative models that quantified and decoded the same information from the Experiment 1 fixation behavior, to infer a searcher's target category.

Method

This was a purely computational experiment: no new behavioral data were collected. Computational methods can be divided into training and testing phases. Two linear-kernel Support Vector Machine (SVM; Chang, & Lin, 2001) classifiers were trained to discriminate teddy bears from non-bears and butterflies from non-butterflies. This procedure involves learning a classification boundary to separate exemplars of the target category (positive samples) from non-target exemplars (negative samples). Positive samples were 136 images of teddy bears and 136 images of butterflies; negative samples were 500 images of random objects unrated for visual similarity to the target categories. The same negative samples were used for training the bear and butterfly classifiers. Negative and positive samples were selected from the same databases of images used to assemble the Experiment 1 search displays, although none of the 772 objects used for training appeared as targets or distractors in the search tasks.

Objects were represented using two complimentary types of computer vision operators: SIFT features (Lowe, 2004) and color histogram features (see Swain & Ballard, 1991, for general details). The SIFT feature captures information about oriented edges and corners, but ignores the color information known to be important for guiding search (Hwang, Higgins, & Pomplun, 2009; Rutishauser, & Koch, 2007; Williams, 1966). The color histogram feature, defined in DKL color space (Derrington, Krauskopf, & Lennie, 1984), used 10 equally spaced bins to code color variability but ignored orientation information. Color and SIFT features were each normalized to be in the 0–1 range prior to SVM training, then concatenated and combined with the bag-of-words technique (Lazebnik, Schmidt, & Ponce, 2006) to create 1250-dimensional bear and butterfly classifiers, based on a vocabulary of 200 visual words spanning two layers of a spatial pyramid.

Testing used the identical target-present and target-absent search displays from Experiment 1. SIFT+color

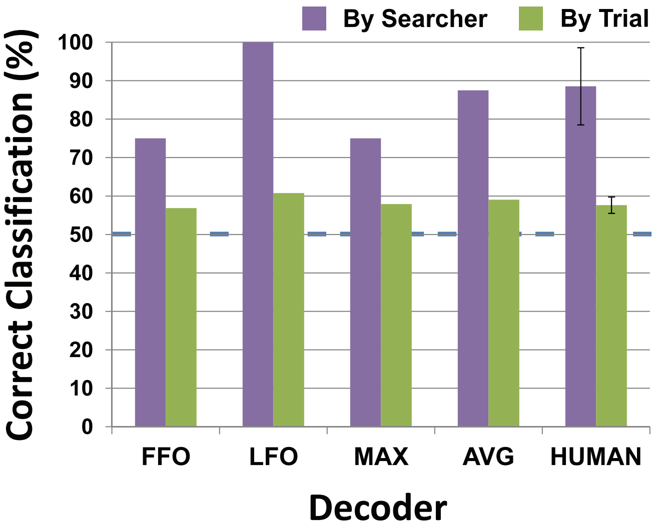

Figure 5. Classification rates for machine decoders (Experiment 3), collapsed over target category and target-absent display type and plotted with the corresponding rates from human decoders (Experiment 2). Purple bars indicate by-searcher classification rates; green bars indicate by-trial classification rates. FFO = First-fixated object; LFO = Longest-fixated object; MAX = Maximum of classifier response from FFO and LFO; AVG = Average classifier response from FFO and LFO. The dashed line indicate chance. Error bars around the behavioral mean indicate a 95% confidence interval.

features were extracted from one preferentially fixated object from each search display. Preferential fixation was again defined as the object fixated first (FFO) or longest (LFO), as in Experiment 1. Our testing of the bear and butterfly classifiers differed from standard methods in one key respect: rather than attempting to recognize positive from negative samples of a target class, we obtained from each classifier estimates of the visual similarity between the preferentially fixated object during search and the bear and butterfly categories learned during training. Similarity estimates were based on distances from the SVM classification boundaries, where larger distances indicate more confident classification decisions. Separate similarity estimates were obtained for the first-fixated and longest-fixated objects (when they were different), yielding four similarity estimates per trial: FFO-bear, FFO-butterfly, LFO-bear, and LFO-butterfly. These similarity estimates were then used to classify the first-fixated and longest-fixated objects as indicating bear or butterfly search.

In addition to FFO and LFO, we also explored two methods of combining information from the first-fixated and longest-fixated objects on a trial-by-trial basis for classification. The MAX method classified a trial as a bear or butterfly search based on the classifier having the larger distance to its decision boundary, irrespective of whether this distance was from the first-fixated or longest-fixated object. The AVG method made the bear or butterfly classification based on the averaged FFO and LFO distance: if this averaged distance was greater for the bear classifier, that trial was classified as a bear search. Each of the Experiment 1 participants was also classified as a bear or butterfly searcher based on the mean distances averaged over trials using each of these four classification methods. Note that this by-searcher classification is comparable (as much as possible) to the method described in Experiment 2, where confidence ratings from human decoders were also averaged over trials.

Results and discussion

How closely do target classification rates from machine decoders compare to those from human decoders? Figure 5 shows by-searcher and by-trial classification rates for each machine decoding method plotted with the corresponding human decoder rates from Experiment 2 (see Table 2 for by-trial rates grouped by display type). There are several noteworthy patterns. First, by-subject decoding rates were high: the target of a person's search was classified far above chance using only the visual features of preferentially fixated distractors. By-trial classification rates were again much lower, although on par with the classification performance from human decoders. Second, there was considerable variability between methods. Target category was correctly classified for each of the 16 searchers using the LFO method, but classification rates dropped to 75% using the FFO and MAX methods.[4] This advantage for the LFO method reaffirms the existence of search-relevant information in the longest-fixated objects, and demonstrates that this information can be extracted using purely automated techniques from computer vision. Third, the AVG method best described human decoder performance, with by-subject classification rates matching almost perfectly (87.5% vs. 88.5%). This congruency raises the intriguing possibility that human decoders

	Display type			
Decoder	TP	TA-HML	TA-HM	TA-MED
FFO	86.1	58.8	56.1	55.4
LFO	100	66.5	58.8	55.5
MAX	100	61.8	56.5	55.1
AVG	98.7	62.4	57.8	56.4
Human	n/a	61.2 (±4.4)	56.7 (±2.0)	55.0 (±1.5)

Table 2. By-trial classification rates (%) by decoder and display type. *Note*: Values in parentheses indicate one standard error of the mean.

may have extracted information from both the first-fixated and the longest-fixated distractors (and perhaps others), then averaged this information to make their classification decisions— despite this averaging being suboptimal compared to LFO alone.

It is also possible to analyze the bear and butterfly classifiers to estimate the relative contributions that color and SIFT features made to the classification decisions. Underlying our use of these two types of features was an assumption that each would be helpful in classifying the target category, but this need not be the case. It may be that only one of these features was solely or disproportionately responsible for classification. This would be in line with suggestions that color, because it is relatively immune to information loss in the visual periphery, is of singular importance in the preferential fixation of targets during search (e.g., Rutishauser & Koch, 2007). Dominance by color features might also explain the better classification observed for teddy bears than for butterflies. If the teddy bear category had a narrower distribution of color features than the butterfly category, this difference may have made it easier for the bear classifier to find color features that discriminate this category from the random category objects used as distractors.

To address this possibility we computed the proportions of color and SIFT features over ten bins, where each bin indicates the top *x%* of features that were weighted most strongly by the bear and butterfly SVM classifiers (Figure 6A and 6B).[5] The leftmost bar in each plot indicates the features weighted in the top 10%, and the proportions of these features that were color and SIFT. The rightmost bars are identical and reflect the unequal numbers of SIFT (1000) and color (250) features used by the classifiers, with the other bins describing intermediate weightings to illustrate the changing contributions of color and SIFT features. If color features dominated classification, one might expect them to appear in large number among the topmost weighted features. Indeed, if classification was based solely on color, then all of these features should have been color features. This pattern was clearly not found. Rather, the opposite trend emerged—the proportions of color features among the topmost weighted features was *less* than their proportion in the total number of features. Contrary to the prediction that color was more important than SIFT features for classification, this analysis suggests that it is the SIFT features that were the more important. However, the bear classifier did rely more on color features than the butterfly classifier, as can be seen by comparing the several leftmost bars in each plot. Whether this pattern might explain in part the better classification performance for bears, or whether human decoders were adopting similar weightings on comparable features when making their classifications, we cannot say.

The previous analysis gauged the relative importance of color and SIFT features in the teddy bear and butterfly classifiers, but it did not address the role that these features played in decoding the target category from first-fixated and longest-fixated distractors. Given the demonstrated importance of color in guiding search to targets and target-similar objects (Hwang, et al., 2009; Motter & Belky, 1998; Rutishauser & Koch, 2007; Williams, 1966), and that guidance is commonly measured in terms of first-fixated objects (Yang & Zelinsky, 2009; Alexander & Zelinsky, 2011, 2012; Maxfield & Zelinsky, 2012), it may be that color features play a larger role in FFO decoding success compared to LFO decoding. To directly test this hypothesis, we borrowed a technique from Zhang, Samaras, and Zelinsky (2008) and trained four new SVM classifiers using the same SIFT+color features and computational procedures described in the Experiment 3 Methods section, with one exception. Rather than using images of actual teddy bears and butterflies as positive samples for training, these classifiers were trained on the distractors that were fixated either first or longest in Experiment 1. For example, the new FFO-bear classifier was trained by finding the distractors from Experiment 1 that were fixated first in the teddy bear search task and using these objects as positive training samples; the distractors that were not fixated first were used as negative samples. Doing this training for both target categories and methods of preferential fixation gave us the four classifiers used in this analysis: FFO-bear, LFO-bear, FFO-butterfly, and LFO-butterfly. We then analyzed the proportion of color and SIFT features selected by these classifiers as a function of their weighting in the classification, just as we did in Figure 6A and 6B.

These results are shown in Figure 6, C–F. There are several patterns to note. First, these new classifiers again overwhelmingly selected SIFT features over color features, and in roughly the same proportions as those reported in Figure 6A and 6B. This consistency is interesting in that it suggests that these features are weighted similarly regardless of whether training was done using actual target exemplars or merely objects that were visually similar to the target classes (replicating Zhang, et al., 2008). Second, and as hypothesized, comparing the leftmost bars in Figure 6C and 6E reveals that color features were indeed more important for FFO bear classification (6C) compared to LFO bear classification (6E). This pattern is consistent with the literature suggesting that color is important for guiding search (Hwang, et al., 2009; Motter & Belky, 1998; Rutishauser & Koch, 2007; Williams, 1966). Lastly, there was no meaningful difference in the proportion of color features selected by the FFO-butterfly (6D) and LFO-butterfly (6F) classifiers, and in general the proportion of color

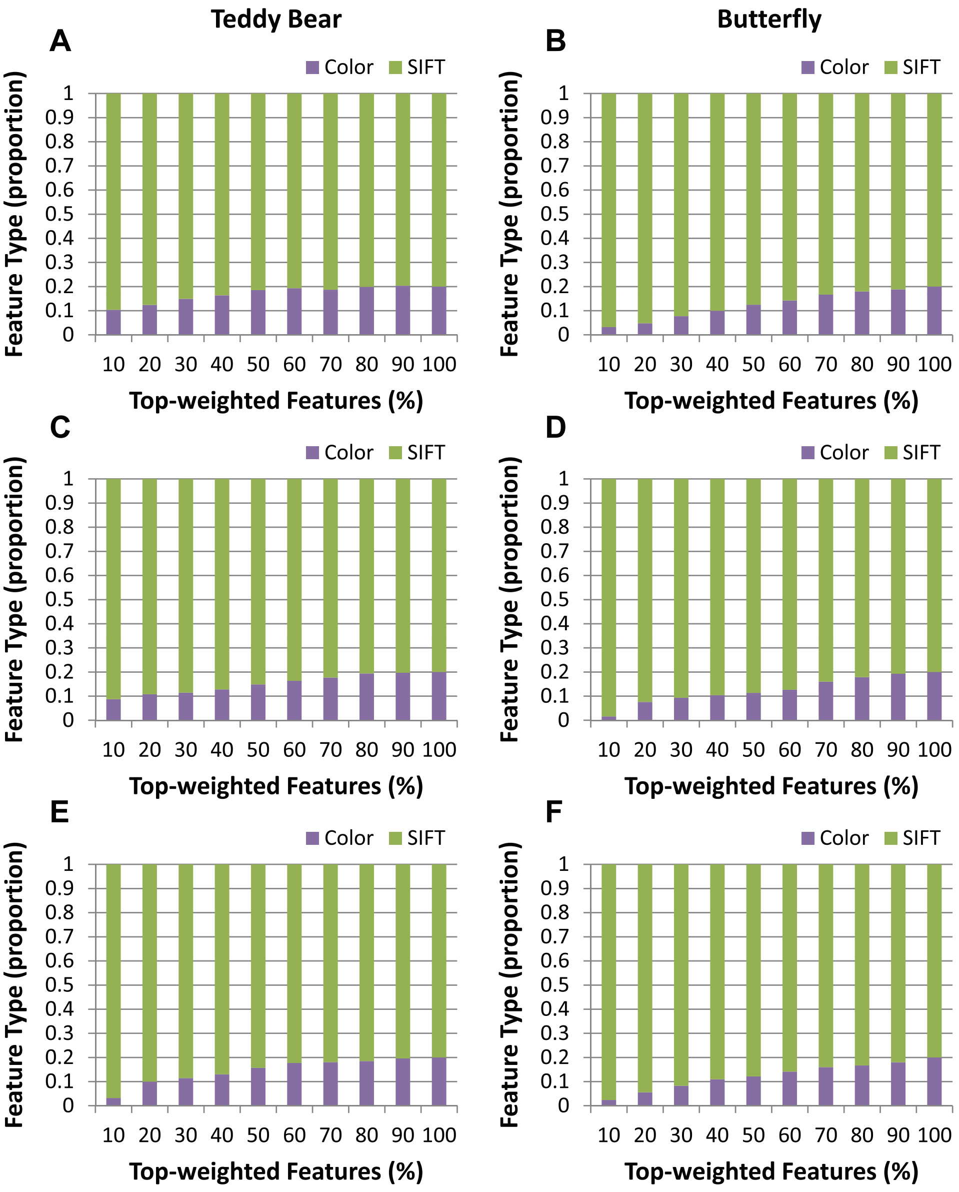

Figure 6. Proportions of color (purple) and SIFT (green) features used by the SVM-based classifiers, plotted as a function of the weight given to these features. Classifier weight is binned and shown along the *x*-axis, where each bin indicates a percentage of the topmost weighted features. The leftmost bars therefore show the proportions of color and SIFT features in the top 10% of features receiving the strongest weights, whereas the 20% bins show the proportions of these features weighted in the top 20%. The rightmost, 100%, bars indicate the proportions of color and SIFT features irrespective of weight in the 1,250-dimensional classifiers, based on 250 color features and 1,000 SIFT features. (A) Teddy bear classifier trained on teddy bears and non-bears. (B) Butterfly classifier trained on butterflies and non-butterflies. (C) Classifier trained on the first-fixated distractors from the Experiment 1 teddy bear search task (FFO-bear). (D) Classifier trained on the first-fixated distractors from the Experiment 1 butterfly search task (FFO-butterfly). (E) Classifier trained on the longest-fixated distractors from the Experiment 1 teddy bear search task (LFO-bear). (D) Classifier trained on the longest-fixated distractors from the Experiment 1 butterfly search task (LFO-butterfly).

features in the top-ranked bins was very low. This result is consistent with the suggestion that color was less useful in classifying butterflies compared to teddy bears, undoubtedly due to the fact that most teddy bears are some shade of brown whereas the class of butterflies is highly variable in color. Collectively, the analyses shown in Figure 6 suggest that color features were less useful than SIFT features in the Experiment 3 results, although it is not possible to rule out the existence of a small number of highly distinctive color features contributing to classification and decoding performance in our tasks.[6]

General discussion

In this study we demonstrated that information not only exists in the fixations that people make as they search, but that it is possible to behaviorally decode these fixations on distractors to infer the category of a person's search target. Human decoders were able to classify a searcher's target with an average accuracy of 89%. This exceptional decoding rate stood in sharp contrast to the by-trial rates, which hovered at or near chance. The implication of this finding is that little information for decoding exists on an individual trial, but that it is possible to accumulate this information over trials to achieve outstanding decoding success.

Decoding rates also depended on the method of selecting the preferentially fixated distractor used for decoding. Although both first-fixated and longest-fixated distractors carried information about the target category, searchers expressed fixation preference more consistently in their gaze durations. Target-similar distractors were more likely to be fixated longer than first, and this preference translated into additional information and a decoding advantage for LFO over FFO. This suggests that distractor rejection times may be more useful to the behavioral decoding of search targets than the information about preferential selection typically associated with search guidance.

Machine vision decoding mirrored closely the performance of human decoders. Not only were by-trial classification rates comparable between the two, they both showed subtle effects of target-distractor similarity. Although we cannot rule out the possibility that human decoders might also have used semantic information when making their target classification judgments, this congruity suggests that the information extracted and used for decoding was indeed visual. By-searcher machine decoder rates varied more with the decoding method, with the LFO method yielding a perfect classification of a searcher's target category. However, the machine decoding method that best described human decoder performance averaged classification distance over the first-fixated and longest-fixated objects. To the extent that human decoders did the same, this finding would suggest that people are very adept in integrating different types of information about the target category over multiple fixations when making their classification decisions.

Our findings also have implications for search theory. The fact that decoding was possible using the first-fixated object (albeit less so than when using the longest-fixated object) quantitatively proves that search can be guided to categorically-defined targets, an issue that had been debated in the search literature (Castelhano, Pollatsek, & Cave, 2008; Schmidt & Zelinsky, 2009; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004; Yang & Zelinsky, 2009)—if categorical guidance did not exist, the decoding of search targets using first-fixated objects should have been at chance. Moreover, the fact that a searcher's fixation preferences and a decoder's classification success varied directly with target-distractor similarity means that these behaviors are sensitive to arguably subtle categorical visual similarity relationships. Although a pumpkin looks very different from a teddy bear, the visual similarity between these objects was sufficient to attract and hold gaze on the distractor, and in so doing created information that decoders were able to extract and use to infer the target category.[7] We also evaluated the contribution of color and texture (as approximated by SIFT) features in mediating this categorical search behavior by analyzing the proportions of each feature that were most heavily weighted by SVM classifiers. Although this suggested a clear preference for texture over color features, an analysis focused on color showed that bear targets could be better decoded from first-fixated distractors than longest-fixated distractors, and that color features likely played a role in this difference. This finding suggests that SVM-based classifiers might be useful in characterizing the discriminative features used to code different categories of search targets.

How surprising are these findings? Certainly, sophisticated decoding techniques would not be needed to determine whether a searcher was looking for a red or a green T among red and green Ls, as the color of the fixated distractors would easily reveal the target (e.g., Motter & Belky, 1998). At the other end of the decoding spectrum, Greene, Liu, and Wolfe (2012) attempted to decode from eye movements the viewing instruction given to observers and concluded that this classification was not possible. There are clearly some tasks that can be decoded from fixation behavior and others that cannot.

The decoding task used in the present study likely falls near the middle of this range. The fact that by-trial classification rates were at or near chance for both human and computer decoders attests to the difficulty of this task, which is undoubtedly due in part to the categorical designation of the targets. An interesting direction for future work will be to compare the decoding rates reported here to those from a search task using a picture preview of the target, where knowledge of the target's exact appearance should yield a stronger guidance signal (Yang & Zelinsky, 2009). We also adopted in this study several design constraints that likely increased the preferential fixation of target-similar distractors: the use of only two target classes, the assessment of fixation preference relative to only four distractors, and the fact that search displays consisted of segmented objects rather than objects embedded in scenes. It will be interesting to learn how decoding success changes as these constraints are

systematically relaxed. More generally, future work will also combine fixation behavior with other behavioral measures in an attempt to improve decoding success. There is a universe of questions to be asked using this exciting blend of psychology and computer science, and the present study describes just one point in this space of tasks, mental states, and perhaps even clinical conditions that are potentially amenable to behavioral decoding.

Keywords: decoding, fixation duration, categorical search, computer vision, classification

Acknowledgments

This work was supported by NSF grant IIS-1111047 and NIMH grant R01-MH063748 to GJZ.

Commercial relationships: none.

Corresponding author: Gregory J. Zelinsky.
Email: Gregory.Zelinsky@stonybrook.edu.
Address: Department of Psychology, Stony Brook University, Stony Brook, NY.

Footnotes

[1] Our LFO measure has sometimes been referred to as first-pass gaze duration (Henderson, Weeks, & Hollingworth, 1999), the sum of all fixation durations on an object during its initial viewing.

[2] In pilot work we determined that showing this slightly abstracted representation of a searcher's viewing behavior was preferable to showing the actual scanpath detailing each fixation (e.g., Figure 2A). This preference was motivated by the fact that information about the first-fixated object and the longest fixated object is often difficult to discern from scanpaths—the first fixation is not always on an object, and sometimes small differences in viewing time separate the longest-fixated object from the others.

[3] Trials in which a searcher did not fixate at least one object, or responded incorrectly, were excluded, so the actual number of displays viewed by each Experiment 2 participant was 1,914.

[4] Classification rates for MAX were lower than for LFO because FFO, a contributor to MAX, often yielded larger distances, even for incorrect classifications.

[5] It is unclear whether it is correct or meaningful to evaluate the relative proportions of features selected by an SVM classifier as we are doing here. SVMs learn a classification boundary, and the relationship between this boundary and the individual weights of any given feature is not straightforward; it may be the case that the same classification boundary would have formed irrespective of the specific weights set on a given feature. However, to the extent that it can be assumed that the most heavily weighted features are contributing the most to the formation of a classification boundary then it is meaningful to group feature type by classification weight and to characterize the data in terms of the relative proportions of each feature type.

[6] Although fewer color features in general were strongly weighted by the classifiers, it may be that these few selected features were highly discriminative and therefore instrumental in determining classification and decoding success. We thank Marc Pomplun for this observation.

[7] Note that these claims about categorical guidance do not imply the existence and use of semantic features of the target category to guide search. Indeed, our focus on visual features in this study, and our finding that these features are sufficient to guide search, suggest that the many demonstrations of categorical guidance in the literature may have very little to do with actual semantic properties of the target category. It is our position that any visual feature that has been learned to discriminate a target category from non-target categories, whether this feature is shape, texture, or color, is by definition a categorical feature.

References

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

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

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research, 46,* 2824–2833.

Becker, S. I. (2011). Determinants of dwell time in visual search: Similarity or perceptual difficulty. *PLoS ONE, 6*(3).

Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proceedings of the National Academy of Sciences, USA, 101(35),* 13086–13090.

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.

Chang, C. C., & Lin, C. J. (2001). LIBSVM: A library for support vector machines (Version 3.11) [Software]. Retrieved from http://www.csie.ntu.edu.tw/cjlin/libsvm.

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

Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of Physiology, 357*(1), 241–265.

Eckstein, M. P., & Ahumada, A. J. (2002). Classification images: A tool to analyze visual strategies. *Journal of Vision, 2*(1):i, http://www.journalofvision.org/content/2/1/i, doi:10.1167/2.1.i. [PubMed] [Article]

Eckstein, M. P., Beutter, B. R., Pham, B. T., Shimozaki, S. S., & Stone, L. S. (2007). Similar neural representations of the target for saccades and perception during search. *Neuron, 27,* 1266–1270.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition. *Vision Research, 41,* 2261–2271.

Grandoni, D. (2012, July 13). "Mind Reading" helmet reads EEG to see if wearer recognizes faces or objects. *The Huffington Post*. Available from http://www.huffingtonpost.com/2012/07/13/mind-reading-helmet-eeg_n_1671210.html.

Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns. *Vision Research, 62,* 1–8.

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

Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, *9*(5):25, 1–18, http://www.journalofvision.org/content/9/5/25, doi:10.1167/9.5.25. [PubMed] [Article]

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature, 452,* 352–355.

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

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

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9*(11):8, 1–13, http://www.journalofvision.org/content/9/11/8, doi:10.1167/9.11.8. [PubMed] [Article]

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

Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research, 38,* 1805–1815.

Rutishauser, U., & Koch, C. (2007). Probabilistic modeling of eye movement data during conjunction search via feature-based attention. *Journal of Vision, 7*(6):5, 1–20, http://www.journalofvision.org/content/7/6/5, doi:10.1167/7.6.5. [PubMed] [Article]

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.

Shen, J., Reingold, E. M., Pomplun, M., & Williams, D. E. (2003). Saccadic selectivity during visual search: The influence of central processing difficulty. In J. Hyona, R. Radach, & H. Deubel (Eds.), *The mind's eyes: Cognitive and applied aspects of eye movement research* (pp. 65–88). Amsterdam: Elsevier Science.

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

Tavassoli, A., van der Linde, I., Bovik, A. C., & Cormack, L. K. (2009). Eye movements selective for spatial frequency and orientation during active visual search. *Vision Research, 49,* 173–181.

Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology, 63,* 483–509.

Williams, L. G. (1966). The effects of target specification on objects fixated during visual search. *Acta Psychologica, 27,* 355–360.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–71). London: University College London Press.

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. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society,* (pp. 1856–1861) Washington, DC.