Visual Search is Guided to Categorically Defined Targets

Hyejin Yang and **Gregory J. Zelinsky**
Department of Psychology, Stony Brook University

Abstract

To determine whether categorical search is guided we had subjects search for teddy bear targets either with a target preview (specific condition) or without (categorical condition). Distractors were random realistic objects. Although subjects searched longer and made more eye movements in the categorical condition, targets were fixated far sooner than was expected by chance. By varying target repetition we also determined that this categorical guidance was not due to guidance from specific previously viewed targets. We conclude that search is guided to categorically-defined targets, and that this guidance uses a categorical model composed of features common to the target class.

Keywords

categorical guidance; eye movements; categorization; object detection; visual attention

Visual search is one of our most common cognitive behaviors. Hundreds of times each day we seek out objects and patterns in our environment in the performance of search tasks. Some of these are explicit, such as when we scan the shelves for a particular food item in a grocery store. Other search tasks are so seamlessly integrated into an ongoing behavior that they become all but invisible, such as when gaze flicks momentarily to each ingredient when preparing a meal (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003).

Such a widely used cognitive operation requires a highly flexible method for representing targets, a necessary first step in any search task. In many cases a search target can be described in terms of very specific visual features. When searching for your car in a crowed parking lot or your coffee cup in a cluttered room, relatively specific features from these familiar objects can be recalled from long term memory, assembled into a working memory description of the target, and used to guide your search (Wolfe, 1994; Zelinsky, 2008). However, in many other cases such an elaborated target description is neither possible nor desirable. Very often we need to find any cup or any pen or any trash bin, not a particular one. In these cases different target-defining features are required, as the features would need to represent an entire class of objects and cannot be tailored to a specific member. How does the search for such a categorically defined target differ from the search for a specific member of a target class?

Although several studies have used categorically defined targets in the context of a search task (e.g., Bravo & Farid, 2004; Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, in press; Fletcher-Watson, Findlay, Leekam, & Benson, 2008; Foulsham & Underwood, 2007; Henderson,

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

Weeks, & Hollingworth, 1999; Mruczek and Sheinberg, 2005; Newell, Brown, & Findlay, 2004; Torralba, Oliva, Castelhano, & Henderson, 2006), surprisingly few studies have been devoted specifically to understanding categorical visual search (Bravo & Farid, 2009; Castelhano, Pollatsek, & Cave, 2008; Schmidt & Zelinsky, 2009). Early work on categorical search used numbers and letters as target classes. For example, Egeth, Jonides, and Wall (1972) had subjects search for a digit target among a variable number of letter distractors, and found nearly flat target present and target absent search slopes (see also Jonides & Gleitman, 1972). Brand (1971) also showed that the search for a digit among letters tended to be faster than the search for a letter among other letters. The general conclusion from these studies was that categorical search is not only possible, but that it can be performed very efficiently, at least in the case of stimuli having highly restrictive feature sets (Duncan, 1983).

Wolfe, Friedman-Hill, Stewart, and O'Connell (1992) attempted to identify some of the relevant categorical dimensions that affect search in simple visual contexts. They found that the search for an oriented bar target among heterogeneous distractors could be very efficient when the target was categorically distinct in the display (e.g., the only "steep" or "left leaning" item), and concluded that categorical factors can facilitate search by reducing distractor heterogeneity via grouping. Wolfe (1994) later elaborated on this proposal by hypothesizing the existence of categorical features, and incorporating these features into his influential Guided Search Model (GSM). According to GSM, targets and search objects are represented categorically, and it is the match between these categorical representations that generates the top-down signal used to guide attention in a search task. However, left unanswered from this work was whether this evidence for categorical guidance would extend to more complex object classes in which the categorical distinctions between targets and distractors are less apparent.

Levin, Takarae, Miner, and Keil (2001) directly addressed this question and provided the first evidence that categorical search might be possible for visually complex object categories. Subjects viewed 3-9 line drawings of objects and were asked to search for either an animal target among artifact distractors or an artifact target among animal distractors. They found that both categorical searches were very efficient, particularly in the case of the artifact search task. Levin and colleagues concluded that subjects might learn the features distinguishing targets from distractors for these two object classes (e.g., rectilinearity and curvilinearity), then use these categorical features to efficiently guide their search.

More recent work suggests that categorical guidance may in fact be quite limited in tasks involving fully realistic objects (Vickery, King, & Jiang, 2005; Wolfe, Horowitz, Kenner, Hyle, and Vasan, 2004, Experiments 5-6). Using a target preview, subjects in the Wolfe et al. study were shown either an exact picture of the target (e.g., a picture of an apple), a text label describing the target type (e.g., "apple"), or a text label describing the target category (e.g., "fruit"). They found that search was most efficient using the picture cue, less efficient using a type cue, and least efficient using a categorical cue (see also Schmidt & Zelinsky, 2009). Contrary to the highly efficient guidance reported in the Levin et al. (2001) study, these results suggest that categorical guidance may be weak or non-existent for search tasks using common real-world object categories.

At least two factors may have contributed to the discrepant findings from previous categorical search studies. First, most of these studies measured categorical guidance exclusively in terms of manual search efficiency. However, this measure makes it difficult to cleanly separate actual guidance to the target from decision processes needed to reject search distractors (Zelinsky & Sheinberg, 1997). For example, it may be the case that subjects were very efficient in rejecting animal distractors in the Levin et al. (2001) study, thereby resulting in shallow search slopes for artifact targets (see also Kirchner & Thorpe, 2006). Given that object verification times vary widely for categorically-defined targets (Castelhano et al., 2008), differences in search

efficiency reported across studies and conditions might reflect different rejection rates for distractors, and have very little to do with actual categorical guidance. Second, previous studies using a categorical search task have invariably repeated stimuli over trials. For example, Wolfe et al. (2004, Experiment 5) used only 22 objects as targets, despite having 600 trials in their experiment. Such reuse of stimuli might compromise claims of categorical search, as subjects could have retrieved instances of previously viewed targets from memory and used these as search templates. To the extent that object repetition speeds categorical search (Mruczek & Sheinberg, 2005), differences in search efficiency between studies might be explained by different object repetition rates.

By addressing both of the above-described concerns, the present study clarifies our capacity to guide search to categorically defined targets. In Experiment 1 we removed the potential for target and distractor repetition to affect categorical search by using entirely new targets and distractors on every trial. If stimuli are not reused from trial to trial, no opportunity for object-specific guidance would exist. In Experiment 2 we explicitly manipulated target repetition so as to determine whether guidance results from the categorical representation of target features or from previously viewed targets serving as specific templates. Eye movements were monitored and analyzed in both experiments so as to separate actual categorical guidance from decision factors relating to distractor rejection. To the extent that search is guided to categorical targets, we expect these targets to be fixated preferentially by gaze (e.g., Chen & Zelinsky, 2006; Schmidt & Zelinsky, 2009). However, finding no preference to fixate targets over distractors would suggest that differences in search efficiency are due to different rates of distractor rejection or target verification under categorical search conditions.

Experiment 1

Can search be guided to categorical targets? To answer this question we had subjects search for a teddy bear target among common real-world objects under specific and categorical search conditions. Following Levin et al. (2001), if categorical search is highly efficient we would expect relatively shallow manual search slopes, perhaps as shallow as those in the specific search condition where the target is designated using a preview. We would also expect categorically-defined targets to be acquired directly by gaze, again perhaps as directly as those under target specific conditions. However, if categorical descriptions cannot be used to guide search to a target (Castelhano et al., 2008; Vickery et al., 2005; Wolfe et al., 2004), we would expect steep manual search slopes in the categorical condition, and a chance or near chance probability of looking initially to the categorical target.

Method

Participants—Twenty-four students from Stony Brook University participated in the experiment for course credit. All had normal or corrected to normal visual acuity, by self report, and were naïve to the goals of the experiment.

Stimuli & Apparatus—Targets were 198 color images of teddy bears from *The teddy bear encyclopedia* (Cockrill, 2001). Of these objects, 180 bears were used as targets in the search task, and 18 bears were used as targets in practice trials. The distractors were 2,475 color images of real-world objects from the Hemera Photo Objects Collection (Gatineau, Quebec, Canada). Of these objects, 2,250 were used as distractors in the search task, and 225 were used in practice trials. No object, target or distractor, was shown more than once during the experiment. All were normalized to have the same bounding box area (8,000 pixels), where a bounding box is defined as the smallest rectangle enclosing an object. Normalizing object area roughly equated for size, but precise control was not possible given the irregular shape of real-world objects. Consequently, object width varied between 1.12° and 4.03°, and object height varied between 1.0° and 3.58°. Figure 1 shows representative targets and distractors.

Objects were arranged into 6, 13, and 20-item search displays, which were presented in color on a 19-inch flat screen CRT monitor at a refresh rate of 100 Hz. A custom-made program written in Visual C/C++ (v. 6.0) and running under Microsoft Windows XP was used to control the stimulus presentation. Items were positioned randomly in displays, with the constraints that the minimum center-to-center distance between objects, and the distance from center fixation to the nearest object, was 180 pixels (about 4°). Approximate viewing angle was 26° horizontally and 20° vertically. Head position and viewing distance (72 cm) were fixed with a chinrest, and all responses were made with a Game Pad controller attached to the computer's USB port. Eye position was sampled at 500 Hz using the EyeLink II eye tracking system (SR Research Ltd.) with default saccade detection settings. Calibrations were not accepted until the average spatial error was less than .49° and the maximum error was less than .99°.

Design—The 180 experimental trials per subject were evenly divided into 2 target presence conditions (present/ absent) and 3 set size conditions (6/ 13/ 20), leaving 30 trials per cell of the design. The type of search, specific or categorical, was a between-subjects variable. Half of the subjects were shown a preview of a specific target bear at the start of each trial (specific search), the other half were instructed to search for a non-specific teddy bear (categorical search). The search displays viewed by these two groups of subjects were identical (i.e., the same targets and distractors in the same locations); the only difference between these groups was that subjects in the specific condition were searching for a particular target.

Procedure—The subject's task was to determine, as quickly and as accurately as possible, the presence or absence of a teddy bear target among a variable number of real-world distractors. Each trial began with subjects looking at a central fixation point and pressing a "start" button, which also served to drift correct the eye tracker. In the specific search condition a preview of the target teddy bear was displayed for 1 second, followed by the search display. In the categorical search condition subjects were instructed at the start of the experiment to search for any teddy bear, based on the bears viewed during the practice trials and their general knowledge of this category; there were no target previews. Target present judgments were registered using the left trigger of the Game Pad, target absent judgments were registered using the right trigger. Accuracy feedback was provided after each response. The experiment consisted of one session of 18 practice trials and 180 experimental trials, lasting about 30 minutes.

Results and Discussion

Manual data—Error rates were generally low. In the categorical condition the false alarm and miss rates were 0.5% and 2.5%, respectively. Corresponding errors in the specific condition were 0.4% and 1.6%. These trials were excluded from all subsequent analyses.

Figure 2 shows manual reaction times (RTs) for the categorical and specific search conditions, as a function of set size and target presence. There were significant main effects of target presence, $F(1, 11) = 34.0$, $p < .001$, and set size, $F(2, 22) = 36.7$, $p < .001$, in the categorical search data, as well as a significant target × set size interaction, $F(2, 22) = 20.5$, $p < .001$. Similar patterns characterized the specific search data. There were again significant main effects of target presence, $F(1, 11) = 18.6$, $p < .01$, and set size, $F(2, 22) = 39.1$, $p < .001$, as well as a significant interaction between the two, $F(2, 22) = 16.1$, $p < .001$. However, we also found a significant three-way interaction between target presence, set size, and search condition, $F(2, 44) = 5.30$, $p < .01$. Target present and absent search slopes in the categorical condition were 33.8 ms/item and 111.4 ms/item, respectively. Search slopes for a specific target were only 14.9 ms/item in the target present condition and 48.5 ms/item in the target absent condition. With respect to manual measures, categorical search was much less efficient than the search for a specific target designated by a preview.

Although the manual data can be interpreted as evidence for search guidance only when the target's specific features are known in advance, such a conclusion would be premature. First, as already noted there is an inherent ambiguity in the relationship between manual search slopes and the decision processes involved in distractor rejection. Search slopes may be shallower in the specific condition, not because of better guidance to the target, but rather because the availability of a specific target template makes it easier to classify objects as distractors. Second, although slopes were steeper in the categorical condition compared to the specific, all that one can conclude from this difference is that target specific search is more efficient, not that categorical search is un-guided. To more directly compare guidance under specific and categorical search conditions, we therefore turn to eye movement measures.

Eye movement data

If search is guided to specific targets but not to categorical targets we should find a high percentage of immediate target fixations under specific search conditions, and a chance level of such fixations under categorical search conditions. Finding above-chance levels of immediate target fixations in either condition would constitute evidence for search guidance. Figure 3 shows the percentage of trials in which the target was the first object fixated after onset of the search display. Immediate target fixations were quite common in both search conditions, but more so when a target was specified, $F(1, 22) = 23.3$, $p < .001$. As expected, the frequency of these fixations also declined with increasing set size under both categorical, $F(2, 22) = 76.7$, $p < .001$, and specific, $F(2, 22) = 90.4$, $p < .001$, search conditions. Together, these patterns are highly consistent with the manual data and indicate stronger search guidance to targets defined by a preview. However, we also compared these immediate fixation rates to the rates expected by a random movement of gaze to one of the display objects. These correspond to the chance baselines of 16.7%, 7.7%, and 5% in the 6, 13, and 20 set size conditions, respectively. Targets in the specific search conditions were clearly fixated initially more often than what would be expected by chance, $t(11) \geq 8.93$, $p < .001$. This is unsurprising given the already strong evidence for guidance under specific conditions provided by the other search measures. More interestingly, a similar pattern of above-chance immediate target fixations was also found in the categorical search conditions. These preferential fixation rates were significantly greater than chance at each set size, $t(11) \geq 9.04$, $p < .001$, with guidance estimates ranging from 31.9% at a set size of 6 to 15.1% at a set size of 20. This preference to look initially to the target, even in the absence of knowledge about the target's specific appearance, constitutes strong evidence for categorical guidance.

Guidance might also increase during the course of a search trial, and this evidence for guidance would be missed if one focuses exclusively on immediate target fixations. To better capture this dimension of guidance we analyzed the number of distractor objects that were fixated prior to fixation on the target. The results from this analysis, shown in Figure 4, indicate that fixations on distractors were rare; fewer than two distractors were fixated even in the relatively dense twenty-object displays. Nevertheless, the average number of distractors fixated before the target increased significantly with set size in both the categorical, $F(2, 22) = 109$, $p < .001$, and specific, $F(2, 22) = 68.7$, $p < .001$, search conditions. The number of fixated distractors was also slightly smaller in the target specific condition than in the categorical condition, $F(1, 22) = 14.6$, $p < .01$, with this difference interacting with set size, $F(2, 44) = 13.6$, $p < .001$. We also compared these distractor fixation rates to baselines reflecting the number of pre-target distractor fixations that would be expected by chance. These baselines, 2.5, 6, and 9.5 in the 6, 13, and 20 set size conditions, respectively, assumed a random fixation of objects in which no fixated object was revisited by gaze (i.e., sampling without replacement). Consistent with our analysis of immediate target fixations, distractor fixation rates were well below chance levels for both categorical and target specific search conditions, with significant differences

obtained at each set size, $t(11) \leq -41.5$, $p < .001$. These below-chance rates of distractor fixation provide additional direct evidence for guidance during categorical search.

In addition to search being guided to targets, search efficiency might also be affected by how long it takes to initiate a search (Zelinsky & Sheinberg, 1997), and the time needed to verify that an object is a target once it is located (Castelhano et al., 2008). To examine search initiation time we analyzed the latencies of the initial saccades, defined as the time between search display onset and the start of the first eye movement (Table 1). Although we found small but highly significant increases in initial saccade latency with set size (Zelinsky & Sheinberg, 1997), $F(2, 22) \geq 14.1$, $p < .001$, search initiation times did not reliably differ between categorical and specific conditions, $F(1, 22) = 0.49$, $p > .05$. To examine target verification time we subtracted the time taken to first fixate a target from the manual RT, on a trial by trial basis (Table 1). This analysis revealed only marginally reliable effects of set size, $F(2, 22) \leq 3.21$, $p \geq .06$, and search condition, $F(1, 22) = 3.54$, $p = .07$. Perhaps more telling is that an analysis of just the time to fixate the target revealed large effects in both set size, $F(2, 22) \geq 57.0$, $p < .001$, and search condition, $F(2, 44) = 17.8$, $p < .001$. Although search initiation times and target verification times can affect search efficiency, in the current task search efficiency was determined mainly by guidance to the target.

To summarize, the goal of this experiment was to determine whether target guidance exists in a categorical search task, and on this point the data were clear; although search guidance was strongest when subjects knew the target's specific features, search was also guided to categorically-defined targets. To our knowledge this constitutes the first oculomotor-based evidence for categorical search guidance in the context of a controlled experiment (but see also Schmidt & Zelinsky, 2009). Also noteworthy is the fact that the observed level of categorical guidance was quite pronounced, larger than the difference in guidance between the specific and categorical search conditions. Rather than being unguided, categorical search more closely resembles the strong guidance observed under target-specific search conditions.

Experiment 2

We know from Experiment 1 that categorical search guidance exists, but how does it work? There are two broad possibilities. One is that subjects were guiding their search based on the features of a specific, previously viewed teddy bear. Although Experiment 1 prevented the repetition of objects, thereby minimizing the potential for this sort of bias from developing over the course of trials, it is nevertheless possible that subjects had in mind a specific target (perhaps a favorite teddy bear from childhood) and were using this pattern to guide their categorical search. Rather than guiding search based on the features of a specific target template, another possibility is that subjects assembled in their working memory a categorical target template based on visual features common to their teddy bear category. Search might therefore have been guided by teddy bear color, texture, and shape features, even if the subject had never viewed the exact combination of these features in a specific teddy bear.

To distinguish between these two possibilities we manipulated the number of times that a target would repeat during the course of the experiment. The net effect of this manipulation is to vary the featural uncertainty of the target category, and the capacity for subjects to anticipate or predict the correct features of the target bear. If the same target repeated on every trial, this prediction would be easy and roughly equivalent to target specific search (e.g., Wolfe et al., 2004); if each of ten targets repeated on a small percentage of trials, a correct prediction becomes more difficult. According to a specific-template model, categorical guidance should decrease monotonically with the number of potential targets, and their accompanying lower repetition rates, due to the lower probability of selecting the correct target to be used as a specific guiding template. A categorical-template model makes a very different prediction. If

a subject builds a target template from all of the teddy bears existing in their long-term memory, then the repetition of specific target bears would be expected to affect categorical guidance only minimally, or not at all.

Method

Participants—Forty Stony Brook University students participated in the experiment for course credit. All had normal or corrected to normal visual acuity, and none had participated in the previous experiment.

Stimuli and Apparatus—The stimuli and the apparatus were identical to those used in Experiment 1, except for the substitution of teddy bear targets to create the repetition conditions (described below). Otherwise, the same distractors appeared in the same search display locations.

Design—The experiment used a 2 target presence (present/ absent) × 3 set size (6/ 13/ 20) × 5 target repetition (1 target/ 2 target/ 3 target/ 5 target/ 10 target) mixed design, with target repetition being a between-subjects factor. In the 1-target (1T) condition, subjects searched for the same teddy bear target on every trial. This target therefore repeated 90 times, appearing once on each of the 90 target-present trials. In the 2-target (2T) condition, two teddy bears were used as targets, with the actual target appearing in each target-present display divided evenly between the two. Each target therefore repeated 45 times throughout the experiment. In the 3-target (3T) condition, 3 teddy bears were used as targets, resulting in each target repeating 30 times. In the 5-target (5T) condition, 5 teddy bears were used as targets, resulting in each target repeating 18 times, and in the 10-target (10T) condition, 10 teddy bears were used as targets, resulting in each target repeating 9 times. Target bears for each condition were randomly selected (without replacement) from the 90 bears used in the categorical condition from Experiment 1, and different target bears were selected and assigned to conditions for each subject. There were 180 trials per subject, and 30 trials per cell of the design.

Procedure—The procedure was identical to what was described for the categorical search task from Experiment 1; no previews of specific targets were shown prior to the search displays.

Results and Discussion

Manual data—Error rates were uniformly low across all target repetition conditions, with miss rates less than 2.6% and false alarm rates less than 0.8%. These error trials were excluded from all subsequent analyses.

Figure 5 shows the mean RT data for the 5 repetition conditions, along with data re-plotted from the Experiment 1 categorical search task, which we now refer to as the 90T condition (reflecting the absence of target repetition). The search slopes for each condition are provided in Table 2. We found significant main effects of target presence, $F(1, 45) = 144$, $p < .001$, and set size, $F(2, 90) = 162$, $p < .001$, as well as a significant target × set size interaction, $F(2, 90) = 83.3$, $p < .001$. Target presence and set size also interacted with the between-subjects repetition factor, $F(10, 90) = 2.60$, $p < .01$; the set size difference between present and absent search increased with the repetition rate. Tukey HSD tests further revealed that the 1T and 3T conditions differed from the 90T condition ($p \leq .05$); no other post-hoc comparisons were reliable (all $p > .2$). With respect to our motivating hypotheses, the manual RT data are therefore mixed. Search efficiency did improve with target repetition, but this repetition effect was limited to comparisons involving the extremes of our manipulation, cases in which the target repeated on every trial (1T) or never repeated (90T). In light of this, perhaps the more compelling conclusion from these analyses is that no reliable differences were found between

any conditions in which a target was repeated, suggesting a relatively minor role of target repetition rate on search.

Eye movement data—To clarify the level of guidance to categorical targets we again turned to eye movement measures. Figure 6 shows the percentage of trials in which the first fixated object was the target, grouped by repetition condition and set size. Immediate target fixation rates for a given set size and repetition condition were all well above the chance baselines defined for each set size, $t(7) \geq 6.37$, $p < .001$. This was expected, as repeating targets should not reduce guidance relative to the categorical levels reported in Experiment 1. We also found significant main effects of set size, $F(2, 90) = 203$, $p < .001$, and repetition, $F(5, 45) = 7.88$, $p < .001$, as well as a repetition $\times$ set size interaction that approached significance, $F(10, 90) = 1.83$, $p = .067$. More interestingly, through post-hoc comparisons we determined that the percentage of immediate target fixations in the 1T condition was significantly greater than any other repetition condition, and the no-repetition condition from Experiment 1 (all $p < .05$). None of the other conditions reliably differed. Notably, the rate of immediate target fixations in the 90T condition was similar to the rates observed in the 2T, 3T, 5T, and 10T repetition conditions (all $p > .2$). Immediate search guidance to a target is therefore strongest when the target pattern is entirely predictable (1T), but any ambiguity in the target description results in a drop to a relatively constant level of above-chance guidance that does not vary with repetition rate.

Figure 7 shows the number of distractors that were fixated before the target, again grouped by repetition condition and set size. As in Experiment 1 distractors were rarely fixated in this task, and their fixation rates in all conditions were well below chance levels, $t(7) \leq -29.8$, $p < .001$. Yet despite their infrequent occurrence we found significant main effects of distractor fixation in both set size, $F(2, 90) = 357$, $p < .001$, and target repetition, $F(5, 45) = 5.41$, $p < .01$, as well as a significant repetition $\times$ set size interaction, $F(10, 90) = 4.66$, $p < .001$. Post-hoc tests again confirmed that these differences were limited to comparisons involving the 1T repetition condition; fewer distracters were fixated in the 1T condition compared to all of the rest ($p < .01$ for contrasts with 2T and 10T; $p < .09$ for contrasts with 3T and 5T). No reliable differences were found between any other pairings of conditions (all $p > .4$). These patterns parallel almost exactly those found for immediate target fixations; search is guided best to highly predictable targets, with decreasing predictability resulting in above-chance guidance that does not change with target repetition rate.

As in Experiment 1 we again analyzed search initiation times and target verification times, and obtained very similar results. Search initiation times, as measured by initial saccade latency, increased with set size for each of the target repetition conditions, $F(2, 14) \geq 5.56$, $p < .05$, but no reliable differences were found between conditions, $F(4, 34) = 0.93$, $p > .05$. Similarly, target verification times were not significantly affected by target repetition, $F(4, 34) = 2.19$, $p > .05$. Repetition effects were found only in the time taken to first fixate targets, $F(4, 34) = 9.98$, $p < .001$, and these differences were limited to contrasts between the 1T condition and all others (all other $p > .05$). Once again, although repetition effects might have been expressed in terms of search initiation or target verification times, these effects, to the extent that they existed at all, appeared only in a measure of search guidance.

To summarize, Experiment 2 tested whether target guidance in a categorical search task is due to the use of a specific or categorical target template. We did this by manipulating target repetition, which varied the predictability of the target. If categorical guidance relies on a specific target template our expectation was that guidance should improve with target predictability, as the template selected for guidance would be more likely to match the search target. This expectation was not confirmed. Rather than finding a graded pattern of guidance that steadily increases with higher target repetition rates, both manual and oculomotor measures

were consistent in suggesting that target repetition effects in our task were small and limited to the extremes of our manipulation. Indeed, a consistent repetition effect was found only between the 1T condition and all of the rest. We can therefore rule out the possibility that categorical guidance is mediated by a template corresponding to a specific, previously viewed instance of the target class, at least under conditions of target uncertainty.[1]

General Discussion

Quite a lot is now known about how attention, and gaze, is guided to targets in a search task. From the seminal work of Wolfe and colleagues (Wolfe, 1994; Wolfe, Cave, & Franzel, 1989) it is known that search is not random, but rather is guided by knowledge of the target's specific features or appearance. This appearance-based search guidance has even been recently implemented in the form of a computational model and applied to realistic objects and scenes (Zelinsky, 2008). Recent work also tells us that this target-specific form of guidance is better than guidance to categorically-defined targets, as demonstrated by increased search efficiency with a specific target preview compared to a target designated by a text label (Castelhano et al., 2008; Foulsham & Underwood, 2007; Schmidt & Zelinsky, 2009; Vickery et al., 2005; Wolfe et al., 2004). However, one question that remained open was whether the search for these categorical targets was also guided, albeit less efficiently than to specific targets. The present study answers this question definitively; search is indeed guided to categorically-defined targets.

Answering this question required adopting in Experiment 1 a dependent measure that could dissociate search guidance from decision-related factors, and a design that included a meaningful no-guidance baseline against which search efficiency could be assessed. Replicating previous work, we found that target-specific search was more efficient than categorical search, indicating that subjects can and do use information about the target's specific appearance to guide their search when this information is available. However, we also found that the proportion of immediate fixations on categorical targets was much greater than chance, and far fewer distractors were fixated during categorical search than what would be expected by a random selection of objects. Importantly, these preferences could not be due to target specific guidance, as targets did not repeat over trials. Together, these patterns provide strong evidence for the existence of search guidance to categorical targets.

In Experiment 2 we showed that this categorical guidance does not use a specific target template in memory even when such templates are readily available. It might have been the case that subjects, when asked to search categorically for *any* teddy bear, would load into working memory a specific, previously viewed teddy bear target and use this pattern's features to guide their search. We tested this hypothesis by manipulating target repetition throughout the experiment, under the assumption that a frequently repeated target would be more likely to be selected as a guiding template. However, except under conditions of perfect target predictability, we found little to no effect of target repetition on search guidance. It appears that specific target templates are not typically used to guide search to categorical targets.

If categorical search does not rely on specific target templates, where else might these guiding features come from? We believe that the reported patterns of guidance are best described by a process that matches a search scene to a categorical model, one created by extracting discriminative features from a pool of previously viewed instances of the target class (Zhang,

---

[1]We also conducted a simulation to test an alternative version of the specific template hypothesis, one in which search is guided based on the specific target that was viewed on the preceding search trial. This simulation produced an immediate target fixation rate that increased with target repetition far more than what we observed in the behavioral data. The fact that guidance behavior in our task did not meaningfully change over the 2T-10T repetition range would seem to rule out any model of search that assumes the use of a specific guiding template.

Yang, Samaras, & Zelinsky, 2006). According to this account the creation of a target model would involve selecting from this pool those features that maximize the discriminability between the target category and expected non-target patterns, given a specific context (random objects, a particular type of scene, etc.). When a target is completely predictable a categorical template degenerates into a form of specific template, as only a single object exists from which to extract guiding features. However, for most categorical search tasks this pool of target instances would be very large, encompassing not only those targets viewed in previous experimental trials but also potentially all of the target instances existing in long term memory. This diversity of instances means that the features ultimately included in the target model will probably not match perfectly the features from the specific target appearing in a given search scene, resulting in the weaker guidance typically found to categorical targets. However, because these features represent the entire target class some match to the target is expected, and it is this match that produces the above chance levels of categorical guidance reported in this study.

A categorical target model also explains why guidance did not vary appreciably with target predictability in Experiment 2. We now know that the injection of any uncertainty into the target description (i.e., all conditions other than 1T) results in subjects choosing, explicitly or otherwise, not to search for a specific teddy bear as viewed on a previous trial. Under these conditions of uncertainty we believe that subjects instead adopt a categorical target model; they choose to search for *any teddy bear* rather than a particular one. With the adoption of a categorical model the repetition of targets over trials diminishes in importance, as even high repetition rates might not add enough new instances to meaningfully affect the model's feature composition. However, the non-significant trend toward less efficient guidance in the 90T condition (relative to the repetition conditions) suggests that all instances may not be treated equally, and that some attempt may be made to weight the features of the repeating targets in the categorical model. Note that a preferential weighting of recently viewed instances in the feature selection process would constitute a form of implicit perceptual priming, and might explain in part previous reports of target repetition on search (see Kristjánsson, 2006, for a review).

In summary, the creation and use of a categorical target model can account for multiple facets of categorical search behavior, including: (1) the fact that search is guided most efficiently when there is a specific target preview, (2) the fact that guidance to categorically-defined targets is generally weaker, but still above chance, and (3) the fact that categorical guidance is largely immune to target repetition effects.

Our evidence for categorical search guidance has important implications for search theory. Models of search often assume knowledge of the target's exact features. With this knowledge, a working memory template can be constructed and compared to the scene, with the degree of search guidance being proportional to the success of this match (e.g., Wolfe, 1994; Zelinsky, 2008). How this process works in the case of real-world categorical search is an entirely open question. Previous solutions to the problem of feature selection and target representation simply no longer work. Although categorical features can be hand-picked when stimuli are simple and uni-dimensional (e.g., "steep" or "left leaning", Wolfe et al., 1992), this approach is obviously undesirable for more complex objects and stimulus classes, which likely vary on a large and unspecified number of feature dimensions. Likewise, methods of representing realistic objects often involve the application of spatio-chromatic filters to an image of a target (e.g., Rao, Zelinsky, Hayhoe, & Ballard, 2002; Zelinsky, 2008), an approach that obviously will not work for categorically-defined targets for which there is no specific target image. To date, there is no computationally explicit theory for how people guide their search to categorical targets.

Our suggestion that this guidance is mediated by a categorical target model creates a framework for how this problem might be solved. Critical to this framework is a method of learning those visual features that are specific to an object class, and those that are shared between classes. In preliminary work we used a machine learning technique (Zhang, Yu, Zelinsky, & Samaras, 2005; see also Viola & Jones, 2001) to select features allowing for the discrimination of teddy bears from random objects in a set of training images (Zhang et al., 2006). We then created from these features a teddy bear classifier that we applied to a categorical search task, one in which new teddy bear targets were presented with new random objects. The output of this classifier provided a guidance signal that, when combined with operations described in Zelinsky (2008), drove simulated gaze to teddy bear targets in the search arrays. Although the behavior of this model captured human gaze behavior in the same categorical search task quite well (Zhang et al., 2006), this effort should be viewed as only a small first step toward understanding the processes involved in categorical search. In truth, research on this problem is only in its infancy, with the space of features and learning methods potentially contributing to categorical search remaining essentially unexplored. All that is currently clear is that the ultimate solution to this problem will require a fundamentally interdisciplinary perspective, one that blends behavioral techniques for studying search with computer vision techniques for detecting classes of objects.

Acknowledgments

References

Brand J. Classification without identification in visual search. Quarterly Journal of Experimental Psychology 1971;23:178–186. [PubMed: 5559706]

Bravo MJ, Farid H. Search for a category target in clutter. Perception 2004;33(6):643–652. [PubMed: 15330362]

Bravo MJ, Farid H. The specificity of the search template. Journal of Vision 2009;9:1–9. [PubMed: 19271904]

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

Chen X, Zelinsky GJ. Real-world visual search is dominated by top-down guidance. Vision Research 2006;46(24):4118–4133. [PubMed: 17005231]

Cockrill, P. The teddy bear encyclopedia. New York: DK Publishing, Inc.; 2001.

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

Egeth H, Jonides J, Wall S. Parallel processing of multielement displays. Cognitive Psychology 1972;3:674–698.

Ehinger K, Hidalgo-Sotelo B, Torralba A, Oliva A. Modeling search for people in 900 scenes: A combined source model of eye guidance. Visual Cognition. in press

Fletcher-Watson S, Findlay JM, Leekam SR, Benson V. Rapid detection of person information in a naturalistic scene. Perception 2008;37(4):571–583. [PubMed: 18546664]

Foulsham T, Underwood G. How does the purpose of inspection influence the potency of visual salience in scene perception? Perception 2007;36(8):1123–1138. [PubMed: 17972478]

Hayhoe MM, Shrivastava A, Mruczek R, Pelz JB. Visual memory and motor planning in a natural task. Journal of Vision 2003;3:49–63. [PubMed: 12678625]

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

Jonides J, Gleitman H. A conceptual category effect in visual search. O as letter of as digit. Perception and Psychophysics 1972;12(6):457–460.

Kirchner H, Thorpe SJ. Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. Vision Research 2006;46(11):1762–1776. [PubMed: 16289663]

Kristjánsson Á. Rapid learning in attention shifts— A review. Visual Cognition 2006;13:324–362.

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

Mruczek REB, Sheinberg DL. Distractor familiarity leads to more efficient visual search for complex stimuli. Perception and Psychophysics 2005;67:1016–1031. [PubMed: 16396010]

Newell FN, Brown V, Findlay JM. Is object search mediated by object-based or image-based representations? Spatial Vision 2004;17:511–541. [PubMed: 15559117]

Rao R, Zelinsky GJ, Hayhoe MM, Ballard D. Eye movements in iconic visual search. Vision Research 2002;42:1447–1463. [PubMed: 12044751]

Schmidt J, Zelinsky GJ. Search guidance is proportional to the categorical specificity of a target cue. Quarterly Journal of Experimental Psychology. 2009

Torralba A, Oliva A, Castelhano M, Henderson JM. Contextual guidance of attention in natural scenes: The role of global features on object search. Psychological Review 2006;113:766–786. [PubMed: 17014302]

Vickery TJ, King LW, Jiang Y. Setting up the target template in visual search. Journal of Vision 2005;5:81–92. [PubMed: 15831069]

Viola, P.; Jones, MJ. Robust object detection using a boosted cascade of simple features. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2001. p. 511-518.

Wolfe JM. Guided search 2.0: A revised model of visual search. Psychonomic Bulletin & Review 1994;1:202–238.

Wolfe JM, Cave KR, Franzel SL. Guided search: An alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human Perception & Performance 1989;15:419–433. [PubMed: 2527952]

Wolfe JM, Friedman-Hill SR, Stewart MI, O'Connell KM. The role of categorization in visual search for orientation. Journal of Experimental Psychology: Human Perception and Performance 1992;18(1): 34–49. [PubMed: 1532193]

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

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

Zelinsky GJ, Sheinberg DL. Eye movements during parallel-serial visual search. Journal of Experimental Psychology: Human Perception and Performance 1997;23(1):244–262. [PubMed: 9090154]

Zhang, W.; Yang, H.; Samaras, D.; Zelinsky, GJ. A computational model of eye movements during object class detection. In: Weiss, Y.; Scholkopf, B.; Platt, J., editors. Advances in Neural Information Processing Systems. Vol. 18. Cambridge, MA: MIT Press; 2006. p. 1609-1616.

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

A

B

Figure 1.
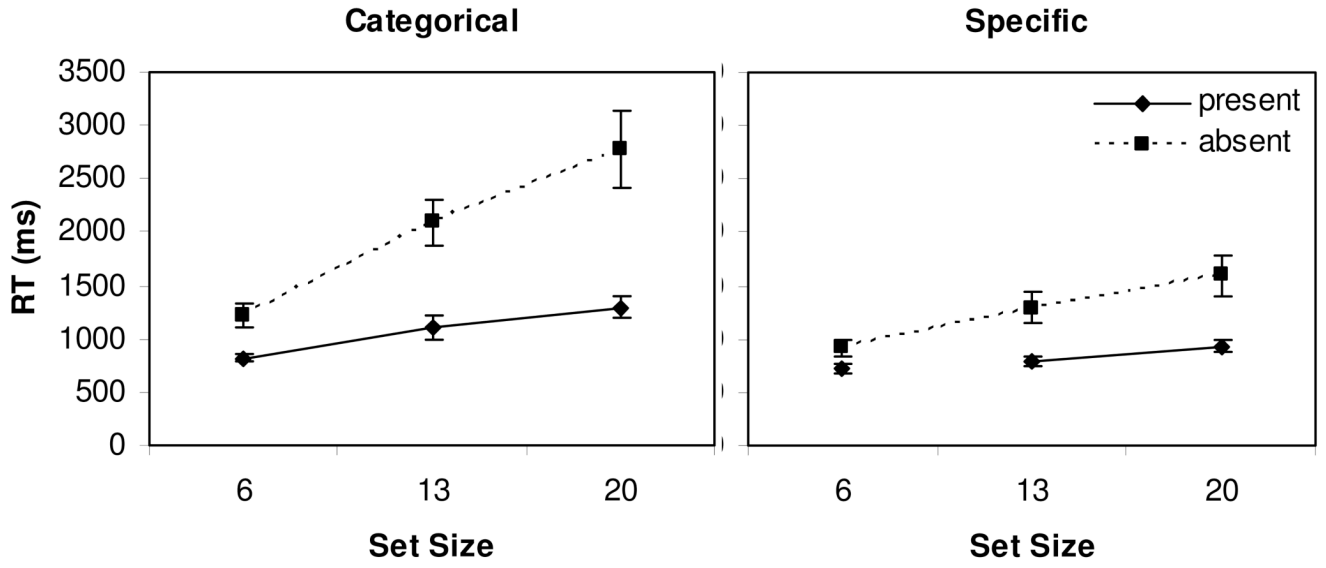Representative targets (A) and distractors (B) used as stimuli.

Figure 2.
Mean manual reaction times from correct trials in Experiment 1. Error bars indicate standard error.
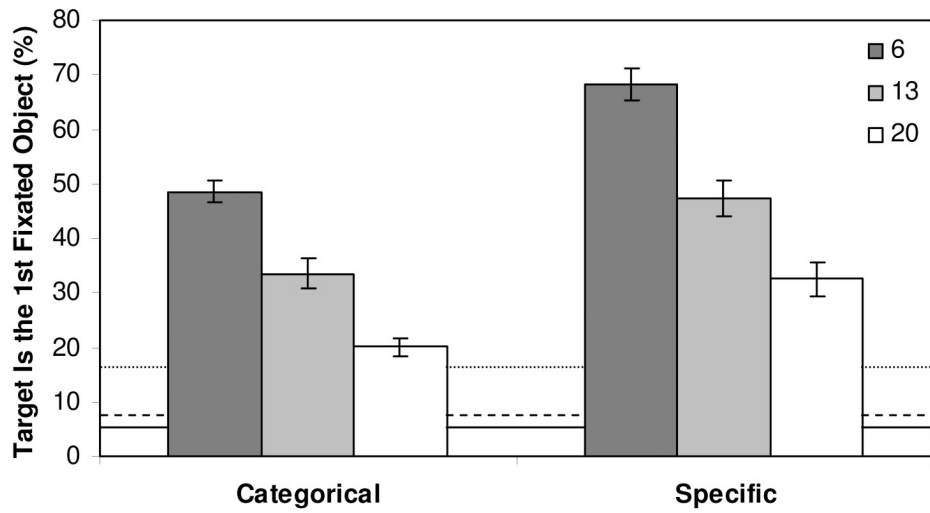
Figure 3.
Percentage of Experiment 1 target present trials in which the first fixated object was the target. Dotted, dashed, and solid lines indicate chance levels of guidance in the 6, 13, and 20 set size conditions, respectively. Error bars indicate standard error.
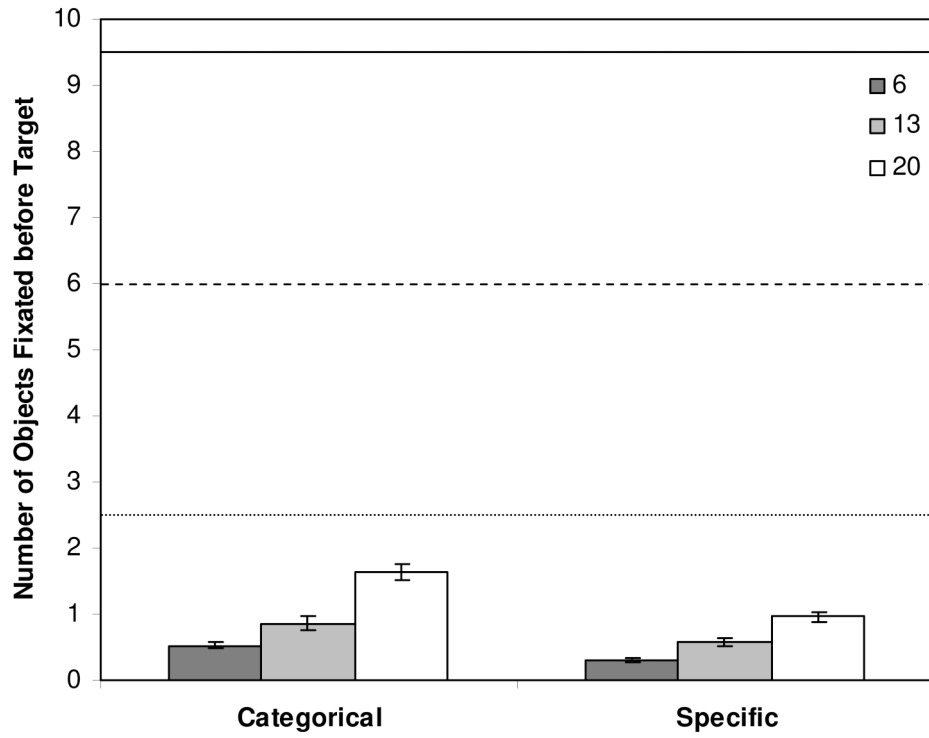
Figure 4.
Average number of distractors fixated before the target (without replacement) from target present Experiment 1 trials. Dotted, dashed, and solid lines indicate chance levels of guidance in the 6, 13, and 20 set size conditions, respectively. Error bars indicate standard error.
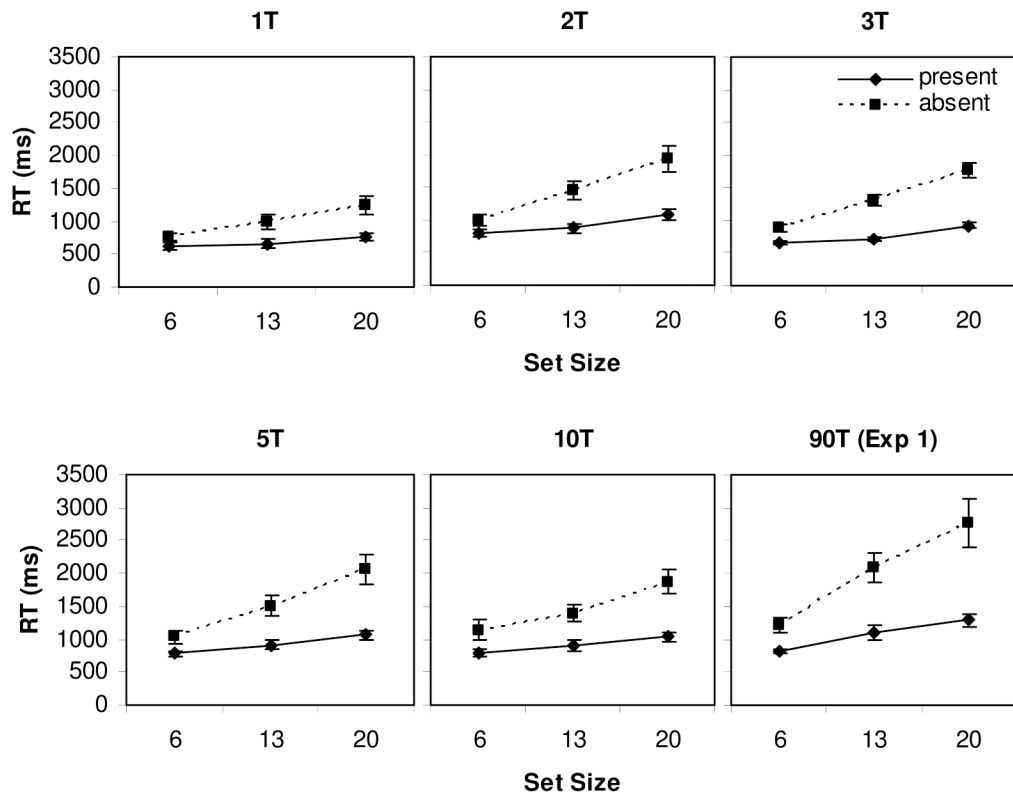
Figure 5.
Mean manual reaction times for the 1 target, 2 target, 3 target, 5 target, and 10 target repetition conditions from Experiment 2. Data from the 90 target condition are re-plotted from Experiment 1. Error bars indicate standard error.
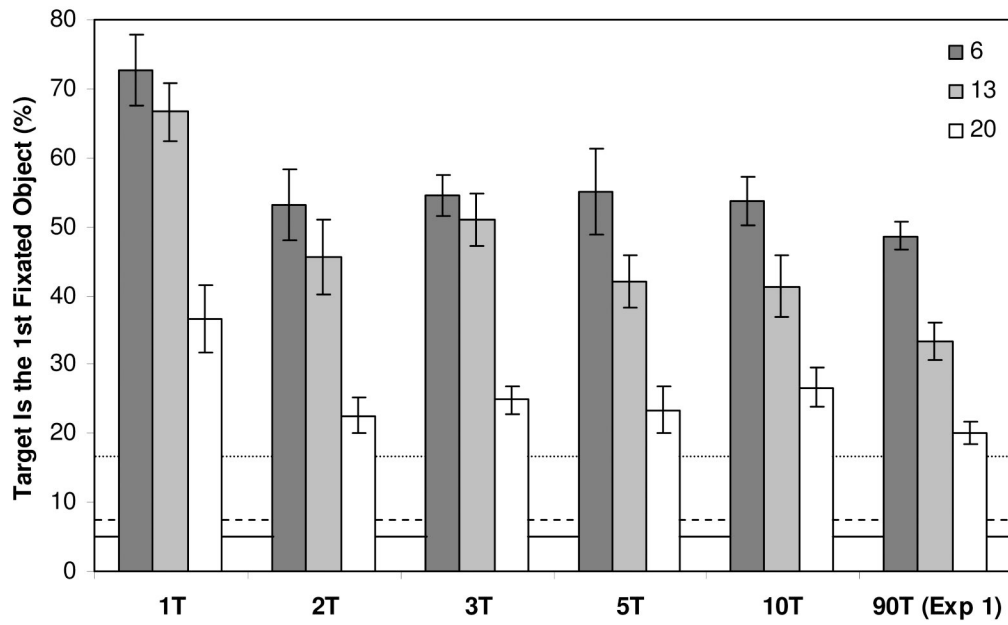
Figure 6.
Percentage of Experiment 2 target present trials in which the first fixated object was the target, grouped by repetition condition. Dotted, dashed, and solid lines indicate chance levels of guidance in the 6, 13, and 20 set size conditions, respectively. Error bars indicate standard error.

Figure 7.
Average number of distractors fixated before the target (without replacement) from target present Experiment 2 trials. Dotted, dashed, and solid lines indicate chance levels of guidance in the 6, 13, and 20 set size conditions, respectively. Error bars indicate standard error.
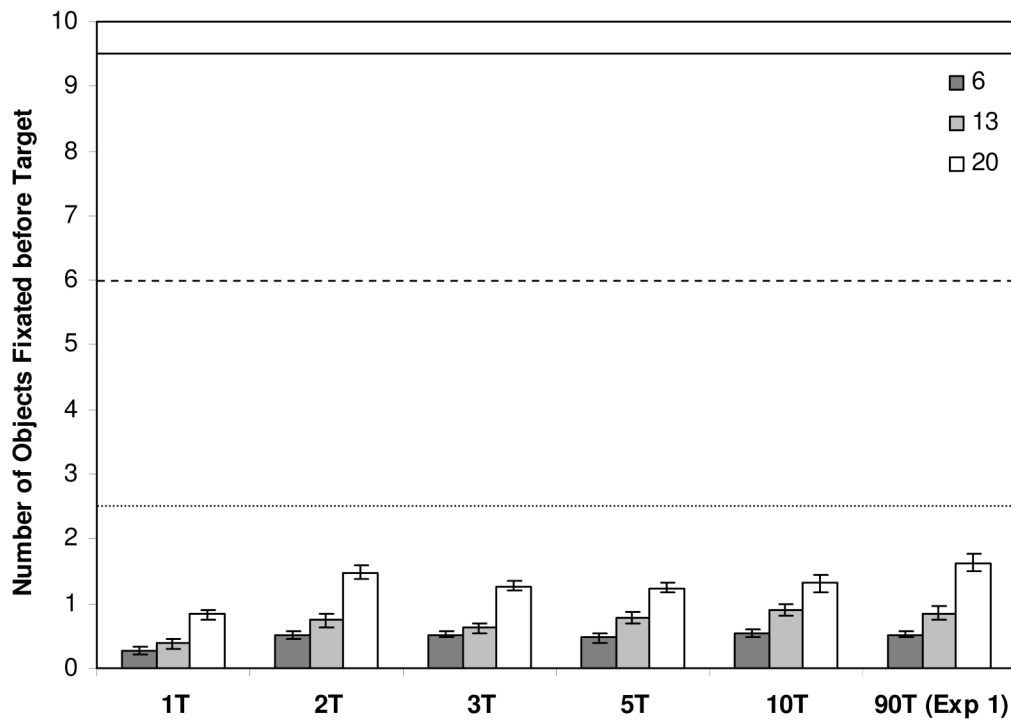
Table 1

Initial saccade latencies, RTs to target, and target verification times from Experiment 1

Set Size	Categorical			Specific		
	6	13	20	6	13	20
Initial Saccade Latency						
Target Present	181 (6.3)	185 (6.9)	191 (8.4)	182 (6.9)	189 (5.8)	196 (7.7)
Target Absent	174 (6.8)	186 (8.1)	193 (8.6)	184 (6.9)	203 (8.0)	198 (5.9)
RT to Target	362 (12.4)	450 (20.3)	667 (21.8)	331 (8.5)	391 (12.1)	495 (20.7)
Target Verification Time	446 (24.4)	622 (107.5)	564 (86.9)	376 (34.4)	396 (37.4)	421 (36.5)

Note: All values in msec. Values in parentheses indicate standard error.

Table 2

Categorical search slopes for manual RT data from Experiments 1 and 2

Repetition	Target Present	Target Absent
1T	9.7 (1.3)	33.3 (6.9)
2T	21.0 (2.1)	67.2 (9.2)
3T	18.0 (1.1)	63.7 (5.6)
5T	20.5 (3.1)	73.0 (13.0)
10T	17.1 (2.2)	52.7 (7.2)
90T (Exp 1)	33.8 (5.6)	111.4 (19.0)

Note: All values in msec/item. Values in parentheses indicate standard error.