

Setting up the target template in visual search

Timothy J. Vickery

Department of Psychology, Harvard University,
Cambridge, MA, USA



Li-Wei King

Department of Psychology, Harvard University,
Cambridge, MA, USA



Yuhong Jiang

Department of Psychology, Harvard University,
Cambridge, MA, USA



Top-down knowledge about the target is essential in visual search. It biases visual attention to information that matches the target-defining criteria. Extensive research in the past has examined visual search when the target is defined by fixed criteria throughout the experiment, with few studies investigating how subjects set up the target. To address this issue, we conducted five experiments using random polygons and real-world objects, allowing the target criteria to change from trial to trial. On each trial, subjects first see a cue informing them about the target, followed 200-1000 ms later by the search array. We find that when the cue matches the target exactly, search speed increases and the slope of response time–set size function decreases. Deviations from the exact match in size or orientation slow down search speed, although they lead to faster speed compared with a neutral cue or a semantic cue. We conclude that the template set-up process uses detailed visual information, rather than schematic or semantic information, to find the target.

Keywords: visual search, target switch, top-down control, visual attention

Introduction

Visual search is a routine human behavior. Finding your friend in a crowd, grabbing a drink from the fridge, and hunting for your lost keys, are some of the routine tasks that exemplify visual search. In this process, we hold in mind a pre-specified target, such as our friend or keys, and move attention in the visual field until a match is spotted. In the past two decades, visual search has been one of the most popular research topics in vision research. Thanks to Anne Treisman, Jeremy Wolfe, John Duncan, Robert Desimone, and others, psychologists and neuroscientists now know a lot about human search behavior. For example, some search tasks are easy: Spotting a red flower among green leaves only takes about 300 ms, even when there are many green leaves in the field (Treisman & Gelade, 1980). Other search tasks are more difficult: Finding a “T” among rotated “Ls” is a slow and deliberate process and takes longer with more “Ls.” Such research has led to excellent models of human attention in search tasks, such as the Feature Integration Theory (Treisman & Gelade, 1980; Treisman & Sato, 1990), Guided Search (Wolfe, 1994), and Biased Competition Model (Desimone & Duncan, 1995). Their differences aside, these models all propose that visual search is an interactive process between top-down knowledge and bottom-up information (i.e., goal-driven and stimulus-driven cues). Attention is biased, or guided, by top-down knowledge about the target. Stimuli that match the target criteria are weighted more heavily.

They dominate neuronal activity, resulting in successful search.

Surprisingly, although top-down knowledge about the target is crucial in visual search, most visual search studies have largely minimized top-down control by asking subjects to search for the same target for hundreds of trials in a row (e.g., Chun & Jiang, 1998; Duncan & Humphreys, 1989). Similarly, in neuroscience, many studies have been devoted to specifying how activity in earlier visual areas is biased by top-down attention (e.g., Reynolds, Pasternak, & Desimone, 2000), but few have looked at the biasing signals themselves (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999). How do we set up the target template for visual search? How long does this process take? What do we have to know about the target to find it efficiently? To answer these questions, one must increase the proportion of trials in which the “biasing signal” is set up. That is, the target of interest must change frequently in the experiment so a new target template needs to be set up on every trial.

Despite their theoretical importance, empirical studies on target set-up processes are only beginning to emerge. Our study aims at facilitating this growing field by addressing the following questions: In difficult visual search tasks, do subjects rely on semantic or visual information to find the target? If visual information is used, what do we have to know about the target to set up an effective template? Can we discard incidental properties such as the target’s size and orientation? Before jumping into the answers to these questions, however, we shall first briefly review relevant studies in the literature.

Changing targets in search

Constantly changing targets from trial to trial slows down reaction time (RT) (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). This basic observation holds even when the target always differs from distractors in a single feature (Treisman, 1988). For example, when subjects search for a uniquely colored target among distractors of another color, their speed in detecting the target is fast when all trials show a red target among green distractors (or vice versa), but slow when a proportion of the trials show a red target among green distractors and the remaining trials show the reverse (Maljkovic & Nakayama, 1994; Wolfe, Butcher, Lee, & Hyle, 2003). The mixed-blocks are slow primarily because the distractor value on some trials – such as red – can become the target value on other trials. If the distractors are always in blue, and the target is sometimes red and sometimes green, the cost of switching targets is negligible (Found & Muller, 1996; Muller, Heller, & Ziegler, 1995).

Observing the cost of target switch in feature search, Wolfe et al. (2003) conclude that top-down attentional control is entailed even in feature search tasks, which are traditionally considered as requiring little attention (Treisman & Gelade, 1980). The cost of switching in these tasks results from both active attentional control and passive priming. Active control allows one to set up the exact target-template for search. Passive priming, in the form of positive priming from repeated targets and negative priming from alteration, also modulates RT (Kristjansson, Wang, & Nakayama, 2002).

Target switching costs also apply to difficult search tasks where the target is defined by a conjunction of two distractor features. In these tasks, advanced knowledge about the target facilitates search. Wolfe, Horowitz, Kenner, Hyle, and Vasan (2004) show that such knowledge needs to convey visual pictorial representation of the target. Whereas a picture of a duck facilitates search when it is shown only 100 ms ahead of the search display, the word “duck” is never as effective, even when it is presented 500 ms ahead of time. The word cue helps visual search, but it is not as helpful as an exact cue. These results suggest that the visual system prefers to use visual, rather than semantic, knowledge to set up the target template.

Between an exact visual cue and an abstract semantic cue lies a wide range of cues differing from the exact target object in various properties. Suppose I ask you to find an apple from oranges, and suppose I show you an image of the apple that differs from the actual target in size or orientation, will you be able to find the apple just as fast as if I've shown you the exact image? More generally, what information about the target is used to set up its template? What differences between the cue and the target can be tolerated?

We conducted five experiments on visual search to study the template set-up process. These experiments share a similar design structure: On each trial subjects first view a cue object that informs them about the target on that trial.

Then approximately 200 to 1000 ms later the search display is presented. The search display contains 5 to 15 items, one of which is the target. We measure how search speed changes as a function of the cue type. The cue may be identical to the target (“exact cue”), smaller than the target (“small cue”), the same size as the target but rotated by various angles (“rotated cue”), a semantic label (“word cue”), or an uninformative shape (“uninformative cue”). Random polygons are used in Experiments 1-3, while three-dimensional (3D) models of real-world objects are used in Experiments 4-5.

Methods

Participants

Sixty-two subjects from Harvard University participated in the experiments for payment or course credit: 7 in Experiment 1, 12 in Experiment 2, 12 in Experiment 3, 15 in Experiment 4, and 16 in Experiment 5. They were 18 to 35 years old; all had normal or corrected-to-normal visual acuity and passed the color blindness test. Most subjects were tested only in a single experiment, although some (about 2-3) participated in more than one.

Stimuli

Experiments 1-3 tested visual search for 2D random polygons. These stimuli were selected because they were visually complex but novel, and they could not be verbally labeled. These properties ensured that subjects would have virtually no experience searching for such objects prior to the experiment. Each polygon subtended approximately $2.5^\circ \times 2.5^\circ$.

Experiments 4-5 used 3D models of real-world objects. These items were selected from the Object Databank, including a set of pictures of 3D models of various objects viewed from several angles. The images were created by Scott Yu and are provided by Michael Tarr (<http://www.cog.brown.edu/~tarr/projects/databank.html>). Items were selected from this set and converted into gray-scale images.

Trial sequence

Each trial started with a fixation point for approximately 400 ms. Then a cue was presented for 200-500 ms. After a blank interval of 0-1000 ms, the search display was presented until subjects made a response. In Experiments 1 and 2, subjects were asked to search for a vertically symmetric object among tilted objects, so the target was not defined by the cue. In this case, the cue provided incidental information about the target. In Experiments 3-5, the target could be any object that matched the cue. In all experiments, the target was always present on every trial. Subjects were told to press the spacebar as soon as they found the target. This response cleared the screen and brought up an

array of letters (“A” or “B”). Subjects then typed in the letter at the position of the target, providing a measure of accuracy. We didn’t test target absent trials because the decision about when to abandon search complicated RT interpretation (Chun & Wolfe, 1996).

We varied three factors: cue type, cue leading time, and search set size. *Cue type* could be *exact cue* (the cue was identical to the target), *small cue* (the cue was half the size of the target), *rotated cue* (the cue had different orientations than the target), *word cue* (the cue was a word describing the target), and *uninformative cue* (the cue carried no information about the target). *Cue leading time* was the interval between the onset of the cue and the onset of the search display. It ranged from 200 ms to 1000 ms. *Search set size* referred to the number of items on the display and ranged from 5 to 15.

Experiment 1: Incidental cues: Exact vs. uninformative

Task

In this experiment, subjects were told to search for the unique, vertically symmetric polygon, among tilted polygons. There were a total of 32 possible targets and 512 distractors. Because search could be completed without relying on the cue, the cue was incidental to task performance. This did not preclude subjects from actively using the cue to find the target, but the use of the cue was not required.

Design

On each trial a cue was presented for 200 ms, followed by a blank interval of 0 or 300 ms, and then the search display. The cue leading time was thus 200 ms or 500 ms. The cue was either identical to the target object (“exact cue”) or a square (“uninformative cue”). There were 5, 10, or 15 items on each trial. All factors – cue leading time, cue type, and set size – were randomly changed from trial to trial. Subjects completed 12 practice trials and 384 experimental trials. We were interested in the RT difference between exact cue and uninformative cue conditions.

Experiment 2: Incidental cue: Size and orientation change

Task

Just like [Experiment 1](#), subjects were told to search for a vertically symmetric object among tilted objects such that the cue was incidental to the task.

Design

The cue leading time was 200 ms on all trials. There were 8 types of cues: exact cue, uninformative cue, small cue (the cue was half the size of the target), and five types of rotated cues (the cue was rotated by 30°, 60°, 90°, 120°, or 150° from the vertical). There were 5, 10, or 15 items on

the display. All conditions were randomly intermixed in the experiment. Of interest is how RT would be affected by the size and orientation differences between the target and the cue.

Subjects completed 12 practice trials and 648 experimental trials.

Experiment 3: Deliberate cue: Size and orientation change

Task

Unlike [Experiments 1](#) and [2](#) where the target was defined independently of the cue, in [Experiment 3](#) the target was defined by the cue. The cue itself was always oriented at 0°, while the target and distractors could be presented at any of the orientations. Subjects searched for the cued shape. They were informed that the target might be viewed from a different angle than the cue and that they should ignore orientation or size changes.

Design

The cue leading time was always 200 ms and the set size was 5, 10, or 15. Seven cue types were tested: exact cue, small cue, and five rotated cues (30°, 60°, 90°, 120°, and 150°). We did not test uninformative cue because the cue must convey target information for the task to be completed.

All conditions were randomly intermixed in presentation. Subjects completed 12 practice trials and 504 experimental trials.

[Figure 1](#) is a schematic illustration of the presentation sequence used in [Experiments 2](#) and [3](#).

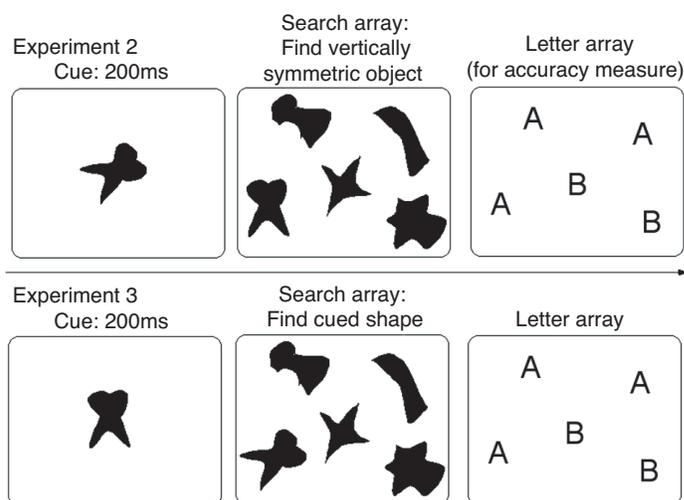


Figure 1. Presentation sequence used in [Experiments 2](#) and [3](#). In [Experiment 2](#), subjects searched for a vertically symmetric object. They pressed the spacebar when the target was spotted, and then typed in the letter at the target’s location. In [Experiment 3](#), subjects searched for a shape defined by an upright cue. Above arrays not drawn to actual scale.

Experiment 4: Deliberate cues: 3D shapes

Task

To ensure that results from novel, 2D random polygons generalize to familiar, 3D objects, we tested subjects using grayscale 3D models of frequently encountered objects. Subjects searched for an object depicted by the cue.

Design

The cue leading time was 200, 400, or 1000 ms. The cue could be one of three types: exact cue, word cue, or 90° rotated cue. Unlike Experiments 2-3 where rotation occurred only in the 2D plane, in Experiment 4 rotation could occur in the depth plane or in the 2D plane. For example, on trials when subjects searched for a computer, the cue might be identical to the target (exact cue), the word “COMPUTER” (word cue), or a 90° side view of the computer (rotated cue). There were 8 or 16 items on each search display. All conditions were randomly intermixed in the experiment. Subjects completed 12 practice trials and 360 experimental trials.

Prior to the experiment, subjects were first familiarized with the verbal label and visual images. They read aloud a word (e.g., “COMPUTER”), then saw two views of the named object that differed by 90°.

Figure 2 shows a schematic trial sequence for Experiment 4.

Experiment 5: Deliberate cue: 3D shapes with various orientations

Task

Just like Experiment 4, subjects searched for a 3D model of a real-world object depicted by a cue. Once they found the target, they pressed the spacebar and then typed in the letter behind the target object.

Design

The cue leading time was 1000 ms on all trials. We did not test the word cue (which was already tested in Experiment 4), but used 6 types of cues differed from the target object by 0° (exact cue), 30°, 60°, 90°, 120°, 150°, or 180°. For half of the subjects (Experiment 5A), all rotation occurred in the depth plane. For the other half (Experiment 5B), rotation occurred in the 2D plane. There were 5, 10, or 15 items on each search display.

Subjects completed 12 practice trials and 720 experimental trials. All conditions were randomly intermixed in the experiment.

Equipment

Subjects were tested individually in a room with normal interior lighting. They sat at an unrestrained distance of about 57 cm from the computer screen at which distance 1 cm corresponds to 1° visual angle. Experiments 1-3 were coded in MacProbe (Hunt, 1994), and Experiments 4-5

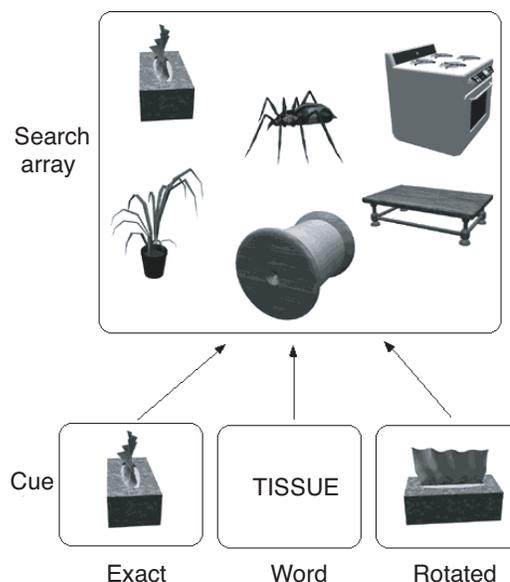


Figure 2. Examples of three cue types and a search display used in Experiments 4 and 5. The stimuli are not drawn to the proper scale.

were coded in MATLAB with the help of the Psychophysics Toolbox (Brainard, 1997). Each experiment lasted for approximately 45 min.

Results

For each subject, we analyzed accuracy for all trials and mean RT for correct trials. Trials with extreme RTs longer than 5000 ms or shorter than 100 ms were not included in the RT analysis. To calculate search slope, we analyzed the slope of RT as a linear function of set size in each condition for each subject. The group mean was reported in the following analyses.

Experiment 1: Incidental cue: Exact vs. uninformative

Because the target was defined by a separate criterion (“vertical symmetry”) than the cue, the cue provided incidental information. There were two types of cue: an *exact* cue and an *uninformative* cue (the cue was a square), two cue leading times (200 or 500 ms), and three set sizes (5, 10, or 15).

Accuracy

Accuracy ranged from 96.4% to 99.1% in different conditions. It was significantly affected by cue type, with higher accuracy to an *exact* cue than an *uninformative* cue, $F(1, 6) = 9.26, p < .023$. Accuracy was not affected by cue leading time, set size, or any interaction effects (all p values $> .15$). Because accuracy was close to ceiling, our conclusions will be drawn primarily on the basis of RT data.

RT

Figure 3 shows mean RT as a function of cue type, cue leading time, and set size.

A repeated-measures ANOVA on cue type, cue leading time, and set size revealed significant main effects of all factors. RT was faster when the cue leading time was 500 ms than when it was 200 ms, $F(1, 6) = 12.16$, $p < .013$, when the cue was exact rather than uninformative, $F(1, 6) = 85.68$, $p < .001$, and when there were fewer items on the display, $F(2, 12) = 439.06$, $p < .001$. There was also a significant interaction between cue type and set size, $F(2, 12) = 21.15$, $p < .001$, as shown by the shallower search slope for the *exact cue* than the *uninformative cue* condition. No other interaction effects were significant ($F_s < 1$).

We calculated search slope of RT as a linear function of set size. For uninformative cues, search slope was 110 ms/item with 200-ms cue and 98 ms/item with 500-ms cue. These were significantly steeper than slopes for exact cues: 65 ms/item with 200-ms cue and 59 ms/item with 500-ms cue.

To investigate whether there was any advantage for repeating the same target object on sequential trials, we separated trials whose target object was the same as the previous trial from other nonrepeated trials. This analysis revealed no difference in mean RT between repeated and nonrepeated trials ($p < .20$). This lack of an effect was due primarily to the small number of repeated trials (of about 10), making our data unsuited for analyzing sequential effects. Thus, for the rest of the article, we will focus only on the within-trial cue effect.

Our results suggest that specific visual information about the target speeds up visual search. This advantage is reflected in accuracy, search RT, and the slope of RT-set size function. In addition, the visual system is highly efficient at using target-specific information: Cueing the target 200 ms ahead of the search display provides nearly as much benefit as cueing the target 500 ms ahead of time.

Experiment 2: Incidental cue: Size and orientation changes

The first experiment showed a significant advantage in visual search when subjects knew the exact shape of the target. But do we really need to have an exact target template to search efficiently? What differences between the cue and the target can the visual system tolerate? To address these questions, in Experiment 2 we showed subjects a cue that differed from the target in size or orientation. Subjects continued to search for a unique, vertically symmetric object such that the cue was incidental to the task.

Accuracy

Accuracy ranged from 96% to 99% in different conditions. It was not significantly affected by cue type, set size, or their interaction (all p values $> .10$).

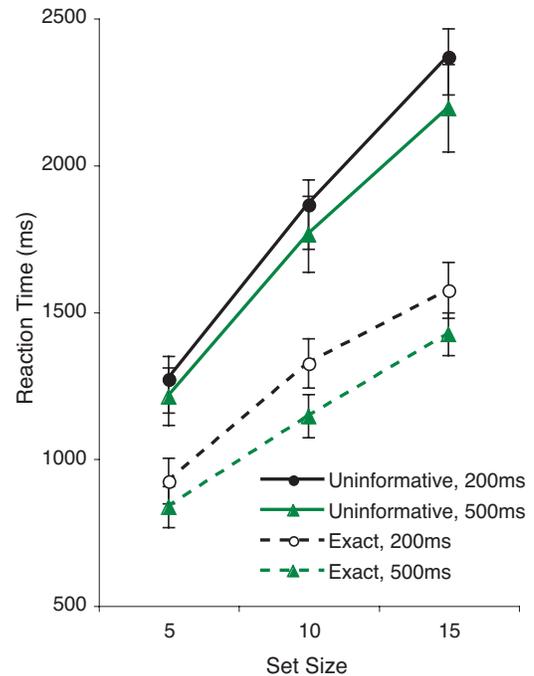


Figure 3. Mean RT data from Experiment 1 ($N = 7$). Search RT was faster and search slope shallower in the *exact cue* than the *uninformative cue* conditions.

RT: Exact, uninformative, small, & rotated cues

When all eight cue types were entered into an ANOVA, there was a significant main effect of cue type, $F(7, 77) = 17.13$, $p < .001$, a significant main effect of set size, $F(2, 22) = 387.77$, $p < .001$, and a significant interaction between the two variables, $F(14, 154) = 2.13$, $p < .013$. Figure 4 shows the RT data.

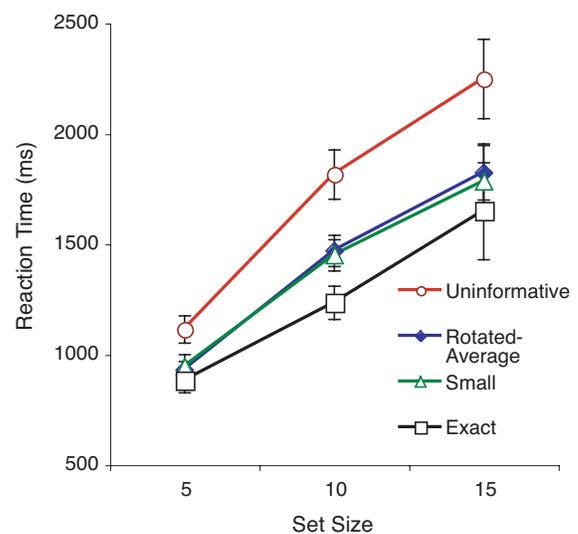


Figure 4. Mean RT data from Experiment 2 ($N = 12$). Size and orientation mismatched led to intermediate RT between the *exact cue* and the *uninformative cue* conditions.

Just like [Experiment 1](#), RT was faster and search slope shallower in the *exact cue* condition than the *uninformative cue* condition. The main effect of condition (*exact* vs. *uninformative*) was significant, $F(1, 11) = 68.31$, $p < .001$, as was the interaction between condition and set size, $F(2, 22) = 10.61$, $p < .001$. Search slope was 77 ms/item with an *exact cue* and 113 ms/item with an *uninformative cue*.

Compared with an *uninformative cue*, a *small cue* sped up RT as well as search slope. The main effect of condition (*small* vs. *uninformative*) was significant, $F(1, 11) = 42.77$, $p < .001$, as was the interaction between condition and set size, $F(2, 22) = 5.68$, $p < .01$. Search slope was 84 ms/item with a small cue.

Similarly, a *rotated cue* sped up RT and search slope when compared with an *uninformative cue*. The main effect of condition (*rotated-average* vs. *uninformative*) was significant, $F(1, 11) = 38.55$, $p < .001$; the interaction between condition and set size was also significant, $F(2, 22) = 5.05$, $p < .016$. Search slope was 89 ms/item with a rotated cue.

There was no statistical difference between the *small cue* and the *rotated cue*, in either overall RT or search slope ($F_s < 1$). Although both types led faster speed than an *uninformative cue*, they were both significantly slower in RT than the *exact cue* ($p_s < .003$). In addition, search slope was numerically larger for *small cue* and *rotated cue* than for *exact cue*, although the slope effects failed to reach statistical significance ($p > .30$ for *small cue*, and $p > .10$ for *rotated cue*).

RT: Rotated cues – angle of rotation

In the previous analysis, we averaged all rotated-cue conditions together and found that orientation discrepancy between the cue and the target led to an RT cost. In this analysis, we examined whether the cost was smaller when the angular difference between the cue and the target was smaller. [Figure 5](#) shows the RT data.

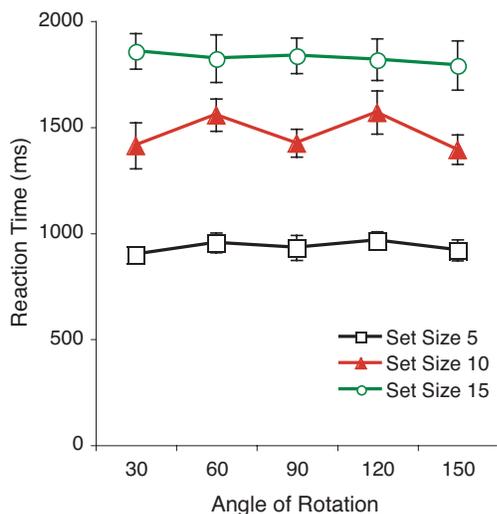


Figure 5. [Experiment 2](#). RT was not unaffected by the angle of rotation between the cue and the target.

An ANOVA on rotation angle (30° to 150°) and set size (5, 10, or 15) revealed a significant main effect of set size, $F(2, 22) = 348.01$, $p < .001$. However, the main effect of rotation angle was not significant, $F(4, 44) = 1.57$, $p > .19$, nor was the interaction between rotation angle and set size significant, $F(8, 88) < 1$.

Discussion

This experiment reveals several properties of the role of visual information in setting up the target template. First, an exact template of the target is most beneficial for search. Compared with a more general description of the target (e.g., “vertical symmetry”), an exact cue facilitates search RT and reduces the slope of the RT-set size function. Second, a cue that is smaller or viewed from a different angle than the target conveys an advantage during search, although not as large as that of an exact cue. This observation suggests that in addition to the target’s shape, incidental properties such as size and orientation are incorporated in the target’s template. This finding is important because it suggests that the visual system does not hold an invariant description of the target during search. Instead, visual details of the target, including its size and orientation, are included in setting up the template and become important cues during the search process.

Experiment 3: Deliberate cue: Size and orientation changes

[Experiment 2](#) suggests that visual details about the target, such as its size and orientation, are included in the setup of a target template. A puzzling finding is that although a rotated cue resulted in slower RT than an exact cue, the specific angular disparity between the cue and the target had little effect on RT. Subjects were just as slow responding to a target that is 30° away from the cue as one that is 90° away from the cue. This observation is puzzling because many studies have found that the time required to recognize an object is proportional to the angular disparity between the current view and the object’s canonical view (Tarr, 1995; Tarr & Pinker, 1989).

One reason for the lack of an orientation effect might be that subjects did not find it necessary to engage in mental rotation when the target was defined by an additional criterion (i.e., vertical symmetry). Results might have been different if subjects were compelled to rely on the cue shape to find the target.

In [Experiment 3](#), subjects were first shown a cue that was always at 0° (i.e., the cue was vertically symmetric). Then they saw an array of different polygons at various orientations and searched for the object that matched the cue. Because the target was no longer uniquely defined by a separate criterion, the cue was the target-defining criterion. We did not test the *uninformative cue* condition, but all other conditions tested in [Experiment 2](#) were also tested here.

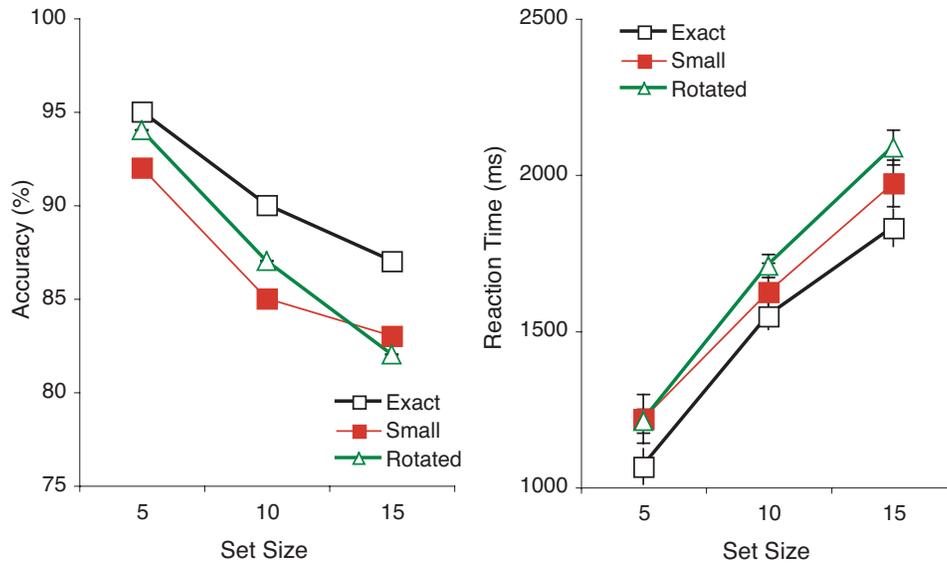


Figure 6. Mean accuracy and RT data from Experiment 3 (N = 12). When subjects searched for the cued object, their performance was the best with *exact cues*.

Accuracy

Accuracy in this experiment was substantially lower than that in Experiments 1-2, presumably because subjects might forget the cue shape and were therefore unable to identify the target. Figure 6 (left) shows the mean accuracy data.

An ANOVA revealed a significant main effect of cue type (*exact*, *small*, and *rotated-average*), $F(2, 22) = 5.96, p < .009$, a significant main effect of set size, $F(2, 22) = 20.27$, but no interaction ($F < 1$). Planned contrasts showed that accuracy in the *exact cue* condition was higher than both the *small cue* and the *rotated cue* conditions ($ps < .05$), but the latter two were not significantly different from each other ($p > .20$).

RT: Exact, small, & rotated cues

RT results were similar to accuracy results. There were main effects of cue type (*exact*, *small*, and *rotated*), $F(2, 22) = 5.89, p < .009$, and set size, $F(2, 22) = 216.94, p < .001$, but no interaction between the two ($F < 1$). In particular, *exact cue* led to faster RT than both *small cue* and *rotated cue* ($ps < .02$), but the latter two did not differ significantly from each other ($F < 1$). The slope of RT as a linear function of set size was 77 ms/item for an *exact cue*, 75 ms/item for a *small cue*, and 88 ms/item for a *rotated cue*.

RT: Rotated cues – angle of rotation

To examine whether search RT depends on the angular disparity between the cue and the target, we compared the five rotated-cue conditions (results shown in Figure 7).

An ANOVA on angle of rotation (30° to 120°) and set size revealed no main effect of rotation ($F < 1$), and no interaction between rotation and set size, $F(8, 88) = 1.45, p > .18$. The angle of rotation also had no effect on accuracy.

Discussion

Experiment 3 replicated several findings from Experiment 2. First, visual search was the fastest when the cue matched the target exactly. If the cue deviated from the target in size or orientation, RT increased significantly. Second, the angle of rotation between the cue and the target had virtually no effect on RT: Whether the cue differed from the target by 30°, 90°, or 150°, search RT was approximately the same. We will discuss the effect of visual mismatch between the cue and the target after presenting results from Experiments 4 and 5.

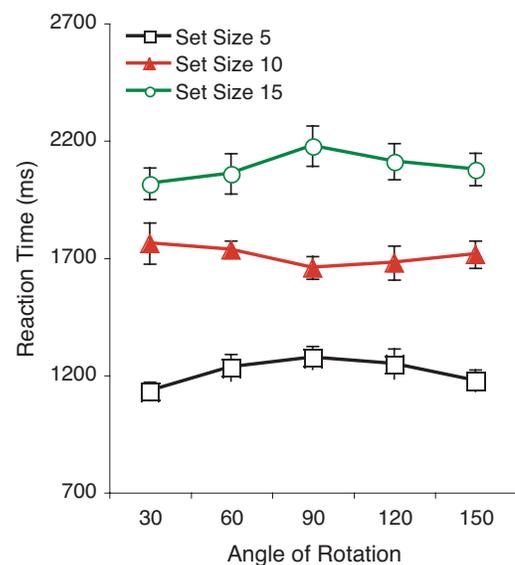


Figure 7. Experiment 3. RT was unaffected by the angle of rotation between the cue and the target.

Experiment 4: 3D objects: Exact, word, & rotated cues

The previous experiments employed novel, meaningless shapes. Here we wish to extend the advantage of an exact cue over other cue types to real-world objects. [Experiment 4](#) used 3D models of real-world objects. We tested subjects in three cue types: *exact cue*, *word cue*, and *90° rotated cue*. In addition, there were three cue leading times (200 ms, 400 ms, and 1000 ms) and two set sizes (8 and 16).

Accuracy

Accuracy ranged from 92% to 99% under various conditions. It was significantly affected by cue leading time, in that it was lower when the cue led by 1000 ms compared with shorter cue times, $F(2, 28) = 4.99$, $p < .014$. Accuracy was also higher for exact cue than rotated and word cues, $F(2, 28) = 7.40$, $p < .003$. No other effects were significant on accuracy.

RT

Mean RT by set size for each cue and cue leading time are shown in [Figure 8](#). An ANOVA on cue type (*exact*, *word*, and *rotated*), cue leading time (200, 400, and 1000 ms), and set size (8 and 16) revealed a significant main effect of cue type, $F(2, 28) = 162.82$, $p < .001$, with fastest RT for the *exact cue*, and slowest RT for the *word cue*. The main effect of cue leading time was not significant, $F(2, 28) = 1.35$, $p > .25$, but the main effect of set size was significant, $F(1, 14) = 412.46$, $p < .001$. Only one interaction effect was significant, that between cue type and cue leading time, $F(4, 56) = 3.26$, $p < .05$. This was accounted for by the fact that whereas the *word cue* became more effective as the cue leading time increased, the *exact cue* became less effective. No other interaction effects were significant (all p s $> .10$).

Comparing *exact cue* with *word cue*, RT was significantly slower in the *word cue* condition ($p < .001$). In addition,

there was a significant interaction between cue type and cue leading time ($p < .017$), in that the *word cue* was more effective, whereas the *exact cue* was less effective with longer cue leading time. The effect of an exact cue diminished with increasing cue-target interval ($p < .03$), perhaps because the exact cue produced both perceptual priming and advanced cueing effects (Wolfe et al., 2004). Presumably perceptual priming would decay at longer cue-target intervals, accounting for the reduction in the cueing effect.

Comparing *exact cue* with *rotated cue*, RT was significantly slower in the *rotated cue* condition ($p < .001$), but this factor did not interact with other factors.

Finally, comparing *word cue* with *rotated cue*, RT was significantly slower in the *word cue* condition ($p < .001$). That is, a visual cue that did not match the target exactly was still more advantageous than a semantic cue. These results suggest that even with 3D models of real-world objects, the target template contains primarily visual details of the target, rather than abstract, semantic labels.

Experiment 5: 3D objects: Various angles of rotated cue

[Experiment 4](#) shows that RT slowed down when the cue and the target differed in orientation by 90°. To further examine the role of orientation, in [Experiment 5](#) we parametrically manipulated the angle of disparity between the cue and the target. The disparity was 0° (*exact cue*), 30°, 60°, 90°, 120°, 150°, and 180°. We ran two versions of the experiment. In [Experiment 5A](#), all rotation occurred in the depth plane, whereas in [Experiment 5B](#), all rotation occurred in the 2D plane.

Experiment 5A: Rotation in the depth plane

Accuracy ranged from 92.7% to 100%. It was significantly affected by angle of rotation, $F(6, 42) = 4.26$, $p < .002$, but not by set size or the interaction between the two (p s $> .10$).

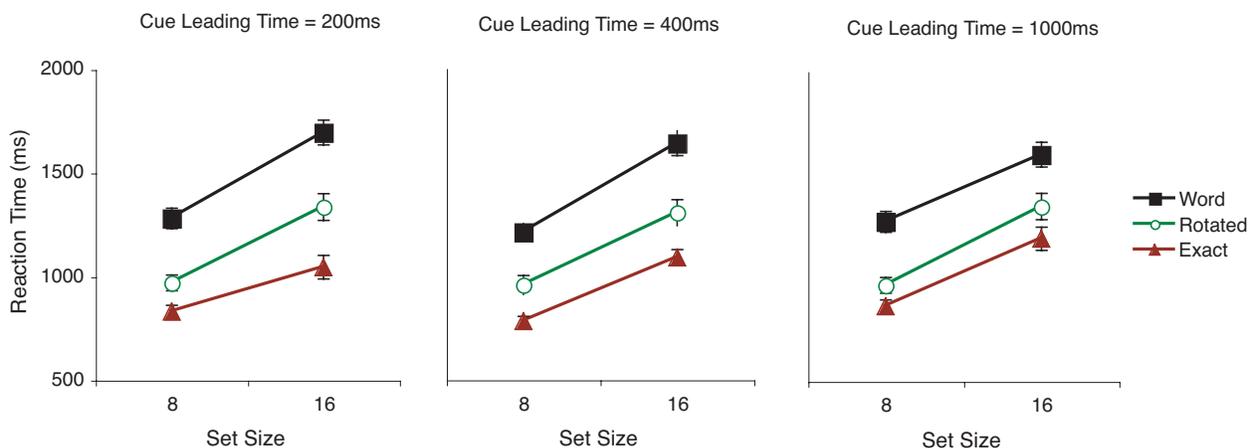


Figure 8. Mean RT data from [Experiment 4](#). The *exact cue* was more effective than the *rotated cue*, which was in turn more effective than the *word cue* ($N = 8$).

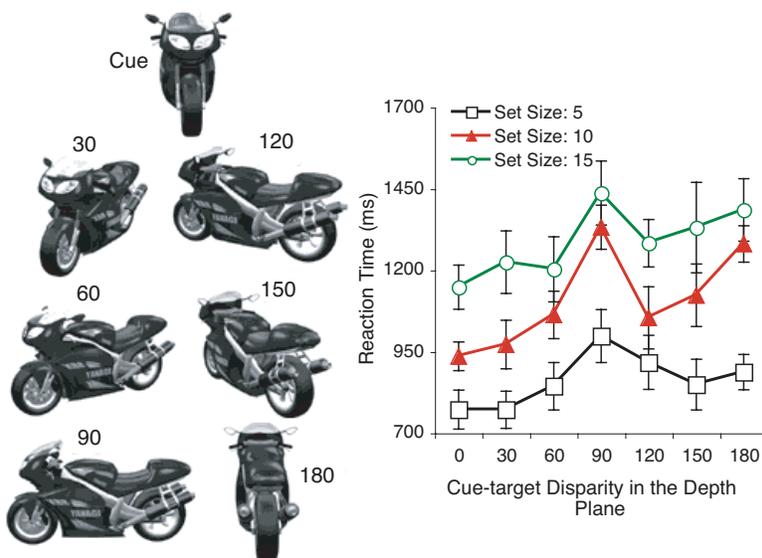


Figure 9. Left. An example of images of a motorcycle changing its view from 0° to 180° in depth. Right. Mean RT data from Experiment 5A: As the orientation difference between the cue and the target increased from 30° to 90°, RT increased ($N = 8$).

When all angles of rotation – including 0° (*exact cue*) – were included in the analysis, RT was significantly affected by orientation, $F(6, 42) = 73.50, p < .001$, set size, $F(2, 14) = 3.52, p < .05$, and their interaction, $F(12, 84) = 3.92, p < .001$.

Even when the 0° condition (*exact cue*) was omitted from the analysis, RT was still significantly affected by orientation, $F(5, 35) = 83.38, p < .001$. The interaction between orientation and set size, however, was only marginally significant, $F(10, 70) = 1.74, p < .09$.

Figure 9 shows RT results from Experiment 5A. As the orientation disparity between the cue and the target increased from 0° to 90°, RT increased significantly. RT then dropped as the disparity increased further from 90° to 120°.

RT increased from 120° to 180° for set sizes 10 and 15.

The fact that 90° disparity produced the largest cost is perhaps not surprising. Because most real-world objects are left-right symmetric, but not front-side symmetric, a 90° view change resulted in the greatest loss of visual detail seen from one angle, and the greatest appearance of new visual details not previously seen.

Experiment 5B: Rotation in the 2D plane

Accuracy ranged from 96.4% to 99%. It was not significantly affected by angle of rotation ($F < 1$), set size, $F(2, 14) = 1.51, p > .25$, or their interaction ($F < 1$).

Figure 10 shows RT results from Experiment 5B. When all angles of rotation – including the *exact cue* condi-

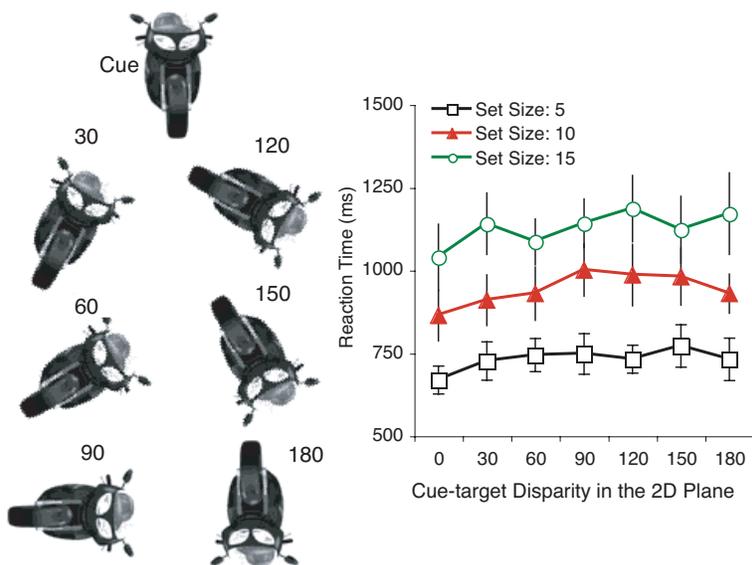


Figure 10. Left. An example of images of a motorcycle changing its view from 0° to 180° in 2D plane. Right. Mean RT data from Experiment 5B: As the orientation difference between the cue and the target increased from 30° to 90°, RT increased ($N = 8$).

tion (0°) – were included in the analysis, RT was significantly affected by orientation, $F(6, 42) = 53.99$, $p < .001$, and set size, $F(2, 14) = 8.56$, $p < .004$, but not their interaction, $F(12, 84) = 1.34$, $p > .20$. The *exact cue* led to significantly faster RT than rotated cues averaged together, $F(1, 7) = 15.32$, $p < .006$.

Even when the *exact cue* was excluded from data analysis, a significant main effect of orientation (from 30° to 180°) remained, $F(5, 35) = 63.38$, $p < .001$. RT increased gradually as the angle of disparity between the target and the cue increased from 30° to 90° . The rotation effect between 90° and 180° was less orderly.

3D versus 2D rotation

To find out whether rotation in depth produced different effects from rotation in the 2D plane, we conducted an ANOVA using rotation plane (2D vs. 3D) as a between-subject factor and angle of rotation (0° to 180°) and set size (5, 10, or 15) as within-subject factors. This analysis revealed no main effects of rotation plane, $F(1, 14) = 2.58$, $p > .13$. The interaction between angle of rotation and rotation plane was not significant ($F < 1$). However, there was a significant interaction between set size and rotation plane, $F(2, 28) = 14.46$, $p < .001$. This observation attributed to the steeper search slope for the 3D rotation experiment (42 ms/item) than for the 2D rotation experiment (17 ms/item). The three-way interaction was not significant ($p > .35$). Thus, rotation in 3D resulted in a steeper search slope than rotation in 2D. In both cases, RT was slowed down by rotation. The larger the angular disparity between the cue and the target (between 0° and 90°), the larger the RT cost.

Discussion

Setting up the target template: Efficiency

What do we need to know about the target object for efficient visual search? Our study suggests that when subjects are given a short time to prepare, cueing the exact target object is most advantageous. We find that setting up an exact cue does not take a long time: Most of the advantage due to the exact cue is gained with only a 200-ms cue leading time, although subjects are still slightly improving in the interval from 200 ms to 500 ms.

The speed of setting up an exact target template should be contrasted with the speed of setting up the target template on the basis of a semantic cue. When provided with a word cue (e.g., “MOTORCYCLE”) or with a general description of the target (e.g., “vertical symmetry”) subjects were slow using such information. In the case of a *word cue*, search was faster if the cue led by 1000 ms rather than 200 ms, suggesting that using semantic labels to set up a target template requires more time.

The speed in switching target templates on the basis of an exact cue is also fast when compared with the speed in switching *tasks*. Studies by Rogers and Monsell (1995), Meiran (1996), Wylie and Allport (2000), and others suggest that *switching tasks* is slow. When subjects have to switch between reporting the vertical position of a dot and reporting its horizontal position, they show a large switching cost even with an advanced cue leading time of 1200 ms. Although switching perceptual targets and switching tasks may rely on similar cognitive and neural mechanisms (Jiang & Vickery, 2004), their efficiency is not the same. Target switch on the basis of an exact cue is fast and efficient, whereas task switch is slow.

Mismatches in size and orientation

The above conclusions – that setting up an exact target template is faster than setting up a semantic template – have also been reached by Wolfe et al. (2004) using different stimuli, such as colored lines. Our study has gone beyond Wolfe et al. by comparing exact cue and semantic cue with cues that are progressively more dissimilar to the target. When subjects search for the target without the correct information about the target’s size or orientation, how well can they tolerate such changes?

Our results suggest that size or orientation differences between the cue and the target are tolerated to some degree. A small cue or a rotated cue still provides an advantage compared with a semantic cue or an uninformative cue. This advantage is reflected in search speed when compared with a word cue, and in both speed and search slope when compared with an uninformative cue. Shape (or size) matches between the target and the cue provide this benefit. Nonetheless, the visual system does not set up an orientation - or size - invariant description of the target. A small cue or a rotated cue leads to slower RT compared with an exact cue, presumably because subjects use the exact information provided by the cue, and the mismatch in size or orientation between the cue and the target slows down RT (for an exception, see Vickery & Jiang, 2004).

Various rotation angles

For rotated cues, Experiments 2-3 suggest that the angular disparity between the target and the cue did not affect search RT. Whether the orientation difference was 30° or 90° , the cost of mismatch in orientation was constant. However, in Experiment 5, we found that the cost was larger when the cue and the target differed by 90° rather than 30° . The latter results held whether rotation occurred in the 2D plane or in depth. How can we reconcile the difference between Experiments 2-3 and Experiment 5?

One possibility is that subjects did not do any mental rotation in Experiments 2-3. Instead, they picked out key features (e.g., a really sharp corner next to two blunt corners) and looked for shapes with these features as the tar-

get. Assuming that in addition to feature-matching, subjects could also do exact template-matching for a 0° rotation; these strategies could lead to a fast exact match, but a slow response to rotated targets that was unaffected by the angle of rotation. Subjects did not use rotation, perhaps because the stimuli used in [Experiments 2-3](#) did not support effective rotation. These random polygons do not have intrinsic up, down, left, or right orientations. Because subjects had no way to determine the canonical upright for an object, they would have difficulty knowing how much to rotate, or in which direction – clockwise or counter-clockwise – an object should be rotated. This ambiguity might have prevented subjects from performing rotation. Alternatively, subjects might have rotated the target to match the cue, but an object that was tilted 30° to the left might have to be rotated by 30° to match the cue on some trials, and by 330° on other trials. Similarly, an object that was tilted 90° to the left might be rotated by 90° to match the cue on some trials, and by 270° on other trials. On average, the amount of rotation would be 180°, independent of the angular disparity between the target and the cue. With 3D models of real-world objects, the situation would be different. Because these objects have a canonical axis, they would readily support efficient rotation. A 30° object would be rotated by 30° most of the time, while a 90° object would be rotated by 90° most of the time. Thus, while we cannot distinguish lack-of-rotation from inefficient rotation in [Experiments 2-3](#), we believe that the difference between these experiments and [Experiments 4-5](#) lies primarily in the presence or absence of a canonical orientation for objects.

A second reason why angular disparity was important in [Experiments 4-5](#) was that rotation in the depth plane resulted in the loss of visual details about the target. This would also contribute to an orientation effect because the 90° rotation produced the greatest loss of visual details. In this condition, the limited amount of visual match between the target and the cue might have turned the 90° cue into a semantic cue, thus slowing down RT.

Top-down control

Our results suggest that the process of setting up a target template can be best considered as a top-down control established on the basis of visual information about the target. This process relies heavily on holding a visual template that matches the target exactly. It works best when the cue does not differ from the target in size or orientation. Furthermore, if the cue is presented too far ahead of the trial (e.g., it is presented 1000 ms ahead of the target, with an interstimulus interval of 800 ms), the cue can be forgotten, resulting in reduced accuracy and increased RT. In this respect, the target-biasing signal is not abstract. It must be supported by visual details of the target. Although we did not separate these components, we suspect that cueing the target leads to both an automatic priming effect and a controlled target set up process. Future studies are needed to separate these components.

Conclusions

To conclude, by asking subjects to search for a different target object on each trial and cueing the target object at various durations prior to the search display, this study clarifies the process that allows humans to set up the target template during visual search. Our results show that the template includes visual details of the target, including its size, orientation, and shape. Although a semantic cue can promote successful visual search, it is not nearly as effective as an exact visual cue. Deviation in size or orientation is only partially tolerated, suggesting that the target template does not contain an object-invariant description. These findings should be incorporated in existing visual search models, such as Guided Search (Wolfe, 1994) and the Biased Competition Model (Desimone & Duncan, 1995). Future studies should separate effects of passive priming from active template set up and examine the commonalities and distinctions between target switch and task switch.

Acknowledgments

This research was supported by National Institutes of Health Grant 1R01 MH071788-01 and Army Research Office Grant 46926-LS (YJ). We thank Sidney Burks for data collection, Diyu Chen and Hing Yee Eng for comments on a previous draft, and an anonymous reviewer for suggestions.

Commercial relationships: none.

Corresponding author: Tim Vickery or Yuhong Jiang.

Email: vickery@wjh.harvard.edu or yuhong@wjh.harvard.edu.

Address: 33 Kirkland Street, Cambridge, MA 02138.

References

- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433-436. [[PubMed](#)]
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36, 28-71. [[PubMed](#)]
- Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, 30, 39-78. [[PubMed](#)]
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193-222. [[PubMed](#)]
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review* 96, 433-458. [[PubMed](#)]

- Found, A., & Muller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a "dimension weighting" account. *Perception & Psychophysics*, *58*, 88-101. [PubMed]
- Hunt, S. M. J. (1994). MacProbe: A Macintosh based experimenter's workstation for the cognitive sciences. *Behavior Research Methods, Instruments, & Computers*, *26*, 345-351.
- Jiang, Y., & Vickery, T. J. (2004). *Common neural and cognitive mechanisms for perceptual set switching and task set switching*. Manuscript in preparation.
- Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*, 751-761. [PubMed]
- Kristjansson, A., Wang, D., & Nakayama, K. (2002). The role of priming in conjunctive visual search. *Cognition*, *85*, 37-52. [PubMed]
- Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out. I. Role of features. *Memory & Cognition*, *22*, 657-672. [PubMed]
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 1423-1442.
- Muller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics*, *57*, 1-17. [PubMed]
- Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron*, *26*, 703-714. [PubMed]
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207-231.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing. I. Detection, search and attention. *Psychological Review*, *84*, 1-66.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing. II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127-190.
- Tarr, M. J. (1995). Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review*, *2*, 55-82.
- Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*, 233-282. [PubMed]
- Treisman, A. (1988). Features and objects: The fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, *40*(2), 201-237. [PubMed]
- Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception & Performance*, *16*, 459-478. [PubMed]
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97-136. [PubMed]
- Vickery, T. J., & Jiang, Y. (2004). *Knowing what to look for: Advanced target knowledge in visual search*. Manuscript submitted for publication.
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202-238.
- Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: On the contribution of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception & Performance*, *29*, 483-502. [PubMed]
- 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. [PubMed]
- Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs." *Psychological Research*, *63*, 212-233. [PubMed]