

Target templates: the precision of mental representations affects attentional guidance and decision-making in visual search

Michael C. Hout · Stephen D. Goldinger

© The Psychonomic Society, Inc. 2014

Abstract When people look for things in the environment, they use *target templates*—mental representations of the objects they are attempting to locate—to guide attention and to assess incoming visual input as potential targets. However, unlike laboratory participants, searchers in the real world rarely have perfect knowledge regarding the potential appearance of targets. In seven experiments, we examined how the precision of target templates affects the ability to conduct visual search. Specifically, we degraded template precision in two ways: 1) by contaminating searchers' templates with inaccurate features, and 2) by introducing extraneous features to the template that were unhelpful. We recorded eye movements to allow inferences regarding the relative extents to which attentional guidance and decision-making are hindered by template imprecision. Our findings support a dual-function theory of the target template and highlight the importance of examining template precision in visual search.

Keywords Visual search · Eye movements · Target templates · Attentional guidance · Decision making

The world is replete with visual information of such richness and complexity that our cognitive systems are subject to information overload. It is impossible to simultaneously process all available information on a moment-to-moment basis.

Electronic supplementary material The online version of this article (doi:10.3758/s13414-014-0764-6) contains supplementary material, which is available to authorized users.

M. C. Hout (✉)
Department of Psychology, New Mexico State University,
P.O. Box 30001 / MSC 3452, Las Cruces, NM 88003, USA
e-mail: mhout@nmsu.edu

S. D. Goldinger
Department of Psychology, Arizona State University,
P.O. Box 871104, Tempe 85287, AZ, USA

Imagine yourself sitting in a coffee shop; as you look around, your eyes are flooded with different colors, shapes, textures, contours, and motion. You spot brightly colored boxes of tea on shelves, round reclining chairs in your periphery, the smooth surface of a linoleum floor, and the sharp edges of a table to avoid as you exit. All this raw information is assembled in your mind such that you perceive coherent objects, rather than features. Rather than “squat, opaque, and round,” you see your coffee mug. “Shiny, silver, and cylindrical” is immediately appreciated as a container of milk on the condiment bar. Despite the fluency of converting visual features into meaningful objects, people are limited regarding how much information they can process in any given moment. As such, people process visual scenes through a series of alternating fixations and saccadic eye movements. Over time, regions of space are briefly foveated and objects are stored in visual working memory (VWM), giving rise to stable representations of the world. Chaotic as saccades may seem, eye movements are not random; they are highly governed by low-level, visual characteristics of the environment, prior experience viewing similar scenes, and (most germane to the current investigation) by the mental representation of whatever the observer is seeking.

Guidance of attention by bottom-up and top-down information

When people look for things in the environment, they use three primary sources of information: low-level salience, scene context, and target template information (Malcolm & Henderson, 2010). Low-level salience is a bottom-up information source that helps select regions of contrast, such as changes in color or intensity (Itti & Koch, 2000, 2001; Koch & Ullman, 1985). Higher-level (top-down) knowledge helps identify informative regions of a scene, most likely to contain

target objects (Castelhano & Henderson, 2007; Neider & Zelinsky, 2006). Finally, target “templates” in memory are used to assess visual information, comparing input to representations stored in VWM. Regions that share features with the template are selected for fixation (Rao et al., 2002; Zelinsky, 2008) and, generally, the more similar an item is to the target, the more likely it will be fixated (Eckstein et al., 2007; Findlay, 1997; Becker, 2011; Olivers, Meijer & Theeuwes, 2006; Mannan et al., 2010). Returning to the example of searching for milk at a coffee shop, low-level salience information decomposes the store into regions of coherence, contextual knowledge makes you look on the condiment bar, rather than merchandise display shelves, and a target template helps direct your eyes toward objects that potentially match the target.

Although bottom-up information is necessary for basic visual processing (see Wolfe & Horowitz, 2004), it seems that, relative to the influence of high-level knowledge, the guidance of attention by low-level features has rather limited utility (Einhäuser et al., 2008; Henderson et al., 2007, 2009; Tatler & Vincent, 2008, 2009). For instance, top-down attentional control is entailed even in simple feature-search tasks, which have been traditionally viewed as requiring little overt attention (Wolfe et al., 2003). People can perform remarkable feats of top-down attentional control; when only a subset of items are ever pertinent in a visual search task, people learn to restrict attention to relevant stimuli (Kunar, Flusberg & Wolfe, 2008; Frings, Wentura, & Wühr, 2012) and can learn the features of distractors in order to guide attention away from them (Yang, Chen & Zelinsky, 2009; see also the “preview benefit” in Watson & Humphreys, 1997; Watson & Humphreys, 2000; Watson et al., 2003). Top-down guidance is even strong enough to override attention capture by low-level salience (Chen & Zelinsky, 2006). In essence, top-down guidance works in two ways: it biases attention toward important features or regions, and it biases attention away from undesirable features (or objects that have already been inspected; Al-Aidroos et al., 2012; Arita, Carlisle, & Woodman 2012). It is unsurprising that successful models of visual search, such as Guided Search (Wolfe et al., 1989; Wolfe, 1994; Wolfe & Gancarz, 1996; Wolfe, 2007; Palmer et al., 2011), incorporate top-down guidance as a key mechanism controlling attention.

Target template

The idea of a target template was first proposed by ethologists studying the feeding behavior of birds. When birds feed on insects, they tend to sample the common bugs disproportionately often, suggesting that their behavior is biased in favor of target features that previously resulted in rewards (Tinbergen, 1960; Pietrewicz & Kamil, 1979; Bond, 1983). Regarding humans, the target template (also called the “attentional

template” or “search template”) refers to the VWM representation of the target item and how it facilitates detection (Wolfe et al., 2004; Malcolm & Henderson, 2009; Bravo & Farid, 2009, 2012; Vickery et al., 2005). Research on primates (Evans et al., 2011; Chelazzi et al., 1993, 1998) has shown that activating a search template involves neurons in the prefrontal cortex (PFC) that select and maintain behavioral goals (such as finding a target among distractors). These neurons project to inferotemporal (IT) cortex, where visual objects are believed to be represented (Mruzek & Sheinberg, 2007; Peelen, Fei-Fei, & Kastner, 2009). Importantly, top-down input from PFC enhances the gain in IT neurons that are selective for the target object; in essence, the PFC “tells” IT cortex which representations to favor and which to inhibit (Stokes et al., 2009; Zhang et al., 2011). Moreover, this bias may be relayed to V4 or other early visual areas that encode basic stimulus features (Hon et al., 2009). Neurons in V4 are sensitive to stimulation in one specific region of the retina, the cell’s “receptive field.” Beyond spatial selectivity, these cells also have preferred colors and/or shapes they selectively favor (Wolfe, 2005).

In a study using macaques, Bichot, Rossi, and Desimone (2005) recorded from V4 neurons. On trials where the target was (for instance) red, they found that red-selective neurons increased their firing rates even before the target was seen (i.e., before the monkey prepared an eye movement to the target location). Moreover, other red-selective neurons began to synchronize their activity, as if preparing to respond to feature presence. More recently, Eimer et al. (2011) used event-related potentials (ERP) to study the benefits of advanced preparation in visual search (in human participants). Their results suggested that holding a target template in VWM accelerates target selection and resolves attentional competition by inhibiting neurons that code for irrelevant features. Taken together, these findings suggest a mechanism by which target templates may guide visual search; holding a template in mind enhances firing in cells that respond to relevant features and may inhibit cells that respond to irrelevant features (Desimone & Duncan, 1995; Usher & Neiber, 1996).

Problem of template imprecision

In typical search experiments, participants look for a single, unambiguous target in every trial. For instance, participants may see a picture of a bicycle and try to find that same bicycle in a cluttered array of other objects. In other cases, precise target descriptions are given, such as “find a vertical red bar” in a display with lines of various orientations and colors. In such cases, search templates are highly precise. In natural examples of visual search, people rarely enjoy such template precision. We may look for things defined very loosely (e.g., “something to use for a doorstep”). Other times, we might

have a good target template but cannot anticipate the exact target appearance, such as looking for a particular friend at a high-school reunion. Perhaps you remember your friend as being thinner than he is now, etc. In such cases, activated template features will draw attention to a relevant subset of individuals but likely will not allow the correct person to “pop out” from the crowd. Additionally, perhaps you are hoping to see several old friends and therefore have multiple templates in mind. If so, larger subsets of people will likely draw attention as potential targets, making search more difficult. The present investigation addressed the effects of such template imprecision on visual search. Specifically, we tested how degrees of dissimilarity between expected and discovered targets affected: 1) the efficiency of guiding attention to targets, and 2) the fluency of appreciating targets once attention has fallen upon them.

Manipulating template precision using template-to-target similarity

The concept of similarity is critical to virtually all theories of perception, attention and memory (Goldstone & Medin, 1994; Hout, Goldinger & Ferguson, 2013; Medin, Goldstone, & Gentner, 1993). With respect to visual search, theories have long emphasized how efficiency is affected by the similarity of targets to distractors, and by the similarity of distractors to each other (Duncan & Humphreys, 1989; 1992; Hwang, Higgins, & Pomplun, 2009; Wolfe, 1994; Zelinsky, 2008). Although vast literature exists on target-to-distractor similarity effects, there is relatively little work on template-to-target similarity effects. Prior experiments have compared template-guided search (using picture cues) to categorical search (using word cues). In such experiments, we assume that searchers’ VWM representations differ across conditions. In template-guided search, observers have (nearly) veridical target representations in mind, whereas categorical search affords less precision.

Yang and Zelinsky (2009), for example, had people search for teddy bears using picture or word cues and found that, in the latter case, people searched longer and required more fixations to complete search. In a similar investigation, Schmidt and Zelinsky (2009) argued that the amount of guidance that a cue elicits is directly related to the cue’s categorical specificity. Participants were given five different types of cues: an exact picture of the target (e.g., a picture of a pair of boots), an abstract textual description (e.g., “footwear”), a precise textual description (e.g., “boots”), an abstract plus color textual description (e.g., “brown footwear”), or a precise plus color textual description (e.g., “brown boots”). They found that guidance (indexed by fixation and saccade metrics) increased as more information was added to the cue. Their findings suggest that guidance improves to the extent that visual information can be extracted from the cue and loaded

into VWM (these assumptions were incorporated in two recent search models; Zhang et al., 2006; Navalpakkam & Itti, 2005).

Studies of priming also show the utility of target templates. In a recent study, Wilschut, Theeuwes, and Olivers (2014; see also Wilschut, Theeuwes, & Olivers, 2013) had people search for a target line segment defined by color (a red horizontal bar, or a green vertical bar). Participants were first shown both potential target colors as filled circles (pre-cues) in different spatial locations, followed a fixation dot, and then an “*exact*” or “*neutral*” post-cue. The post-cue indicated which color to search for; it was another circle, presented in the same spatial location as one of the pre-cues. On exact-cue trials, the circle was shown in color (green or red), and on neutral-cue trials, the circle was black. Search arrays were shown very briefly, followed by a backwards mask. Wilschut and colleagues found better search accuracy for exact cues, relative to neutral cues. Importantly, however, they found that the effects were equivalent across cue types when priming effects were accounted for. In a second experiment, target pre-cues were shown either verbally (the words “RED” and “GREEN”) or visually (colored circles), and the spatial post-cue was always a neutral black dot, indicating which color to search for. In this way, visual priming was controlled for, because both target cues were presented on each trial and because the search array was temporally aligned to the neutral post-cue. No differences were found between the verbal or visual cue conditions, suggesting that visual target cues engender superior attentional selectivity due largely to the priming of visual features.

In the present investigation, we employed two new techniques for directly manipulating the precision of searchers’ target templates, controlling their similarity to targets that eventually appeared in search displays. Our first set of experiments (1a through 1d) involved a paradigm wherein search targets were validly cued in most trials: observers were shown targets that appeared, unaltered, in the display. Instructions indicated that targets would occasionally differ from the cues: “*Please search for this item or something very much like it.*” In the remaining trials, the eventual targets were slight variations of the provided cues. Of key interest, there were different levels of variation. For example, Experiment 1a included two levels of variation from the provided cues, denoted as *imprecise* versus *inaccurate*. In imprecise trials, the observer might be shown a coffee mug, and the target would be the same mug, oriented differently. In inaccurate trials, the target would be a different color mug. (Importantly, despite these differing degrees of variation, targets were always unmistakable and performance was highly accurate.) Performance in valid trials was then contrasted with invalid trials, with two levels of dissimilarity. In later variations of Experiment 1, cue-target similarity was manipulated more directly, using multi-dimensional scaling (MDS).

Our second set of experiments (2a through 2c) involved manipulating the “width” of searchers’ template “feature spaces.” In these experiments, observers searched for multiple targets (although single-target, control trials were also included). Two target cues were shown before search, and participants tried to locate either in the search array (only one ever appeared). Width of the template feature space was manipulated by varying the similarity of the two potential targets to one another. Feature spaces ranged from narrow (e.g., a coffee mug oriented in two ways) to broad (e.g., two starfish with different colors and shapes). Across experiments, we therefore examined situations wherein observers’ templates were made imprecise, either by the inclusion of inaccurate features, or the addition of unhelpful ones.

Experiment 1

Experiment 1 tested the degrees to which inaccurate template features would affect visual search. Theoretically, there are two likely functions of a search template. First, it may contribute to attentional guidance, changing the visual system’s “activation map” that sets priorities for objects or regions to fixate, based on similarity to the features in VWM (Malcolm & Henderson, 2009). Second, templates are likely used in target verification, as the criterion to which visual input is compared (Malcolm & Henderson, 2010). To guide this study, we considered three hypotheses regarding how search behavior might change when target templates are imprecise. First, the *attentional guidance* hypothesis suggests that imprecision will hinder the ability to quickly allocate attention to the correct object, as misleading features in VWM will allow competing objects to draw attention. Second, the *decision-making* hypothesis posits that imprecise templates will hinder the process of comparing visual input to VWM, slowing target verification (and distractor rejection). Third, the *dual-function* hypothesis suggests simply that an imprecise target template will hinder both attentional guidance and decision-making during search.

To preview our findings, Experiment 1a established that template imprecision slows search RTs, then Experiment 1b included eye-tracking to decompose search behavior into separate phases. These were *scanning* (eye movement behavior from search initiation until the target was located) and *decision making* (the time from first target fixation until the overt response). We used two dependent measures to characterize search behavior during these phases. *Scan-path ratios* (SPRs) were obtained by summing the amplitude of all saccades (in degrees of visual angle) prior to target fixation and dividing that value by the shortest possible distance between central fixation and the target. Thus, perfect guidance (e.g., pop-out) would yield a ratio equal to one; ratios > 1 would

indicate imperfect guidance, as other locations were visited prior to the target.¹ *Decision times* (DTs) were measured from target fixation to the spacebar press terminating search. In Experiments 1a and 1b, degrees of imprecision (between items stored in VWM and eventual targets) were operationalized by comparing “state” and “exemplar” pairs (from Brady et al., 2008; Konkle et al., 2010). Examples are shown in Fig. 1. In Experiments 1c and 1d, we manipulated template precision using multidimensional scaling measures of similarity among objects (from Hout, Goldinger, & Brady, *under review*) and again tested both search RTs and eye movements (Alexander & Zelinsky, 2011; Godwin, Hout, & Menneer, 2014).

Experiment 1a: method

Participants

Twenty students from Arizona State University participated in Experiment 1a as partial fulfillment of a course requirement. All participants had normal, or corrected-to-normal, vision and all reported normal color vision. (These aspects of the participant pool were true for every experiment reported in this article, and are not reiterated each time.)

Apparatus

Data were collected on up to 12 identical computers simultaneously, all separated by dividers. The PCs were Dell Optiplex 380 systems (3.06 GHz, 3.21 GB RAM) operating at 1366 x 768 resolution on Dell E1912H 18.5” monitors (60 Hz refresh rate). The operating system was Windows XP, and E-Prime v2.0 software (Schneider, Eschman, & Zuccolotto, 2002) was used to control all procedures.

Design

Three levels of *Template Precision* (precise, imprecise, inaccurate) were manipulated within-subjects. In every condition, three levels of *Set Size* (12, 16, 20) were manipulated in equal proportions.

Stimuli

All stimuli came from the “Massive Memory” database (Brady et al., 2008; Konkle et al., 2010; cvcl.mit.edu/MM/)

¹ We also examined “time to fixate” the target as a measure of scanning behavior in each of the eye-tracking studies (e.g., Castelhamo, Pollatsek, & Cave, 2008). The results were entirely consistent with the findings from scan-path ratios. We therefore chose to report SPRs, owing to their straightforward interpretation; specifically, that a SPR of 1.0 indicates perfect attentional guidance.



Fig. 1 Sample state- and exemplar-pair stimuli from the “Massive Memory” database (cvcl.mit.edu/MM/stimuli.html)

stimuli.html). They were photographs of real-world objects, resized (maintaining original proportions) to a range of 2.0° to 2.5° visual angle (horizontal or vertical), from a viewing distance of 55 cm. The pictures contained no background; a single object or entity was present in each image (e.g., an ice cream cone, a pair of shoes).

Procedure

Visual search At the beginning of each trial, participants were shown a target cue and were asked to “search for this item or something very similar to it.” When the participants were ready, they pressed the spacebar to start the trial. This initiated a 500-ms fixation cross, followed by the visual search display, which remained until a response was recorded or 10-s elapsed (timeouts were coded as errors). Participants rested their fingers on the spacebar during search, quickly pressing it upon target location (RTs were measured from display onset to the spacebar press). Responding cleared the images from view, and each image was replaced with a random number (between one and the set size) for 2 seconds (Navalpakkam & Itti, 2007, for a similar approach). The numbers then disappeared and participants indicated which number appeared at the target location, using 2AFC (Fig. 2). Feedback was provided as either a centrally presented green checkmark or a large red X. Feedback for correct trials lasted 1 second; feedback for incorrect trials lasted 2 seconds. Instructions asked participants to respond as quickly as possible while remaining accurate. After four practice trials, there were 360 experimental trials, presented in 4 blocks of 90. There were 240 trials of the precise condition and 60 trials apiece of the imprecise and inaccurate conditions.

Search array organization A search array algorithm was used to create spatial configurations with pseudo-random organization (Fig. 3; Hout & Goldinger, 2012). An equal number of objects appeared in each quadrant of the display (three, four, or five, depending upon set size). Each quadrant was broken down into nine equal “cells,” making the entire display into a 6x6 grid. On each trial, images were placed in random cells (per quadrant); specific locations were selected to ensure a minimum of 1.5° of visual angle between adjacent images, and between any image and the edges of the screen. No images appeared in the four centermost locations of the screen to ensure the participant’s gaze would never immediately fall on a target at onset. Targets appeared equally often in each quadrant of the display.

Stimulus selection In the 240 “precise” trials, targets appeared exactly as cued. In the remaining 120 trials (60 “imprecise” and 60 “inaccurate”), targets were slight deviations from the initial cues; these were either the state or exemplar partners, respectively, of the cue pictures. Participants were told that targets would appear exactly as shown in most trials but would occasionally be slightly different. This procedure encouraged participants to adopt the cue as a search template. In each trial, the target was selected quasi-randomly from among the 100 exemplar- and 100 state-pairs; each cue-target pair was used once or twice in the experiment. In each trial, distractors were selected quasi-randomly from among 240 object categories, chosen such that only one exemplar per semantic category was represented; across trials, no category was repeated until each had been used at least once. The entire set contained 4,040 images; no picture was used more than twice in the entire experiment.

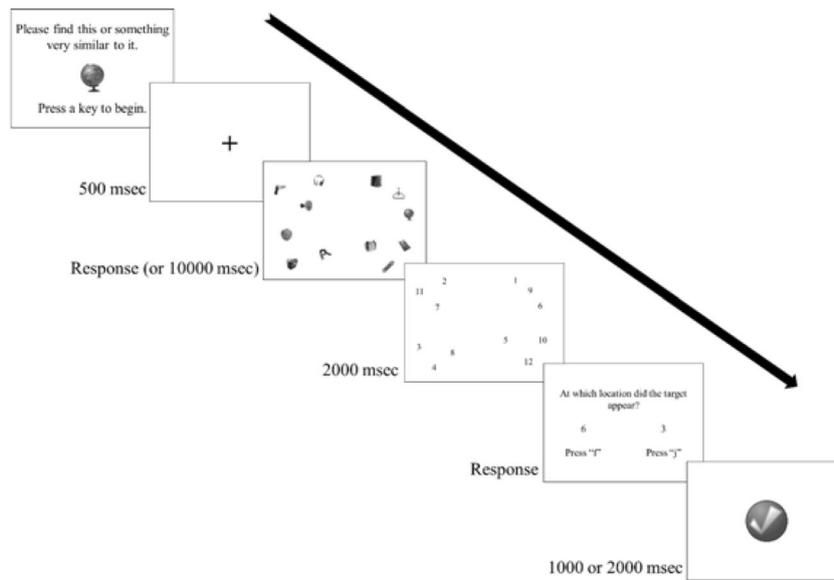


Fig. 2 Visual search trial progression, from Experiment 1a. (Images were presented in full color)

Experiment 1a: results and discussion

Overall, accuracy was very high (in Experiments 1a through 1d, average accuracy was always > 97 %). Although some reliable effects (e.g., Set Size) were observed in accuracy, potential ceiling effects make them uninteresting. Therefore, all reported results for Experiment 1 focus on RT measures and eye-movements. In the RT analyses, although set size and block were included, we focus on the specific findings of interest, involving template precision, and do not unduly focus

on extraneous interactions. Accuracy and RT data were analyzed using 3 (Precision: precise, imprecise, inaccurate) x 3 (Set Size: 12, 16, 20) x 4 Block (1-4) within-subjects, repeated measures ANOVAs. Only RTs from correct trials were analyzed.

An initial validation analysis was conducted to ensure that all stimuli were comparable. When used in precise trials, the stimuli from exemplar and state pairs produced equivalent search RTs (987 and 1025 ms, respectively, $F < 1$). This suggests that there was nothing inherently more difficult about

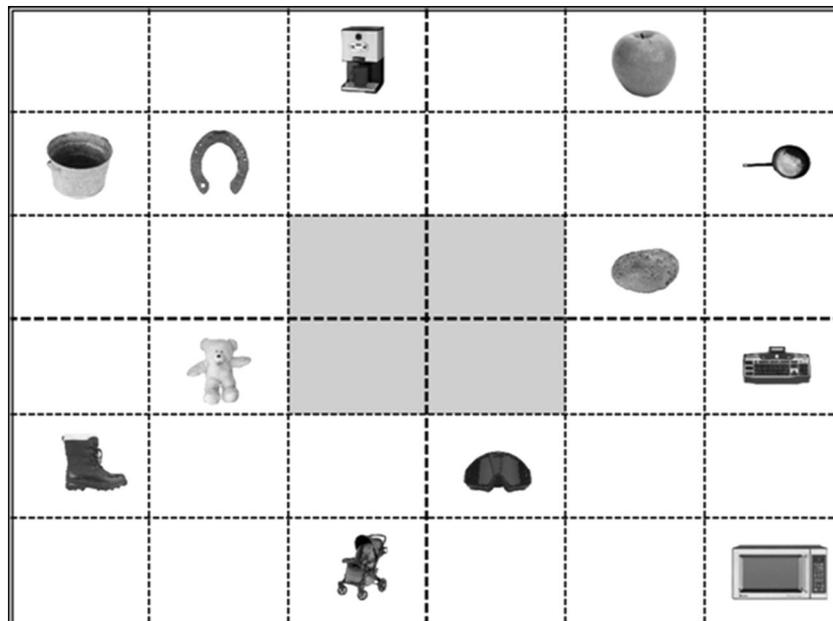


Fig. 3 Sample visual search display, from Experiment 1a, showing the search array organization grid. Gridlines were imaginary; they were added to the figure for clarity. No images ever appeared in the centermost locations, shown here in gray. In the experiment, pictures were shown in full color

the exemplar-pair pictures, relative to the state-pair pictures. RTs as a function of Precision and Block (collapsed across Set Size) are shown in Fig. 4. There was a main effect of Precision, with fastest RTs in precise trials (1006 ms), followed by imprecise (1321 ms) and inaccurate (1941 ms) trials; $F(2, 18) = 77.75, p < 0.01, \eta^2_p = 0.90$. There also were main effects of Set Size (1242, 1424, and 1601 ms for sizes 12, 16, and 20, respectively), $F(2, 18) = 24.05, p < 0.01, \eta^2_p = 0.73$, and Block (1579, 1473, 1360, and 1277 ms for Blocks 1-4), $F(3, 17) = 13.53, p < 0.01, \eta^2_p = 0.70$: RTs increased with increasing set size and decreased across blocks. There was a Precision x Block interaction ($F(6, 14) = 5.89, p < 0.01, \eta^2_p = 0.72$); search RTs improved the most when templates were less precise (improvements of 128, 245, and 533 ms across blocks for the precise, imprecise, and inaccurate conditions, respectively).

Experiment 1a showed that template imprecision impaired search RTs, but that difference diminished over blocks. Search RTs remained relatively flat in the precise condition, improved across blocks in the imprecise condition, and improved most in the inaccurate condition. There are (at least) two reasons for this interaction. First, over time, participants may have become more resilient to inaccurate target cues, implicitly adopting broad (perhaps categorical) templates based on the cue, rather than anticipating the specific features shown before search. This would appear mainly as an improvement in decision making, once targets had been located. (It seems unlikely that adopting a broader template would lead to more efficient attentional guidance.) Second, and less interesting, is a potential ceiling effect: RTs were fast overall, hovering around 1000 ms in all blocks of the precise condition. The interaction may have arisen, because only the harder conditions afforded more “room” to improve. In Experiment 1b, we followed up on this question in two ways. First, eye-tracking

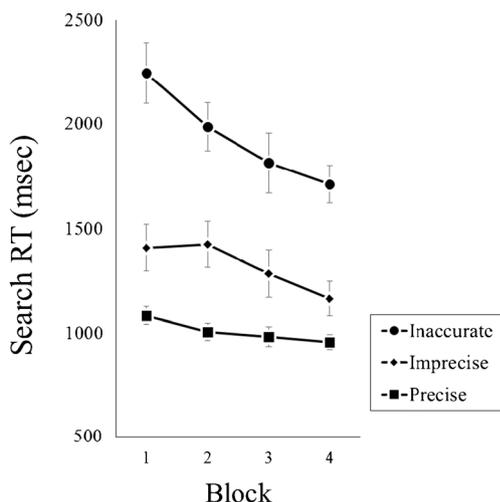


Fig. 4 Mean visual search reaction time (on accurate trials) from Experiment 1a, presented as a function of Template Precision and Block. Error bars represent 1 standard error (SE) of the mean

allowed us to decompose RTs into scanning and decision-making phases. We expected to see more fluent decision-making, without a concurrent increase in scanning efficiency.

Second, we introduced a new manipulation of *cue validity* in Experiment 1b: Across participants, different proportions of mismatch trials were used, based on the expectation that people will adhere to the cues in proportion to their information value. If the cues rarely predict the appearance of the actual targets, an ideal searcher may choose to only loosely base search templates on them. Indeed, a recent study by Machizawa et al. (2012) suggests that the precision of VWM representations can be controlled at will (this study is discussed later in further detail). Experiment 1b introduced a *Match Frequency* manipulation, varying the proportions of perfect matches. For the high match frequency group, we expected faster search RTs in the “precise” condition, relative to the low match frequency group. However, in the “inaccurate” condition, we expected a reversal, wherein the high match frequency group would be slower, due to their strict adherence to the initial cues.

Experiment 1b: method

Participants

Twenty-nine new Arizona State University students participated in Experiment 1b. There were 10, 10, and 9 participants (respectively) in the low, medium, and high match frequency groups.

Apparatus

The stimuli were unchanged from Experiment 1a, but data were now collected using a Dell Optiplex 755 PC (2.66 GHz, 3.25 GB RAM) with a 21-inch NEC FE21111 CRT monitor, (resolution 1280x1024, refresh rate 60 Hz). Eye-movements were recorded using an Eyelink 1000 eye-tracker (SR Research Ltd., Mississauga, Ontario, Canada), mounted on the desktop. Temporal resolution was 1000 Hz, and spatial resolution was 0.01° . An eye movement was classified as a saccade when its distance exceeded 0.5° and its velocity reached $30^\circ/s$ (or acceleration reached $8000^\circ/s^2$). Viewing was binocular, but only the left eye was recorded.

Design

The design was identical to Experiment 1a, with two exceptions. First, the frequency of precise trials was manipulated between-subjects, with three levels of match frequency: the low, medium, and high conditions corresponded to 20 %, 53 %, and 80 % precise trials, respectively. Second, there were now only three blocks, with 90 trials each.

Procedure

The procedure was identical to Experiment 1a, with the exception of details pertaining to eye-tracking. Participants used a chin-rest during all search trials, adjusted so the eyes were fixated centrally on the computer screen when the participant looked straight ahead. The procedure began with a calibration routine to establish a map of the participant's known gaze position, relative to the tracker's coordinate estimate of that position. The routine proceeds by having participants fixate a small black circle as it moves to 9 different positions (randomly) on the screen. Calibration was accepted if the mean error was less than 0.5° of visual angle, with no error exceeding 1° of visual angle. Periodic drift correction and recalibrations ensured accurate recording of gaze position throughout the experiment. Interest areas (IAs) were defined as the smallest rectangular area that encompassed a given image.

The trial procedure was modified to include a gaze-contingent fixation cross. When the fixation cross appeared, participants had to direct their gaze to it for 500 ms to initiate the search display. If they did not do this within 10 seconds, due to human error or calibration problems, the trial was marked as incorrect, and a recalibration was performed before the next trial.

Experiment 1b: results and discussion

Data were analyzed using 3 (Precision: precise, imprecise, inaccurate) \times 3 (Match Frequency: low, medium, high) \times 3 (Set Size: 12, 16, 20) \times 3 Block (1-3) mixed-model, repeated measures ANOVAs. Match Frequency was the only between-subjects factor. We included two new dependent measures, obtained via eye-tracking: 1) scan-path ratios (SPR), and 2) decision-time (DT). Only correct trials were analyzed, and SPRs and DTs were not analyzed for any trials in which the target was not directly fixated.

Response times

The RT results, SPRs, and DTs are shown (in separate panels) in Fig. 5. There was a main effect of Precision, with fastest RTs in precise trials (1146 ms), followed by imprecise (1390 ms) and inaccurate trials (1889 ms); $F(2, 25) = 79.55$, $p < 0.01$, $n_p^2 = 0.86$. There was no main effect of Match Frequency ($F < 1$). There were reliable effects of Set Size (1339, 1449, and 1637 ms, for 12, 16, and 20, respectively), $F(2, 25) = 23.84$, $p < 0.01$, $n_p^2 = 0.66$, and Block (1581, 1531, and 1313 ms for Blocks 1-3), $F(2, 25) = 28.49$, $p < 0.01$, $n_p^2 = 0.70$. There was a Precision \times Block interaction, $F(4, 23) = 7.90$, $p < 0.01$, $n_p^2 = 0.58$, indicating that performance improved to the greatest degree when templates were less precise

(improvements across blocks of 78, 294, and 433 ms for the precise, imprecise, and inaccurate conditions, respectively). No other interactions were significant ($F_s < 2$).

Scan-path ratios

There was a main effect of Precision, with most optimal SPRs in precise trials (1.63), followed by imprecise (1.79) and inaccurate (2.50); $F(2, 25) = 66.98$, $p < 0.01$, $n_p^2 = 0.84$. There was no main effect of Match Frequency ($F < 1$). There also were main effects of Set Size (1.70, 1.96, and 2.26, for 12, 16, and 20, respectively), $F(2, 25) = 25.73$, $p < 0.01$, $n_p^2 = 0.67$, and Block (2.01, 2.03, and 1.88 for Blocks 1-3), $F(2, 25) = 3.42$, $p < 0.05$, $n_p^2 = 0.22$. There was a Precision \times Block interaction, $F(4, 23) = 3.59$, $p < 0.05$, $n_p^2 = 0.38$, indicating that performance improved most across blocks when templates were imprecise or inaccurate (improvements of 0.23 and 0.20, respectively), but not when templates were precise (decrement of 0.05). There was a Set Size \times Block interaction, $F(4, 23) = 4.45$, $p < 0.01$, $n_p^2 = 0.44$, and a Precision \times Match Frequency \times Block interaction, $F(8, 46) = 2.21$, $p < 0.05$, $n_p^2 = 0.28$. No other interactions were significant ($F_s < 2$).

Decision times

There was a main effect of Precision, with fastest DTs in precise trials (443 ms), followed by imprecise (621 ms) and inaccurate (869 ms); $F(2, 25) = 47.04$, $p < 0.01$, $n_p^2 = 0.79$. There were no main effects of Match Frequency or Set Size ($F_s < 1$). A main effect of Block was observed, $F(2, 25) = 22.50$, $p < 0.01$, $n_p^2 = 0.64$, with DTs decreasing over blocks (734, 665, and 534 ms, for Blocks 1-3). There was a Precision \times Block interaction, $F(4, 23) = 5.07$, $p < 0.01$, $n_p^2 = 0.47$, indicating that performance improved across blocks the most when templates were less precise (improvements of 98, 182, and 321 ms, for precise, imprecise, and inaccurate, respectively). No other interactions were significant ($F_s < 2$).

Experiment 1b replicated and extended Experiment 1a. As before, search RTs were slower when targets differed from initial cues, an effect that diminished across blocks. We hypothesized that this interaction might reflect an increased ability to accommodate imperfect target cues over time. However, the eye-tracking data did not support this idea. If participants adopted broader, more tolerant templates, we would expect their decision times to improve across trials, but their scanning to either remain constant or grow worse. In fact, both aspects of search behavior significantly improved with practice. Moreover, we found no interactions of Block with Mismatch Frequency, suggesting that participants did not alter their search templates as a function of cue validity. Indeed, Mismatch Frequency had essentially no effects in any measure (aside from one unsystematic three-way

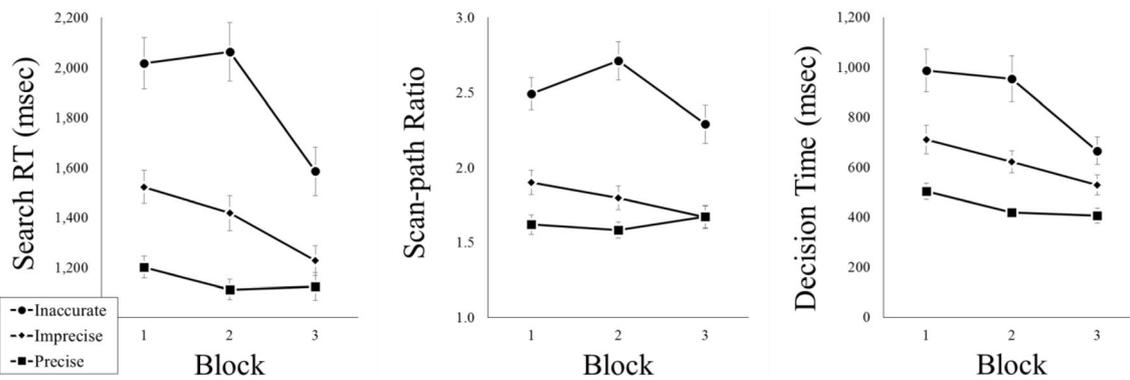


Fig. 5 Mean visual search reaction time, scan-path ratio, and decision time (on accurate trials) from Experiment 1b, presented as a function of Template Precision and Block. Error bars represent 1 standard error (SE) of the mean

interaction). This finding is surprising, and suggests that, even with explicit instructions regarding the trustworthiness of the cues, participants did not alter their approach to the task. Rather, they steadfastly adopted the given cues as their templates and adapted to the challenges imposed by inaccuracy as necessary. (Experiment 1d follows up on this finding by introducing a condition wherein the cues were never reliable indicators of target appearance.)

With respect to the hypotheses outlined previously, the eye-tracking results clearly argue in favor of the dual-function hypothesis. We found clear effects of template precision on both scan-path ratios and decision times, with patterns that mirrored the overall search RTs. The presence of inaccurate template features hinders both the ability to put attention in the correct place, and to identify the target once it is viewed. It is interesting to note that decision time accounted for nearly half of overall RT. In Experiments 1c and 1d, we provide converging evidence for these findings with a different approach to estimating template-to-target similarity. Rather than define similarity based on exemplar and state pairs, we used numerous exemplars of each object category and varied similarity by assessing the distance between items in MDS space. Because these experiments were nearly identical, we present them together.

Experiments 1c and 1d: method

The stimuli and apparatus in Experiments 1c and 1d were identical to Experiments 1a and 1b.

Participants

Data collection for Experiments 1c and 1d were (both) conducted in two phases. During the first phase, participants completed the experiments without eye-tracking. In the second phase, a new group of participants completed the experiments and had their eye movements recorded. There were 30 and 18 participants in phases 1 and 2 of Experiment 1c,

respectively. And there were 60 and 18 participants in phases 1 and 2 of Experiment 1d, respectively. None participated in the prior experiments, nor did any participant complete more than one experiment.

Design

The design of Experiment 1c was identical to Experiment 1a, except for the manner in which template precision was manipulated. Here, there were four levels of precision (precise, similar, moderate, dissimilar). There were four blocks of 60 experimental trials, with 15 trials per precision condition, presented in random order, for a total of 240 trials. Experiment 1d was identical to Experiment 1c, except that the precise condition was removed. There were four blocks of 60 experimental trials, with 20 trials per precision condition.

Procedure

The procedure of Experiment 1c was identical to Experiment 1a, except for the stimulus selection process and inclusion of eye-tracking. As before, *precise* trials involved presenting search targets that were unaltered, relative to their appearance as cues. In other trials, targets deviated slightly from their cues. In contrast to Experiment 1a, which relied on the distinction of state- and exemplar-pairs, similarity was now manipulated by selecting item from MDS spaces obtained in Hout et al. (under review). In each trial, a pair of images was selected from among the 240 image categories for which we acquired MDS data. For every object in every category, its neighboring objects were designated as being close, moderately distant, or far neighbors, based on rank-ordering within the category. When the experimental control program selected a trial type (i.e., *similar*, *moderate*, or *dissimilar*), it randomly selected a cue image and its appropriate target counterpart. Distractors were chosen randomly from the other 239 categories, with no more than one exemplar per semantic category in any trial. Cue-target pairs were used only once per experiment. In Experiment 1d, the *precise* condition was removed;

participants were told that the exact cue objects would never appear in the search display but were instead guides for directing attention to the appropriate target object during search.

Experiments 1c and 1d: results and discussion

For Experiment 1c, all data were analyzed using 4 (Precision: precise, similar, moderate, dissimilar) \times 3 (Set Size: 12, 16, 20) \times 4 (Block: 1-4) within-subjects, repeated measures ANOVAs. For Experiment 1d, data were analyzed in similar fashion but the design was 3 \times 3 \times 4, due to removal of the precise condition. As before, only correct trial RTs, SPRs and DTs were analyzed, and SPRs and DTs were only analyzed for trials in which targets were directly fixated.

In both experiments, we first analyzed the RT data using Phase (1, 2) as a between-subjects factor. In Experiment 1c, there was no main effect of Phase ($F < 3$) and only one interaction including that factor. There was a significant Phase \times Precision interaction ($p < .05$), showing the same qualitative pattern of findings in both phases, but a slightly steeper change in RTs for phase 2, relative to phase 1. For Experiment 1d, there was a main effect of Phase ($p < .05$), wherein RTs were 246 msec faster in phase 1, relative to phase 2, but no interactions with this factor. Therefore, for the sake of brevity, we collapsed all data across Phase in the following analyses.

Response times

Experiment 1c Separate panels of Fig. 6 show mean search RTs, SPRs, and DTs from all conditions of Experiments 1c and 1d. In Experiment 1c, we observed a main effect of Precision, $F(3, 45) = 117.70$, $p < 0.001$, $n_p^2 = 0.89$, with slower RTs as precision decreased (1479, 2023, 2222, and 2306 ms for precise, similar, moderate, and dissimilar conditions, respectively). To assess whether this effect was driven exclusively by the fast RTs to the precise trials, we performed planned-comparisons among other conditions. We found reliable differences between all pairs of conditions (all $p < 0.001$), except the moderate and dissimilar conditions ($p = 0.44$), indicating that this was not the case. There was also a main effect of Set Size, $F(2, 46) = 58.15$, $p < 0.001$, $n_p^2 = 0.72$, with slower RTs as set size increased (1802, 2005, and 2215 ms for 12, 16, and 20, respectively). The main effect of Block was not reliable ($F < 3$), nor were any interactions ($F_s < 3$).

Experiment 1d Again, we found a main effect of Precision, $F(2, 76) = 7.36$, $p < 0.01$, $n_p^2 = 0.16$, with slower RTs as precision decreased (2084, 2143, and 2208 ms for similar, moderate, and dissimilar conditions, respectively). There was a main effect of Set Size, $F(2, 76) = 103.17$, $p < 0.001$, $n_p^2 =$

0.73, with slower RTs at higher set sizes (1945, 2119, and 2370 ms for 12, 16, and 20, respectively). There was no main effect of Block ($F < 3$), nor any interactions ($F_s < 4$).

Scan-path ratios

Experiment 1c There was a main effect of Precision, $F(3, 15) = 68.63$, $p < 0.01$, $n_p^2 = 0.93$, with higher SPRs as precision decreased (2.18, 2.97, 3.35, and 3.50 for precise, similar, moderate, and dissimilar, respectively). Beyond the precise condition, planned comparisons revealed a difference between the similar and dissimilar conditions ($p < 0.05$). There was a main effect of Set Size, $F(2, 16) = 14.46$, $p < 0.01$, $n_p^2 = 0.64$, with higher SPRs at larger set sizes (2.64, 3.03, and 3.23 for 12, 16, and 20, respectively). The main effect of Block was not reliable ($F < 1$), nor were any interactions ($F_s < 1$).

Experiment 1d There was a marginal main effect of Template Precision, $F(2, 15) = 3.57$, $p = 0.05$, $n_p^2 = 0.32$. The numerical trend, however, was not consistent with prior hypotheses, as the moderate condition had the largest SPR (2.97, 3.23, and 3.09 for similar, moderate, and dissimilar, respectively). We found a main effect of Set Size, $F(2, 15) = 23.94$, $p < 0.01$, $n_p^2 = 0.76$, with higher SPRs at larger set sizes (2.76, 3.14, and 3.39 for 12, 16, and 20, respectively). There was no main effect of Block ($F < 2$), nor any interactions ($F_s < 1$).

Decision times

Experiment 1c There was a main effect of Template Precision, $F(3, 15) = 22.49$, $p < 0.01$, $n_p^2 = 0.82$, with longer DTs as precision decreased (345, 644, 736, and 798 ms for precise, similar, moderate, and dissimilar, respectively). Planned-comparisons revealed that this was not due to the precise condition exclusively (there was a significant difference between similar and dissimilar conditions; $p < 0.05$). There was no main effect of Set Size or Block ($F_s < 1$), nor were there any interactions ($F_s < 2$).

Experiment 1d The main effect of Template Precision was not significant ($F < 2$), although the trend was consistent with prior hypotheses (730, 746, and 819 ms for similar, moderate, and dissimilar, respectively). There was no main effect of Set Size ($F < 1$). The main effect of Block was significant, $F(3, 14) = 5.59$, $p < 0.05$, $n_p^2 = 0.55$, with the longest DTs in Block 3 (713, 711, 836, and 802 ms for Blocks 1-4). There was a Set Size \times Block interaction, $F(6, 11) = 6.74$, $p < 0.01$, $n_p^2 = 0.79$, as the Set Size effect was inconsistent across blocks. No other interactions were reliable ($F_s < 2$).

In Experiments 1c and 1d, we again found that imprecise templates hinder search times, now finding that psychologically larger changes (manipulated via distances in MDS

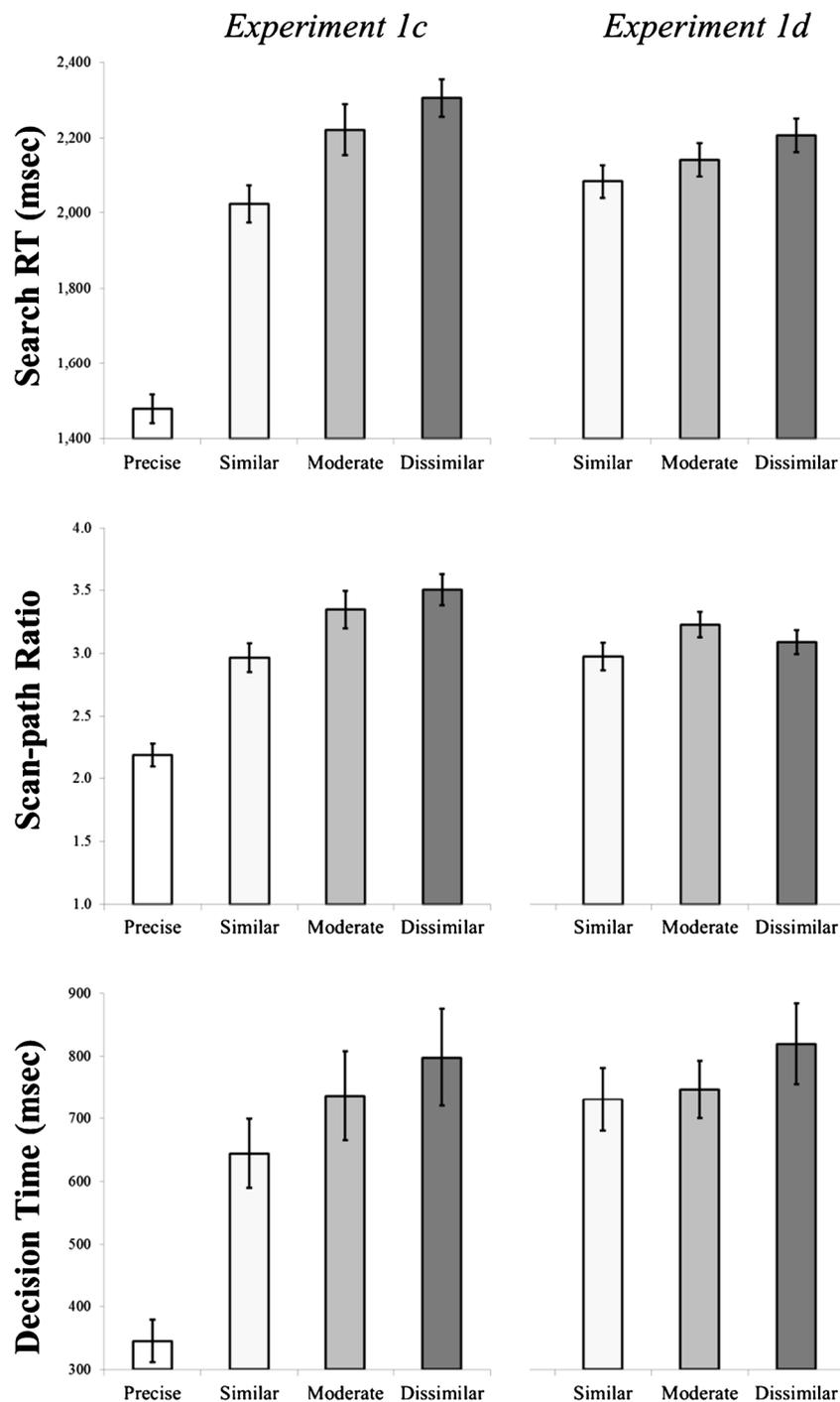


Fig. 6 Mean visual search reaction time, scan-path ratio, and decision time (on accurate trials) from Experiments 1c and 1d, presented as a function of Template Precision. Error bars represent 1 standard error (SE) of the mean

space) have more detrimental effects. In Experiment 1c, search RTs increased as a function of template imprecision and both scan-path ratios and decision times followed suit. These effects were not driven entirely by fast performance in the precise conditions, relative to the others. Rather, it was a graded effect, increasing as dissimilarity between the cues and targets increased. In Experiment 1d, which included no

precise templates, we found consistent trends in the RT and DT data (i.e., slower with greater imprecision), but the latter finding was not reliable (scan-path ratios did not change systematically as a function of template precision).

Together, the results suggest that people hold their search template with great fidelity in VWM and are affected in a continuous fashion by the psychological distance between

expected and encountered targets. The findings from Experiment 1d, in particular, suggest that participants hold the exact cues provided, even when knowing they would never find an exact match. Their RTs still increased monotonically with dissimilarity between cues and targets.

Experiment 2

The key findings from Experiment 1 can be summarized in three points: 1) Template imprecision causes decrements in search RTs that are inversely proportional to the similarity between encoded cues and eventual targets. 2) This finding holds, both when targets are the same exemplars as the cues but in different perceptual states, and when cues and targets are different exemplars altogether. 3) Increased search RTs arise due to deficiencies in both attentional guidance and decision-making, as revealed by eye-tracking.

In Experiment 1, participants (theoretically) formed search templates that mismatched eventual targets. Returning to our earlier example of a high-school reunion, you may have a clear picture of a person in mind but must accommodate mismatching features to recognize the older version of your friend. This is a straightforward situation wherein search templates do not comport very well with actual objects in the environment. In other situations, search templates may not be “wrong,” per se but may contain too many disparate features. As a thought exercise, imagine again that you are attending the reunion and are hoping to see two different old friends. If both friends were men with brown hair, you would be able to visually scan the room and limit consideration to people fitting a fairly narrow candidate set. Alternatively, if one friend was a man with brown hair, and the other was a woman with blonde hair, simultaneously searching for both will allow a far larger candidate set of people to draw your attention. Experiment 2 was focused on this basic contrast, the *feature width* of the search template, asking “how does the inclusion of extraneous template features affect search behavior?”

In Experiment 2, we manipulated template precision by having people search for two targets at once (only one target ever appeared), with varying degrees of mutual similarity. In Experiment 2a, we first tested whether this manipulation (enacted using state versus exemplar pairs) would affect search RTs, then added eye-tracking in Experiment 2b. In Experiment 2c, we again collected search RTs and added eye-tracking measures, with mutual cue similarity estimated using MDS. Data were analyzed in the same manner as Experiment 1, with special focus on main effects and interactions involving feature width. In theoretical terms, we again tested whether having broader search templates would affect attentional guidance, decision making, or both. It seems likely that extraneous features will hinder attentional guidance,

drawing attention to objects that resonate with the “wrong” features in VWM. Additionally, if one assumes that a Sternberg-like comparison process (Sternberg, 1966; 1969; 1975) is completed upon viewing each item, then the similarity of the two potential targets in VWM should have no effect on decision-making times.

Experiment 2a: method

For all the subexperiments under Experiment 2, the apparatus and stimuli were identical to the appropriate counterparts from Experiment 1.

Participants

Twenty-six new students from Arizona State University participated in Experiment 2a.

Design

Three levels of Feature Width (precise, narrow, wider) were manipulated within-subjects. Within conditions, three levels of Set Size (12, 16, 20) were used in equal proportions. There were four practice trials, followed four blocks of 96 experimental trials (384 total). Half the experimental trials presented search cues with only one target object (*precise* trials) and half presented two potential targets. Among the two-target trials, half had *narrow* feature width and half had *wider* feature width.

Procedure

Visual search The procedure for single-target trials was identical to prior experiments. In two-target trials, participants were shown a pair of images at the beginning of each trial and were asked to find one of them. They were informed that only one target would appear per display, so they should respond as soon as they found one. During the cue phase, actual targets were shown equally often on the left and right. Search arrays were configured as before.

Stimulus selection In each trial, a pair of images was selected quasi-randomly from among the 100 exemplar- and 100 state-pair stimuli. In *precise* (i.e., single-target) trials, only one item was randomly selected as the cue and later appeared as the target. In precise trials, stimuli were selected from the exemplar- and state-pair “pools” equally often. The *narrow* and *wider* conditions had two potential targets per trial. For narrow trials, paired images were selected from the state-pair stimuli; in wide trials, paired images were selected from the exemplar-pair stimuli. Figure 7 shows example target cues (note that the *widest* condition was not included in



Fig. 7 Sample single- and two-target visual search cues, from Experiment 2. Participants saw one or a pair of images and were instructed to find one of them. In the *precise* Feature Space condition, only a single item was shown as the cue. For the *narrow* condition a state-pair was used, and for the *wider* condition an exemplar-pair was selected. In the *widest* condition, two images were selected from different semantic categories

Experiment 2a but was added to later experiments). As in prior experiments, distractors were chosen quasi-randomly from 240 picture categories; only one exemplar per category was presented in any given trial. Neither target nor distractor stimuli were used more than twice throughout the entire experiment.

Experiment 2a: results and discussion

The data were analyzed using a 3 (Feature Width: precise, narrow, wider) x 3 (Set Size: 12, 16, 20) x Block (1-4) within-subjects, repeated measures ANOVA. Only correct trial RTs were analyzed and are shown in Fig. 8. One participant was excluded from data analysis for having mean accuracy and RTs that were more than 2.5 standard deviations below the

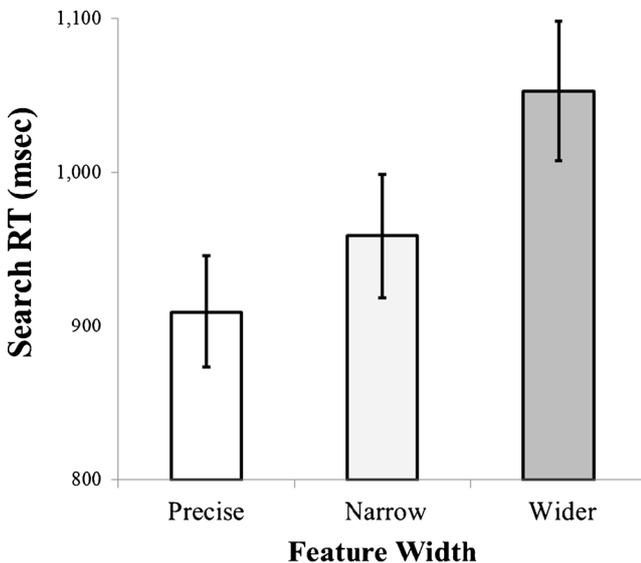


Fig. 8 Mean visual search reaction time (on accurate trials) from Experiment 2a, presented as a function of Feature Width. Error bars represent 1 standard error (SE) of the mean

group mean. There was a main effect of Feature Width, $F(2, 23) = 18.89, p < 0.01, n^2_p = 0.62$, with slower RTs as the feature space widened (909, 958, and 1053 ms for precise, narrow, and wider, respectively). We also observed main effects of Set Size, $F(2, 23) = 82.63, p < 0.01, n^2_p = 0.88$, and Block, $F(3, 22) = 4.00, p < 0.05, n^2_p = 0.35$. RTs lengthened as set size increased (855, 975, and 1090 ms for 12, 16, and 20, respectively) and shortened over blocks (1037, 982, 958, and 917 ms for Blocks 1-4). There were no reliable interactions ($F_s < 2$).

Experiment 2a showed that increasing template feature width slowed visual search. In Experiment 2b, we examined attentional guidance and decision making by adding eye-tracking. We also added a fourth condition to the feature width manipulation (shown in Fig. 7). In the wider condition of Experiment 2a, people searched for two different exemplars of the same core object. In the *widest* condition of Experiment 2b, people search for two different underlying objects simultaneously.

Experiment 2b: method

Participants

Sixteen new students from Arizona State University participated in Experiment 2b.

Design

The design was identical to Experiment 2a, with two exceptions. First, the *widest* condition was added, wherein two images were shown as cues prior to search, showing objects from two semantic categories. Second, there were now only three blocks of 72 trials (216 total).

Procedure

The procedure was identical to Experiment 2a, except that eye-tracking procedures were now added to the experiment. In the *widest* condition trials, two images were selected quasi-randomly, used as search cues, with one item randomly selected to be the actual target. Stimuli for this condition were drawn equally from the state- and exemplar-pairs. Distractors were quasi-randomly selected from 240 item categories, and no image was used more than twice in the entire experiment.

Experiment 2b: results and discussion

Data were analyzed using a 4 (Feature Width: precise, narrow, wider, widest) x 3 (Set Size: 12, 16, 20) x 3 (Block: 1-3)

within-subjects, repeated measures ANOVA. Only correct trial RTs, SPRs, and DTs were analyzed; SPRs and DTs were only analyzed from trials in which targets were directly fixated.

Response times

Figure 9 shows all search RTs, SPRs, and DTs from Experiment 2b. In RTs, we found a main effect of Feature Width, $F(3, 13) = 51.76, p < 0.01, n^2_p = 0.92$, with longer RTs as feature space widened (958, 1016, 1119, and 1563 ms for precise, narrow, wider, and widest, respectively). As before, we conducted planned comparisons among conditions, finding a difference between the precise and wider conditions ($p < 0.05$). There were also main effects of Set Size, $F(2, 14) = 46.32, p < 0.01, n^2_p = 0.87$, and Block, $F(2, 14) = 3.90, p < 0.05, n^2_p = 0.36$: RTs increased with set size (1009, 1154, and 1329 ms for 12, 16, and 20, respectively) and were unequal across blocks (1215, 1134, and 1143 ms for Blocks 1-3). There also was a Feature Width \times Set Size interaction, $F(6, 10) = 4.64, p < 0.05, n^2_p = 0.74$, indicating that increasing set size had a larger effect when people's feature space was wider. We calculated slopes of the best-fitting lines relating RT to Set Size; this indicates the "cost" associated with adding each new item to the display. Slopes were 15, 29, 44, and 72 ms/item for the precise, narrow, wider, and widest conditions, respectively.

Scan-path ratios

There was a main effect of Feature Width, $F(3, 13) = 68.67, p < 0.01, n^2_p = 0.94$, with larger SPRs as feature space widened (1.39, 1.46, 1.62, and 2.24 for precise, narrow, wider, and widest, respectively). Planned-comparisons showed only a marginal difference between the precise

and wider conditions ($p = 0.07$). We also found main effects of Set Size, $F(2, 14) = 30.63, p < 0.01, n^2_p = 0.81$, and Block, $F(2, 14) = 8.28, p < 0.01, n^2_p = 0.54$. SPRs grew as set size increased (1.46, 1.66, and 1.91 for 12, 16, and 20, respectively) and were inconsistent across blocks (1.76, 1.60, and 1.67 for Blocks 1-3). There were no interactions ($F_s < 2$).

Decision times

We found a main effect of Feature Width, $F(3, 13) = 11.78, p < 0.01, n^2_p = 0.73$, with longer DTs as feature space increased (362, 415, 413, and 521 ms for precise, narrow, wider, and widest, respectively). Planned comparisons revealed that this effect was driven exclusively by the difference between the widest condition, relative to all other conditions; no other pairwise comparisons were reliable. There was a main effect of Block, $F(2, 14) = 10.82, p < 0.01, n^2_p = 0.61$, indicating that decisions became faster over the course of the experiment (478, 414, and 391 ms for Blocks 1-3). There was effect of Set Size ($F < 2$), but there was a Feature Width \times Set Size \times Block interaction, $F(12, 4) = 27.72, p < 0.01, n^2_p = 0.99$. No other interactions were reliable ($F_s < 2$).

Experiment 2b showed that search is hindered by target templates that were unnecessarily broad. The eye-tracking measures revealed that wider templates affect both attentional guidance and decision-making. In Experiment 2c, we provide converging evidence by manipulating the template widths using MDS-derived item pairs. In this experiment, participants always searched for two potential targets from the same semantic category, but the psychological distance between those exemplars was manipulated. We again compared these conditions to a single-item (precise template) control condition.

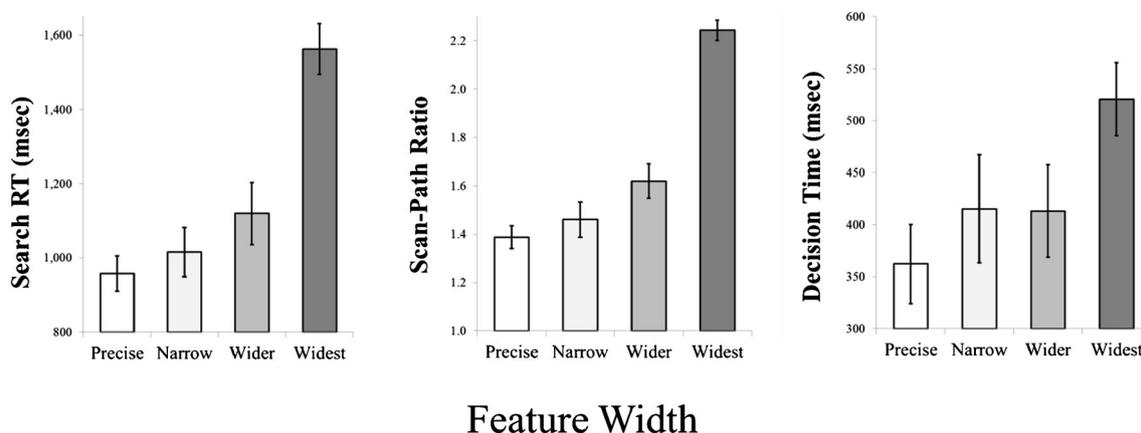


Fig. 9 Mean visual search reaction time, scan-path ratio, and decision time (on accurate trials) from Experiment 2b, presented as a function of Feature Width. Error bars represent 1 standard error (SE) of the mean

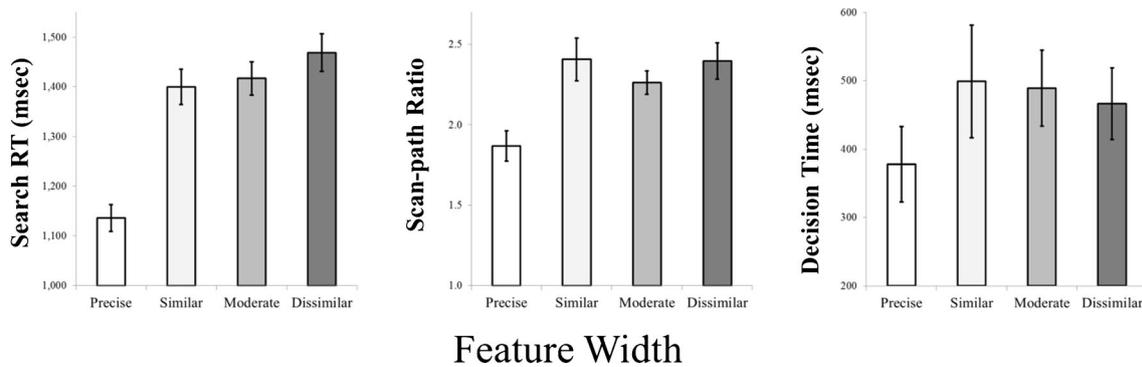


Fig. 10 Mean visual search reaction time, scan-path ratio, and decision time (on accurate trials) from Experiment 2c, presented as a function of Feature Width. Error bars represent 1 standard error (SE) of the mean

Experiment 2c: method

Participants

As in Experiments 1c and 1d, data collection in Experiment 2c was conducted in two phases (without and with the inclusion of eye-tracking, respectively). Sixty-five and 17 new students from Arizona State University participated in phase 1 and 2, respectively.

Design and procedure

The design of Experiment 2c was identical to Experiment 2a, with two exceptions (plus the inclusion of eye-tracking in phase two). First, feature width was manipulated using the MDS spaces from Hout et Al. (under review), with four levels: *precise*, *similar*, *moderate*, and *dissimilar*. Second, there were 4 blocks of 80 trials (320 total), with 20 trials per feature width condition. As before, *precise* trials involved only one image, used as the cue and target. In other trials, two target cues were shown, selected from the MDS spaces, with inter-item distances that had been classified as *similar*, *moderate* and *dissimilar*.

Experiment 2c: results and discussion

The data were analyzed using a 4 (Feature Width: precise, similar, moderate, dissimilar) x 3 (Set Size: 12, 16, 20) x 4 (Block: 1-4) within-subjects, repeated measures ANOVA. Only correct trial RTs, SPRs, and DTs were analyzed, and SPRs and DTs were only analyzed for trials in which targets were directly fixated. Two participants were excluded from analysis for error rates more than 2.5 standard deviations above the group mean.

As before, we first analyzed the RT data using Phase as a between-subjects factor. The main effect of Phase was not significant ($p = 0.11$). There was a Phase x Set Size interaction ($p < 0.05$), showing a steeper increase in RTs as a function of

set size for phase 2, relative to phase 1. There also was an unsystematic 5-way interaction of all factors ($p < 0.05$) but no other interactions with Phase. We therefore collapsed all data across Phase in the following analyses.

Response times

We found a main effect of Feature Width (see Fig. 10), $F(3, 79) = 99.39$, $p < 0.001$, $n_p^2 = 0.79$, with slower RTs as feature spaces widened (1136, 1400, 1417, and 1469 ms for precise, similar, moderate, and dissimilar, respectively; Fig. 1). Planned comparisons showed a reliable difference between similar and dissimilar conditions ($p < 0.01$). There also were main effects of Set Size, $F(2, 80) = 177.99$, $p < 0.001$, $n_p^2 = 0.82$, and Block, $F(3, 79) = 4.28$, $p < 0.01$, $n_p^2 = 0.14$. RTs increased with set size (1202, 1361, and 1503 ms for 12, 16, and 20, respectively), and decreased over blocks (1444, 1347, 1315, and 1317 ms for Blocks 1-4). No other interactions were reliable ($F_s < 2$).

Scan-path ratios

We found a main effect of Feature Width, $F(3, 14) = 8.31$, $p < 0.01$, $n_p^2 = 0.64$, with larger SPRs for dual-target trials, relative to single-target trials (1.87, 2.41, 2.26, and 2.40 for precise, similar, moderate, and dissimilar conditions, respectively). Planned-comparisons showed that the main effect was driven by the difference between single- and dual-target trials (none of the latter different from one another). There was a main effect of Set Size, $F(2, 15) = 12.18$, $p < 0.01$, $n_p^2 = 0.62$, with larger SPRs at higher set sizes (1.99, 2.20, and 2.51 for 12, 16, and 20, respectively). The main effect of Block was not reliable ($F < 1$), nor were any interactions (all $F_s < 3$).

Decision times

There was a main effect of Feature Width, $F(3, 14) = 9.81$, $p < 0.01$, $n_p^2 = 0.68$, again indicating a difference between single- and dual-target trials (378, 499, 489, and 467 ms for

precise, similar, moderate, and dissimilar conditions, respectively). Planned-comparisons showed that none of the dual-target conditions differed from one another. There was a main effect of Set Size, $F(2, 15) = 3.84$, $p < 0.05$, $\eta_p^2 = 0.34$, with longer DTs at higher set sizes (393, 452, and 530 ms for 12, 16, and 20, respectively). The main effect of Block was not reliable ($F < 3$), nor were any interactions (all F s < 2).

Experiment 2c showed converging evidence that broader template features inhibit the ability to quickly perform visual search. Participants always searched for a single semantic category, but template width was increased monotonically with distances in psychological space. Importantly, for RTs, the feature width effect was not driven exclusively by the contrast of single- and dual-target trials. Rather, there was a graded effect within the dual-target conditions alone. With respect to eye-movement measures, we found that the feature width effect was driven by the contrast of the precise condition versus every other condition.

Experiment 2 can be summarized in three main points: 1) Template imprecision, imposed by widening the “feature space” of search templates, slows search RTs monotonically with the dissimilarity of the potential target images. 2) Clear performance changes arise between single-target and dual-target search conditions, and (within dual-target trials) between conditions in which the targets are semantically matched, relative to when they are not. And 3) the eye-movement data showed that slower search RTs are caused by decrements in both attentional guidance and decision-making.

General discussion

The present results are consistent with a large literature, which shows that the contents of VWM bias attention toward target-defining features (Olivers et al., 2011; Woodman et al., 2007; Dowd & Mitroff, 2013; Soto et al., 2008). For instance, it has been shown that search templates work at a functionally “low” level, acting in the service of gaze correction. Hollingworth et al. (2008); Hollingworth and Luck (2009) found that small, corrective saccades tend to be directed toward features that match the search template, typically outside of the observer’s conscious awareness. People make thousands of saccades each day, but these eye movements are ballistic in nature and therefore are highly prone to error. When the eyes miss their intended locations, people make rapid, corrective saccades toward the intended locations or objects (Becker, 1972; Deubel, Wolfe, & Hauske, 1982). These saccades are initiated quickly, almost reflexively (Kapoula & Robinson, 1986).

Using gaze-contingent displays, Hollingworth and colleagues induced saccadic errors: Participants fixated a central cross, and were shown a circular array of different colored patches. One of the color patches changed in size, signaling to

the observer to fixate that item. On some trials, after the saccade was initiated, the circular array was rotated. Participants did not notice the rotation, due to saccadic suppression (Ethel, 1974; Thiele et al., 2002), but this procedure artificially created saccadic error because the targets moved while the eyes were “in flight.” The arrays were rotated only a little, causing participants to fixate the middle region between the intended target and a distractor. At this point, the task becomes a small-scale visual search, wherein the searcher must make a corrective eye movement to the target and avoid the distractor. The results demonstrated that gaze-correction was fast and accurate: Participants correctly moved their eyes to the target more than 90 % of the time, typically in under 200 ms, with no awareness of making two distinct saccades (Hollingworth, Richard, & Luck, 2008). This suggests that the search template—in this instance, a target color—biased attention to allow the visual system to quickly inspect regions of interest.

At the other end of the complexity spectrum, search templates also are used to direct attention in visually rich environments. Malcolm and Henderson (2010) showed people pictures of real-world scenes and asked them to locate targets that were either shown as picture cues or verbal cues. Specific, picture-cued templates allowed observers to more efficiently place their attention at target locations (as indexed by scan-path ratios) and to verify more quickly target identity once attention was situated appropriately (indexed by postfixation RTs). These findings are entirely consistent with the present results.

Our data also are consistent with a smaller literature that shows that imperfect target cues slow visual search. For example, Vickery, King, and Jiang (2005) conducted experiments wherein people searched for either polygon shapes or real-world objects. In some trials, the target cues were imperfect representations of the to-be-located items. Specifically, they manipulated the size and orientation of targets, relative to cues, and found that deviations along either dimension increased search time. Nevertheless, these imperfect pictorial cues still produced faster search RTs than verbal cues, suggesting that detailed visual information improves attentional guidance. Similarly, Bravo and Farid (2009) had people search for tropical fish in scenes of coral reefs. Before conducting the search task, the participants were trained. They were shown single exemplars from five different species and learned to associate the species names with those particular fish. The search task involved determining whether a fish of any species was present in the picture. Across experiments, pictorial and verbal cues indicated which fish was likely to appear. Importantly, there were three conditions: *no variation* (targets were identical to studied images), *2D viewpoint variation* (targets were rotated, flipped, and scaled, relative to studied images), and *subordinate level variation* (targets were different images of the same species). Given picture

cues, participants found the targets most quickly without variation, slower when given 2D viewpoint variations, and slowest when given new exemplars. By contrast, when verbal cues were used, search times were equivalent when the target was unvaried or 2D transformed. However, search was significantly slower when a novel exemplar was shown. This pattern strongly suggests that participants create search templates that are detailed and specific, yet tolerant to deviation.

The present investigation built upon these prior findings by controlling degrees of template imprecision and by tracking eye movements during search. We asked whether imprecise templates might affect attentional guidance to objects on-screen, whether they might affect target detection after fixation, or both. Our results strongly supported the dual-function hypothesis: by decomposing search behavior into two functionally distinct phases—scanning time and decision time—we consistently found that the inclusion of inaccurate template features affected performance in both attentional guidance and decision-making. The present results extend the prior literature because of the way template-to-target similarity was manipulated. Vickery and colleagues (2005) used simple visual variations of targets, relative to cues; Bravo and Farid (2009) involved new exemplars of the target categories. We aimed to add more precision, classifying cross-exemplar similarity using MDS, finding graded decrements in performance that were proportional to the psychological dissimilarity of the items.

Can templates be flexibly controlled?

The utility of precise target templates is well-known. Specific information about the likely appearance of a target can help to guide attention and can help a person fluently verify whether visual input matches whatever is sought. As noted by Bravo and Farid (2012), an effective template must do two things: it must differentiate the target from potential distractors, and it must tolerate variability in target appearance. After all, exemplars from real-world categories vary widely in appearance, are viewed from odd vantage points, etc. An interesting theoretical question, however, is whether people can control the precision (or flexibility) of their search templates.

Indeed, it seems that people can willfully control the precision with which information is held in VWM. Machizawa et al. (2012) used an orientation discrimination paradigm, wherein people were prompted to anticipate a fine- or coarse-level discrimination following a delay interval. Participants were shown a sample display showing several lines, oriented at different angles. The display was then removed for more than 1 s and was replaced by a test probe that looked exactly like the sample display, but one bar changed color and was rotated 15 or 45 degrees (fine and coarse discriminations, respectively). The task was to indicate

whether the bar had rotated clockwise or counterclockwise. Unbeknownst to participants, there was an intermediate condition, wherein the bar was rotated 30 degrees. In this condition, discrimination performance was improved when participants were first prompted to anticipate fine-level discrimination, relative to trials in which they were prompted to anticipate coarse-level discrimination. This suggests that participants used the prompt to adjust the precision with which information was stored in VWM.

In a related study, Bravo and Farid (2012) examined the extent to which people could hone their target templates to meet variable task demands. Participants again looked for tropical fish in underwater coral reef scenes. There were two groups: half of the participants searched for the same target image over and over again, and the other half searched for multiple exemplars of the same species. Everyone came back 1–2 days after the initial session and completed a second session that included new exemplars from the target category. Search times to find these new images were faster for participants who were trained on multiple exemplars, relative to those who were trained on a single image, suggesting more varied training allowed people to adopt more general templates that tolerate variation.

In Experiment 1b of the present investigation, participants were told how often targets would appear exactly as cued. Our hypothesis was that participants who received frequently accurate cues would adopt templates that closely matched that the cues, and those who received frequently inaccurate cues would adopt templates that were more tolerant to variation from the cues. Our findings, however, contradicted this prediction: instructions had no substantive effect on RTs or eye movements. In Experiment 1c, we included a condition in which the cues *never* accurately depicted the targets. We anticipated that participants would adopt general templates that represented categories but were less tethered to the cues. Nevertheless, people still appeared to unduly rely on the presented cues: performance generally decreased as cue-target similarity decreased.

Given our manipulations, we anticipated that participants would construct templates that resembled category prototypes, rather than the cues themselves. This did not occur, which may suggest that people cannot behave optimally. However, upon closer consideration, it seems that our participants simply chose to adopt the presented cues as their templates, because that was the easiest thing to do. Stated plainly, calling to mind prototypical templates might be easy when given verbal cues, but—given visual cues—the easiest strategy is to simply encode it without transformation. This may be a case of availability, such that discounting a visual cue is nearly impossible because it was just seen. We expect that, in conditions with less salient visual cues, we might find that participants are better able to adjust the fidelity of their search templates in accordance with optimal strategies.

Cost of searching for multiple items

People are adept visual searchers. Under most circumstances, we can search for more than one item at a time, seemingly without effort. Before departing for work, you may search for several things simultaneously (e.g., your wallet, keys, and phone) and will happily collect them in any order that they appear. Despite its subjective ease, multiple-target search incurs clear costs, relative to single-target search. In prior research (Hout & Goldinger, 2010), we had people search for one, two, or three potential targets, either in cluttered search displays or RSVP “streams” (wherein single items were centrally presented in rapid succession). Only one target appeared in any given trial and participants were informed of this regularity. We found that multiple-target search affected both accuracy and speed; participants made more misses and false-alarms during multiple-target search and were slower in accurate trials, relative to single-target search. These costs also are revealed in eye movements: when people look for multiple targets, they require more fixations to complete each trial (Hout & Goldinger, 2012).

Arguably, in most situations, it likely feels more natural to conduct one multiple-target search than to conduct several consecutive searches for individual items. For example, if search involves walking around your house, it would be inefficient to sweep the entire home once for your wallet, then again for your keys. In situations where search is confined to a single display, however, intuition and performance do not align so well. Menneer et al. (2007; also Menneer et al., 2009; 2010) compared performance in conditions wherein people either looked for two targets simultaneously or looked for two single targets back-to-back. Despite intuition, dual-target search took longer (and was less accurate) than two consecutive single-target searches. Menneer and colleagues’ work suggests that, when people look for multiple items at once, the fidelity of the target representations cannot be faithfully maintained. Thus, high-stakes searchers (such as airport baggage screeners) may be better served by a divided search strategy. The findings from Experiment 2 clearly showed dual-target costs in visual search, with worse performance when templates held two potential targets, relative to singular (definite) targets.

An open question remains, however, regarding the nature of VWM templates during multiple-target search: do people use separate, discrete representations, or do they merge cues into a single, broad target template? The answer seemingly depends on the task at hand. VWM has limited capacity (Cowan, 2001; Vogel et al., 2005), and different theoretical accounts have addressed its potential organization. Some theories propose that VWM contains limited “slots” in which information can be stored (Anderson et al., 2011; Awh et al., 2007). Others envision VWM as a dynamic resource that is limited by overall precision, rather than number (Bays &

Husain, 2008; Gorgoraptis et al., 2011). Still other theories suggest an answer somewhere in between, as some kind of hybrid discrete-slot/dynamic-resource organization (Alvarez & Cavanagh, 2004; Buschman et al., 2011).

A recent study by Stroud et al. (2011; also Godwin et al., 2010) investigated this question using single- and dual-target search for letter stimuli. People looked for *T*s of a certain color; distractors were *L*s of various colors. Stroud et al. entertained two hypotheses regarding the nature of dual-target templates. First, searchers might maintain two target templates simultaneously (or rapidly alternate between them; Moore & Osman, 1993). Alternatively, they might construct single-target templates that include features for both targets, and possibly values in between (i.e., colors that occupy the “color space” between the targets). Stroud et al. systematically manipulated the similarity of the two potential targets by varying how far apart the items were in “color-steps” (defined using CIE_xyY color space). When people searched for a single target, they exhibited impressive color selectivity, rarely fixating items that did not match the target color. When looking for two targets, however, color selectivity was reduced, as people often fixated colors that did not match either target. However, the data were more nuanced: when the target colors were two steps apart, people fixated the intervening color more often than when the targets were four steps apart. This suggests that, when the targets were similar, they were encoded as unitary representations that meshed together both target colors, and those in between. When the targets are dissimilar, however, they were encoded as separate and discrete representations that did not “absorb” the feature space between them (although see Houtkamp & Roelfsema, 2009).

Following the findings of Stroud et al. (2011), we may be inclined to suggest that when people search for two similar targets (as state pairs), they are represented by meshed, unitary representations, but dissimilar targets (such as exemplar pairs) involve detached representations. In Experiment 2b, the scan-path ratios showed that attentional guidance was worse when people switched from single-target (precise) templates to dual-target (but single-category) templates, and then grew even worse for two-category templates. In Experiment 2c, however, we found an effect of feature width that only arose between single- and dual-target search; none of the dual-target conditions (that varied in similarity defined by MDS space) differed from one another. The decision-time data showed similar results.

In short, we did not observe fine-grained effects of feature width on either guidance or decision-making. However, adding extraneous features to the search template caused problems for both aspects of search performance. It seems likely that, when two potential targets are very similar (as in our narrow feature width conditions), they may be fused into a singular templates, and that less similar targets may be represented by discrete templates.

Conclusions and future directions

The current experiments investigated several important aspects of search templates in VWM, but many theoretical questions remain. In Experiment 1, we provided evidence that, when inaccurate features are included in search templates, both attentional guidance and decision-making are impaired. A question that remains, however, regards the extent to which people have flexible control over the nature of their templates. We attempted to address this issue by manipulating the trustworthiness of the cues and by providing cues that were never fully accurate. Our results, however, suggested that participants still adopted the cues as their templates, perhaps because ignoring such available visual information is impossible. In future experiments, it would be interesting to investigate the extent to which people can create broad, categorical templates, less anchored to the provided search cues. For instance, if participants know that cues are categorically accurate, but dissimilar to the actual targets, can they learn to emphasize category knowledge and deemphasize specific cue features? If so, it would suggest that memory-derived templates are under the flexible control of the observer and can be honed with a high degree of contextual sensitivity.

In Experiment 2, we found further evidence that imprecise templates hinder both search guidance and decision-making. Taken together, these experiments provide strong support for a dual-function theory of target templates, wherein they help select appropriate objects to sample in the environment and help the memory-comparison process that determines when search has been achieved successfully. An unresolved issue, however, concerns the VWM representations in dual-target search. When people look for two things at once, do they operate using a single, meshed template, or two discrete representations? We attempted to answer this question by evaluating eye-movements between single- and dual-target searches, and (in dual-target search) between single- and dual-category search. Our findings suggest that it may be possible, under certain circumstances, to mesh two similar targets into a singular representation, and that with increasing dissimilarity between the targets, they must be maintained separately. At present, the inconsistencies in our eye-tracking measures are not enough to make this claim forcefully. Further evidence is required to support this contention.

Future experiments should be designed to further delineating this issue, perhaps by modeling the decision-making process in analogous fashion to Sternberg memory scanning (Sternberg, 1966). If we stretch out search to three, four, or five potential targets, and if we vary item similarity within those sets, there may be sufficient leverage to model decision-times (and fixation durations), helping to determine whether separate templates are maintained in tandem. Imagine, for instance, that you try to search for three simultaneous targets. These could be three categorically different targets (e.g., a

mug, a lamp, and a dog), or a combination wherein two targets are alike (e.g., a mug and two similar lamps, or a mug and two dissimilar lamps). By contrasting search performance across such conditions, we may better understand when target templates can be collapsed and when they must be maintained separately. For example, searching for one mug and two similar lamps might resemble searching for one mug and one lamp, suggesting that the searcher collapsed both lamps into a single representation. This again would suggest that people have flexible control over search templates and can constrain the template specifics to match changing task demands.

Clearly, it is beneficial to have target templates that faithfully represent the actual appearance of sought-after items out in the world. But everyday search rarely affords the precision provided in laboratory experiments, and people typically search using templates that contain inaccurate features, too many features, or too few. To the extent that search templates deviate from the exact forms of targets, people suffer decrements both in guiding attention to viable candidate objects in space, and in ascertaining whether visual input matches the desired item, such that attention has found its way to the right place. The present approach showed that search templates are surprisingly detailed, such that people are highly sensitive to even small mismatches between expected and discovered targets. Future research will seek to better characterize exactly how objects are represented in VWM when people perform the common task of looking for more than one thing at a time, a topic of critical theoretical and practical importance.

Acknowledgments This work was supported by NIH grant 1 R01 HD075800-01 to Stephen D. Goldinger. We thank Kyle J. Brady for assistance in multidimensional scaling data analysis, and Alexi Rentzis, Lindsey Edgerton, Shelby Doyle, Taylor Thorn, Christina Molitor, Sarah Fialko, Mandana Minai, Deanna Masci, and Taylor Coopman for assistance in data collection. We also thank Carrick Williams and an anonymous reviewer for helpful comments on an earlier version of this manuscript.

References

- Alexander, R. G., & Zelinsky, G. J. (2011). Visual similarity effects in categorical search. *Journal of Vision*, *11*, 1–15. doi:10.1167/11.8.9
- Al-Aidroos, N., Emrich, S. M., Ferber, S., & Pratt, J. (2012). Visual working memory supports the inhibition of previously processed information: Evidence from preview search. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 643–663. doi:10.1037/a0025707
- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106–111. doi:10.1111/j.0963-7214.2004.01502006.x
- Anderson, D. E., Vogel, E. K., & Awh, E. (2011). Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *Journal of Neuroscience*, *31*, 1128–1138. doi:10.1523/JNEUROSCI.4125-10.2011

- Arita, J. T., Carlisle, N. B., & Woodman, G. F. (2012). Templates for rejection: Configuring attention to ignore task-relevant features. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 580–584. doi:10.1037/a0027885
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*, 622–628. doi:10.1111/j.1467-9280.2007.01949.x
- Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854. doi:10.1126/science.1158023
- Becker, S. I. (2011). Determinants of dwell time in visual search: Similarity or perceptual difficulty? *PLoS One*, *6*, 1–5. doi:10.1371/journal.pone.0017740
- Becker, W. (1972). The control of eye movements in the saccadic system. *Bibliotheca Ophthalmologica*, *82*, 233–243.
- Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*, 529–534. doi:10.1126/science.1109676
- Bond, A. B. (1983). Visual search and selection of natural stimuli in the pigeon: The attention threshold hypothesis. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*, 292–306. doi:10.1037/0097-7403.9.3.292
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*, 14325–14329. doi:10.1073/pnas.0803390105
- Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision*, *9*, 1–9. doi:10.1167/9.1.34
- Bravo, M. J., & Farid, H. (2012). Task demands determine the specificity of the search template. *Attention, Perception & Psychophysics*, *74*, 124–131. doi:10.3758/s13414-011-0224-5
- Buschman, T. J., Siegel, M., Roy, J. E., & Miller, E. K. (2011). Neural substrates of cognitive capacity limitations. *Proceedings of the National Academy of Sciences*, *108*, 11252–11255. doi:10.1073/pnas.1104666108
- Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 753–763. doi:10.1037/0096-1523.33.4.753
- Castelhano, M. S., Pollatsek, A., & Cave, K. (2008). Typicality aids search for an unspecified target, but only in identification, and not in attentional guidance. *Psychonomic Bulletin & Review*, *15*, 795–801. doi:10.3758/PBR.15.4.795
- Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*, 2918–2940.
- Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*, 345–347. doi:10.1038/363345a0
- Chen, X., & Zelinsky, G. J. (2006). Real-world search is dominated by top-down guidance. *Vision Research*, *46*, 4118–4133. Real-world search is dominated by.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. doi:10.1146/annurev.ne.18.030195.001205
- Deubel, H., Wolf, W., & Hauske, G. (1982). Corrective saccades: Effect of shifting the saccade goal. *Vision Research*, *22*, 353–364. doi:10.1016/0042-6989(82)90151-1
- Dowd, E. W., & Mitroff, S. R. (2013). Attentional guidance by working memory overrides saliency cues in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 1786–1796. doi:10.1037/a0032548
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458. doi:10.1037/0033-295X.96.3.433
- Duncan, J., & Humphreys, G. W. (1992). Beyond the search surface: Visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 578–588. doi:10.1037/0096-1523.18.2.578
- Eckstein, M. P., Beutter, B. R., Pham, B. T., Shimozaki, S. S., & Stone, L. S. (2007). Similar neural representations of the target for saccades and perception during search. *Neuron*, *27*, 1266–1270. doi:10.1523/JNEUROSCI.3975-06.2007
- Eimer, M., Kiss, M., & Nicholas, S. (2011). What top-down task sets do for us: An ERP study on the benefits of advance preparation in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 1758–1766. doi:10.1037/a0024326
- Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*, 1–19. doi:10.1167/8.2.2
- Ethel, M. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*, 899–917. doi:10.1037/h0037368
- Evans, K. K., Horowitz, T. S., Howe, P., Pedersini, R., Reijnen, E., Pinto, Y., ... & Wolfe, J. M. (2011). Visual attention. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 503–514. doi: 10.1002/wcs.127
- Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, *37*, 617–631. doi:10.1016/S0042-6989(96)00218-0
- Frings, C., Wentura, D., & Wühr, P. (2012). On the fate of distractor representations. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 570–575. doi:10.1037/a0027781
- Godwin, H. J., Hout, M. C., & Menneer, T. (2014). Visual similarity is stronger than semantic similarity in guiding visual search for numbers. *Psychonomic Bulletin & Review*, *21*, 689–695. doi:10.3758/s13423-013-0547-4
- Godwin, H. J., Menneer, T., Cave, K. R., & Donnelly, N. (2010). Dual-target search for high and low prevalence X-ray threat targets. *Visual Cognition*, *18*, 1439–1463. doi:10.1080/13506285.2010.500605
- Goldstone, R. L., & Medin, D. L. (1994). The time course of comparison. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 29–50. doi:10.1037/0278-7393.20.1.29
- Gorgoraptis, N., Catalao, R. F., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *Journal of Neuroscience*, *31*, 8502–8511. doi:10.1523/JNEUROSCI.0208-11.2011
- Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford, UK: Elsevier.
- Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance versus visual salience during search for non-salient objects in real-world scenes. *Psychonomic Bulletin & Review*, *16*, 850–856. doi:10.3758/PBR.16.5.850
- Hollingworth, A., & Luck, S. J. (2009). The role of visual working memory (VWM) in the control of gaze during visual search. *Attention, Perception & Psychophysics*, *71*, 936–949. doi:10.3758/APP.71.4.936
- Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General*, *137*, 163–181. doi:10.1037/0096-3445.137.1.163
- Hon, N., Thompson, R., Sigala, N., & Duncan, J. (2009). Evidence for long-range feedback in target detection: Detection of semantic targets modulates activity in early visual areas. *Neuropsychologia*, *47*, 1721–1727. doi:10.1016/j.neuropsychologia.2009.02.011
- Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception & Psychophysics*, *72*, 1267–1282. doi:10.3758/APP.72.5.1267

- Hout, M. C., & Goldinger, S. D. (2012). Incidental learning speeds visual search by lowering response thresholds, not by improving efficiency. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 90–112. doi:10.1037/a0023894
- Hout, M. C., Goldinger, S. D., & Brady, K. J. (under review). MM-MDS: A multidimensional scaling database with similarity ratings for 240 object categories from the Massive Memory picture database.
- Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, *142*, 256–281. doi:10.1037/a0028860
- Houtkamp, R., & Roelfsema, P. R. (2009). Matching of visual input to only one item at any one time. *Psychological Research*, *73*, 317–326. doi:10.1007/s00426-008-0157-3
- Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, *9*, 1–18. doi:10.1167/9.5.25
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506. A saliency-based search mechanism for overt and covert shifts of visual attention.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, *2*, 194–203. doi:10.1038/35058500
- Kapoula, Z., & Robinson, D. A. (1986). Saccadic undershoot is not inevitable: Saccades can be accurate. *Vision Research*, *26*, 735–743. doi:10.1016/0042-6989(86)90087-8
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227. doi:10.1007/978-94-009-3833-5_5
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*, 558–578. doi:10.1037/a0019165
- Kunar, M. A., Flusberg, S., & Wolfe, J. M. (2008). The role of memory and restricted context in repeated visual search. *Perception & Psychophysics*, *70*, 314–328. doi:10.3758/PP.70.2.314
- Machizawa, M. G., Goh, C. C. W., & Driver, J. (2012). Human visual short-term memory precision can be varied at will when the number of retained items is low. *Psychological Science*, *23*, 554–559. doi:10.1177/0956797611431988
- Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9*, 1–13. doi:10.1167/9.11.8
- Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, *10*, 1–11. doi:10.1167/10.2.4
- Mannan, S. K., Kennard, C., Potter, D., Pan, Y., & Soto, D. (2010). Early oculomotor capture by new onsets driven by the contents of working memory. *Vision Research*, *50*, 1590–1597. doi:10.1016/j.visres.2010.05.015
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*, 254–278. doi:10.1037/0033-295X.100.2.254
- Menner, T., Barrett, D. J. K., Phillips, L., Donnelly, N., & Cave, K. R. (2007). Costs in searching for two targets: Dividing search across target types could improve airport security screening. *Applied Cognitive Psychology*, *21*, 915–932. doi:10.1002/acp.1305
- Menner, T., Cave, K. R., & Donnelly, N. (2009). The cost of search for multiple targets: Effects of practice and target similarity. *Journal of Experimental Psychology: Applied*, *15*, 125–139. doi:10.1037/a0015331
- Menner, T., Donnelly, N., Godwin, H. J., & Cave, K. R. (2010). High or low target prevalence increases the dual-target cost in visual search. *Journal of Experimental Psychology: Applied*, *16*, 133–144. doi:10.1037/a0019569
- Moore, C. M., & Osman, A. M. (1993). Looking for two targets at the same time: One search or two? *Perception & Psychophysics*, *53*, 381–390. doi:10.3758/BF03206781
- Mruczek, R. E. B., & Sheinberg, D. L. (2007). Activity of inferior temporal cortical neurons predicts recognition choice behavior and recognition time during visual search. *Journal of Neuroscience*, *27*, 2825–2836. doi:10.1523/JNEUROSCI.4102-06.2007
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*, 205–231.
- Navalpakkam, V., & Itti, L. (2007). Search goal tunes visual features optimally. *Neuron*, *53*, 605–617. doi:10.1016/j.neuron.2007.01.018
- Neider, M. B., & Zelinsky, G. J. (2006). Searching for camouflaged targets: Effects of target-background similarity on visual search. *Vision Research*, *46*, 2217–2235. doi:10.1016/j.visres.2006.01.006
- Olivers, C. N., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1243–1265. doi:10.1037/0096-1523.32.5.1243
- Olivers, C. N., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*, 327–334. doi:10.1016/j.tics.2011.05.004
- Palmer, E. M., Fencsik, D. E., Flusberg, S. J., Horowitz, T. S., & Wolfe, J. M. (2011). Signal detection evidence for limited capacity in visual search. *Attention, Perception & Psychophysics*, *73*, 2413–2424. doi:10.3758/s13414-011-0199-2
- Peelen, M. V., Li, F.-F., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*, 94–97. doi:10.1038/nature08103
- Pietrewicz, A. T., & Kamil, A. C. (1979). Search image formation in the blue jay (*Cyanocitta cristata*). *Science*, *204*, 1332–1333. doi:10.1126/science.204.4399.1332
- Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, *42*, 1447–1463. doi:10.1016/S0042-6989(02)00040-8
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh, PA: Psychology Software Tools Inc.
- Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *The Quarterly Journal of Experimental Psychology*, *62*, 1904–1914. doi:10.1080/17470210902853530
- Soto, D., Hodsoll, J., Rotshein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences*, *12*, 342–348. doi:10.1016/j.tics.2008.05.007
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*, 652–654.
- Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, *57*, 421–457.
- Sternberg, S. (1975). Memory scanning: New findings and current controversies. *The Quarterly Journal of Experimental Psychology*, *27*, 1–32. doi:10.1080/14640747508400459
- Stokes, M., Thompson, R., Nobre, A. C., & Duncan, J. (2009). Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences*, *106*, 19569–19574. doi:10.1073/pnas.0905306106
- Stroud, M. J., Menner, T., Cave, K. R., Donnelly, N., & Rayner, K. (2011). Search for multiple targets of different colours: Misguided eye movements reveal a reduction of colour selectivity. *Applied Cognitive Psychology*, *25*, 971–982. doi:10.1002/acp.1790
- Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene-viewing. *Journal of Eye Movement Research*, *2*, 1–18.
- Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioral biases in eye guidance. *Visual Cognition*, *17*, 1029–1054. doi:10.1080/13506280902764539

- Thiele, A., Henning, P., Kubischik, M., & Hoffman, K. P. (2002). Neural mechanisms of saccadic suppression. *Science*, *295*, 2460–2462. doi:10.1126/science.1068788
- Tinbergen, N. (1960). The natural control of insects in pine woods: Vol. I. Factors influencing the intensity of predation by songbirds. *Archives Neerlandaises de Zoologie*, *13*, 265–343.
- Usher, M., & Neiber, E. (1996). Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *Journal of Cognitive Neuroscience*, *8*, 311–327. doi:10.1162/jocn.1996.8.4.311
- Vickery, T. J., King, L., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*, 81–92. doi:10.1167/5.1.8
- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500–503. doi:10.1038/nature04171
- Watson, D. G., & Humphreys, G. W. (1997). Visual marking: Prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychological Review*, *104*, 90–122. doi:10.1037/0033-295X.104.1.90
- Watson, D. G., & Humphreys, G. W. (2000). Visual marking: Evidence for inhibition using a probe-dot paradigm. *Perception & Psychophysics*, *62*, 471–481. doi:10.3758/BF03212099
- Watson, D. G., Humphreys, G. W., & Olivers, C. N. L. (2003). Visual marking: Using time in visual selection. *Trends in Cognitive Sciences*, *7*, 180–186. doi:10.1016/S1364-6613(03)00033-0
- Wilschut, A., Theeuwes, J., & Olivers, C. N. L. (2013). The time it takes to turn a memory into a template. *Journal of Vision*, *13*, 1–11. doi:10.1167/13.3.8
- Wilschut, A., Theeuwes, J., & Olivers, C. N. L. (2014). Priming and the guidance by visual and categorical templates in visual search. *Frontiers in Psychology*, *5*, 1–12. doi:10.3389/fpsyg.2014.00148
- Woodman, G. F., Luck, S. J., & Schall, J. D. (2007). The role of working memory representations in the control of attention. *Cerebral Cortex*, *17*, 118–124. doi:10.1093/cercor/bhm065
- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238. doi:10.3758/BF03200774
- Wolfe, J. M. (2005). Watching single cells pay attention. *Science*, *308*, 503–504. doi:10.1126/science.1112616
- Wolfe, J. M. (2007). Guided Search 4.0: Current progress with a model of visual search. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). New York, NY, USA: Oxford University Press.
- 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 and Performance*, *29*, 483–502. doi:10.1037/0096-1523.29.2.483
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided Search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 419–433. doi:10.1037/0096-1523.15.3.419
- Wolfe, J. M., & Gancarz, G. (1996). Guided Search 3.0: A model of visual search catches up with Jay Enoch 40 years later. In V. Lakshminayanan (Ed.), *Basic and clinical applications of vision science* (pp. 189–192). Dordrecht, Netherlands: Kluwer Academic.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews: Neuroscience*, *5*, 1–7. doi:10.1038/nrn1411
- 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. doi:10.1016/j.visres.2003.11.024
- Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, *49*, 2095–2103. doi:10.1016/j.visres.2009.05.017
- Yang, H., Chen, X., & Zelinsky, G. J. (2009). A new look at novelty effects: Guiding search away from old distractors. *Attention, Perception & Psychophysics*, *71*, 554–564. doi:10.3758/APP.71.3.554
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, *115*, 787–835. doi:10.1037/a0013118
- Zhang, Y., Meyers, E. M., Bichot, N. P., Serre, T., Poggio, T. A., & Desimone, R. (2011). Object decoding with attention in inferior temporal cortex. *Proceedings of the National Academy of Sciences*, *108*, 8850–8855. doi:10.1073/pnas.1100999108
- Zhang, Y., Yang, H., Samaras, D., & Zelinsky, G. J. (2006). A computational model of eye movements during object class detection. In Y. Weiss, B. Scholkopf, & J. Platt (Eds.), *Advances in neural information processing systems* (Vol. 18, pp. 1609–1616). Cambridge, MA: MIT Press.