Target "templates":

How the precision of mental representations affects attentional guidance and decision-making in visual search.

by

Michael C. Hout

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2013 by the
Graduate Supervisory Committee:

Stephen Goldinger, Chair
Erik Reichle
Tamiko Azuma
Donald Homa

ARIZONA STATE UNIVERSITY

August 2013

ABSTRACT

When people look for things in their environment they use a target template – a mental representation of the object they are attempting to locate – to guide their attention around a scene and to assess incoming visual input to determine if they have found that for which they are searching. However, unlike laboratory experiments, searchers in the real-world rarely have perfect knowledge regarding the appearance of their target. In five experiments (with nearly 1,000 participants), we examined how the precision of the observer's template affects their ability to conduct visual search. Specifically, we simulated template imprecision in two ways: First, by contaminating our searchers' templates with inaccurate features, and second, by introducing extraneous features to the template that were unhelpful. In those experiments we recorded the eye movements of our searchers in order to make inferences regarding the extent to which attentional guidance and decision-making are hindered by template imprecision. We also examined a third way in which templates may become imprecise; namely, that they may deteriorate over time. Overall, our findings support a dual-function theory of the target template, and highlight the importance of examining template precision in future research.

DEDICATION

This work is dedicated to my mother, who I have to thank for both my curiosity about the world around me, and the intellect to explore it for myself. Each trip to the library, every stroll around a museum, each afternoon spent at the science center, and every bit of encouragement towards my studies helped instill in me an obsession with knowing how the world works. For that I am grateful beyond words.

ACKNOWLEDGMENTS

This work would not have been possible were it not for the profound influence of three friends. Melissa, for her decade-long unwavering support and encouragement towards reaching my goals. Erik, for igniting my interest in cognition, and for helping me make the first steps towards becoming a scientist. And Steve, for training me, for pushing me to excel at every aspect of this profession, and for being an exemplary mentor the likes of which I aspire to someday be.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

The world is replete with information of such richness and complexity that our visual systems are subjected to an overload of information. Simply put, it is impossible to simultaneously process all of the information we are confronted with on a moment to moment basis. Consider the simple act of sitting idly in a coffee shop, enjoying a drink. As you lazily take in your surroundings, your eyes are flooded with different colors, shapes, textures and contours. You spot brightly colored boxes of tea high on a shelf, round reclining chairs in your periphery, the smooth surface of a linoleum floor, and the sharp edges of a table that you'll need to avoid on your way out. All the raw information that is absorbed by your retina is assembled in your cortex in such a way that you perceive not features, but coherent objects. You don't see "*squat, opaque, and round*", you observe your coffee mug. "*Shiny, silver, and cylindrical*" is not a collection of features that your eyes passively absorb; rather, it is the container of milk that you're looking for on the condiment bar.

Despite the efficiency with which our visual system constructs these features into meaningful objects, we are limited in the extent to which this information can be processed at any given point in time. Largely, the limitation comes from the physiology of our retinas. The distribution of high-resolution cone photoreceptors is densely centered in the small *fovea* of our eyes. As such, we process visual scenes through a series of alternating fixations and saccadic eye movements. Over time, informative regions of space are briefly brought into focus on the fovea, and are stored in visual-short term memory, giving rise to the illusion that we have a static representation of our world. Each waking moment, our eyes dart around – usually three to four times per second – in search of objects of interest, familiar faces, or items for which we are searching. Chaotic as this behavior may seem, however, our eye movements are not simply random. We are not passive observers of the world; we are active seekers of information. And the way in which we examine a scene is highly governed by the low-level, visual characteristics of the environment, our experience viewing similar scenes in the past, and by the mental representation of that which we are seeking out.

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

The factors that guide visual attention can broadly be categorized as either "bottom-up" or "top-down" information sources. Under normal circumstances, processing seems to proceed from the "middle-out" (Kinchla & Wolfe, 1979). That is, both sources of information seem to be acted upon in concert. At times, one type of information may dominate the other, such as when a color singleton "pops out" from a display, despite whatever else you may be searching for (cf. Theeuwes & van der Burg, 2011). For instance, as a passenger in your friend's car, it is your job to be on the lookout for a landmark that signals where to make the next turn, yet you may find your attention being inadvertently captured by a changing stoplight.

When people look for things in their environment, they make use of 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 is used to select regions of contrast in a scene, such as changes in color, luminance, or intensity (Itti & Koch, 2000, 2001; Koch & Ullman, 1985). By contrast, higher-level (i.e., top-down) contextual knowledge is used to identify regions of a scene that are informative, or that have a high likelihood of containing the target object (Castelhano & Henderson, 2007; Neider & Zelinsky, 2006). Finally, the searcher's target "template" is used to match information in the scene to a stored representation of the target in visual working memory (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 is that that item will be selected for fixation (Eckstein et al., 2007; Findlay, 1997; Tavassoli et al., 2009; Becker, 2011; Olivers, Meijer & Theeuwes, 2006; Mannan et al., 2010). Returning to the example of searching for a milk canister at a coffee shop, low-level salience information is used to decompose the store into regions of coherence, contextual knowledge is what makes you more likely to look on the condiment bar than the merchandise display shelves, and your target template is what helps you direct your eyes to regions of space that potentially match your target, and to ascertain if the item you are currently fixating is in fact that which you seek.

Early theories of visual search, such as *Feature Integration Theory* (or *FIT*; Treisman & Gelade, 1980), focused largely on how the low level characteristics of a scene influence one's ability to find a target (see also *attentional engagement theory*; Duncan & Humphreys, 1989, 1992; Desimone & Duncan, 1995). Search is easy, for instance, when you are looking for a color "singleton", such as a red letter among distracting green letters. Search becomes more difficult when features have to be combined to locate a conjunction stimulus, as when searching for a red *X* among red *O*s and green *X*s. A key weakness of strictly bottom-up approaches to search, however, is that they constrain search behavior to a strict dichotomy of efficiency. That is, either search is easy, and therefore conducted by a parallel, pre-attentive processing mechanism, or it is difficult, in which case it involves a serial mechanism that maneuvers focal attention. Clearly, however, the reality is that search behavior can fall along a continuum of efficiency, depending on the lower level characteristics of a scene, and the top-down attentional set of the observer (Wolfe, 1998).

Although bottom-up information is undoubtedly necessary for basic visual processing (see Wolfe & Horowitz, 2004), it seems that relative to the influence of higher-level knowledge, the guidance of attention by lower level features has a rather limited utility (see Einhäuser et al., 2008; Henderson et al., 2007, 2009; Tatler & Vincent, 2008, 2009). For instance, it has been shown that 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 demonstrate remarkable feats of top-down attentional control: When only a subset of items are ever pertinent in a visual search task, they can learn to restrict their attention to the relevant stimuli (Kunar, Flusberg & Wolfe, 2008b; Frings, Wentura, & Wühr, 2012), and can learn the features of old distractors in order to successfully guide attention away from them (Yang, Chen & Zelinsky, 2009). The "preview benefit" – the finding that search is facilitated by a brief preview of all or part of a display prior to visual search – shows that people can prioritize the selection of new stimuli via the top-down inhibition of previously seen items (Watson et al., 1997, 2000, 2003). In essence, top-down guidance works in two ways: First, by biasing attention *towards* features or regions of space that are important, and second, by biasing attention *away* from features that are undesirable (or objects that have already been inspected; Al-Aidroos et al., 2012; Arita, Carlisle, & Woodman, in press; Woodman & Luck, 2007).

It has even been shown that top-down guidance is strong enough to override the capture of attention by low-level salience. Chen and Zelinsky (2006) had people search for a target (a picture of a real-world object) with or without a preview. In some conditions, a color singleton was present; without a preview, the singleton attracted attention, as indexed by initial saccades directed to that item. However, when given top-down context in the form of a preview, reaction times quickened, and more initial eye movements were guided to the target (rather than the singleton), suggesting that the top-down set overrode the bottom-up salience signal. It is unsurprising then, that more recent, successful models of visual search – such as the *Guided Search* model (Wolfe et al., 1989, 1994, 1996, 2007; Palmer et al., 2011) – incorporate top-down guidance as a key mechanism controlling attention.

It is reasonable to assume that an optimal system would make use of all available information sources in order to guide search. Indeed, recent computational models – such as the *contextual guidance* model (Torralba et al., 2006) and the *Saliency Using Natural statistics* model (SUN; Kanan et al., 2009) – have been developed to model the fixation positions of real observers (unlike Guided Search, which strictly models RTs), using one or more types of guiding information (see also Ehringer et al., 2009). Both of these models have shown that combining two sources of information (relative to, for instance, using only salience-driven guidance) significantly improves the model's ability to predict human fixation locations. This suggests that humans are using both types of information to guide their attention, even if one is more beneficial than the other. Taken together, the conclusion is clear: Our eyes are not slaves to the characteristics of the environment through which we are searching, but are instead guided by our previous experiences, and by the properties of the template we hold in memory.

The target template.

The idea of a target template was first proposed not by psychologists, but by ethologists, who were 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 a successful meal (Tinbergen, 1960; Pietrewicz & Kamil, 1979; Bond, 1983). In the human literature, the "target template" (also referred to as the "attentional template" or "search template") refers to the process whereby a person holds in visual working memory a representation of the to-be-located item, and uses this representation to facilitate the detection of the target (Wolfe et al., 2004; Malcolm & Henderson, 2009; Bravo & Farid, 2009, 2012; Vickery et al., 2005).

A considerable amount of work has been conducted with the aim of finding neuroscientific evidence of template use. Much of this work has been conducted with primates (see Evans et al., 2011; Chelazzi et al., 1993, 1998). Activation of the template involves neurons in the prefrontal cortex (PFC) that select and maintain behavioral goals (such as finding a target among distracting nontargets). These neurons project to inferotemporal (IT) cortex, where visual objects are believed to be represented (Mruczek & Sheinberg, 2007; Peelen, Fei-Fei, & Kastner, 2009). Importantly, the top-down input from PFC enhances the gain in IT neurons that are selective for the target object; in essence, the PFC appears to tell 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 lower-level 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, however, these cells also tend to have a preferred color and/or shape that they selectively fire in response to (see Wolfe, 2005, for an interesting treatment of this topic).

In a study using macaque monkeys, Bichot, Rossi, and Desimone (2005) recorded from neurons in V4. On trials where the target was (for instance) red, they found that red-selective neurons increased their firing rate even before the target had become the object of attention (i.e., before the monkey prepared to make an eye movement to the target location). Moreover, other neurons that shared selectivity for the target feature (e.g., "red") began to synchronize their activity, as if they were preparing to respond (more vigorously) to the presence of that feature. More recently, Eimer, Kiss and Nicholas (2011) used event-related potentials (ERP) to study the benefits of advanced preparation in visual search, with human participants. Their results suggested that advanced preparation (i.e., the use of a target template) accelerates target selection and resolves attentional competition by preventing irrelevant information from attracting attention or entering working memory. Taken together, these findings suggest an interesting mechanism by which a target template may guide visual search behavior: Implementation of the template enhances the firing of cells that typically respond to the features of interest, and may inhibit cells that typically respond to irrelevant or unhelpful ones (Desimone & Duncan, 1995; Usher & Neiber, 1996).

The ability to guide attention toward task-relevant objects (i.e., those that resemble the target) is critically dependent on the use of working memory (Soto et al., 2005, 2008, Kang et al., 2011), which allows the searcher to "load" information into VWM in as little as 50 ms (Vogel, Woodman & Luck, 2006). The storage capacity of

VWM is thought to be 3-4 objects (Cowan, 2001; Luck, 2008); an “object” does not mean a single feature, but rather, a perceptual whole that may be constructed from the conjunction of several features simultaneously (Luck & Vogel, 1997). At present, there is a debate regarding the extent to which attention can be guided by multiple search templates in tandem. By one account, only a single object can remain in a fully active state at any given time (McElree, 2001; Oberauer, 2002), and when a searcher uses multiple templates, s/he alternates between them, using only one to guide attention at any given moment (Olivers et al., 2011; Huang & Pashler, 2007). By another account, the visual system can store search representations within the visual system itself (e.g., within primary visual cortex), which should make it possible for multiple VWM representations to control attention simultaneously (cf. Stroud et al., 2011, 2012).

A recent article by Beck, Hollingworth and Luck (2012) shines light on this debate. They had participants search for Landolt Cs with gaps on the top or bottom (distractors were oriented with gaps to the left or right; participants indicated the location of the gap using a response box). In some conditions, people were induced to searching via a single template at a time: They were shown two colors (e.g., red and blue), and on 80% of the trials, the target appeared in the former color, and in 20% of the trials, it appeared in the latter color (distractors were of four different colors). Under these conditions, participants searched sequentially through many items of a single color (as indexed by “run lengths”: The number of consecutive items fixated of a single color), and then showed a switch cost before using the second template (as indexed by fixation durations on the last, for instance, red item prior to examination of blue ones). In a different condition, wherein color cues were not reliable (i.e., targets were red on 50% of trials and blue on 50%), participants demonstrated shorter run lengths and no switch costs, suggesting that they used both color templates to guide attention simultaneously. This finding is important because it suggests that there is not a bottleneck in top-down control whereby only a single object can guide attention at any given moment. Rather, the findings indicate that multiple VWM representations may concurrently interact with the flow of information through the visual system (cf., Hollingworth et al., 2008, 2009, 2012; Woodman & Luck, 2007, 2010).

The problem of template imprecision.

In typical laboratory search experiments, participants look for a single, unambiguous target on every trial. For instance, participants may be shown a picture of a bicycle and be asked to find that bicycle (unaltered) in a cluttered array of other objects. In other cases, a picture cue may not be provided; rather, a precise description of the target is given, such as when people are asked to search for a “vertical red bar” in a display that contains lines of various orientations and colors. In both cases, the searcher’s mental template is highly accurate. That is, they can anticipate exactly how the target will appear in the search array, either because they’ve already seen an isolated instance of that item, or because the target is so simply defined as to be unmistakable.

Real-life searches, by contrast, rarely afford the kind of template precision with which we provide our research participants. This is an inevitable truth, because the appearance of objects in the world is frequently variable over time. Consider the following scenario: Your spouse decided to treat your Scottish Terrier to a fun day spent in doggy-daycare. It’s your task to pick up your beloved pet on the way home from work. You arrive at an overcrowded daycare to retrieve your dog, but the place is sprawling, filled with dogs, and unfortunately your mutt does not respond to your calls. So now you’ve got to navigate the area and retrieve your Scottie, hidden among other black dogs (e.g. labradors), other toy breeds (e.g., miniature poodles), and all manner of frolicking, four-legged companions. Naturally, this is fairly easy for you. Your template is precise, because you know all the specific features of your dog. You know his height, the texture of his fur, the color of his collar, and so on. Knowledge of these features allows you to efficiently hone in on dogs with scraggly black fur, or those with thick red collars, and to quickly reject those with mismatching features.

However, it’s possible that your dog has changed in appearance since last you saw him. Consider performing the same undertaking, but now imagine that your spouse has also decided – without informing you – to pamper your pet by getting him a fresh grooming, and a new collar. Now, the specific features that you have in mind may be unhelpful, inaccurate, or even altogether wrong. You may have a hard time directing your attention accurately because you are incorrectly looking for a dog who has not been recently trimmed, or he may go unnoticed when your gaze falls directly upon him because you quickly reject this imposter due to the presence of a new thin blue collar.

In other situations, our templates may be made imprecise not through the intrusion of incorrect features, but by the presence of too many features, or too few. If your task is, for instance, to pick up a pair of dogs at doggy-daycare, you will now have considerably more features in mind, depending on the similarity of your dogs to one another. If you’ve got a pair of Scotties to locate, then your template is simple. It has all the general features of a Scottie, plus a few additional characteristics that may be specific to your pets (e.g., a blue collar for the boy and a pink one for the girl). By contrast, if you are trying to locate a Scottie and a Basset Hound, the task becomes considerably more difficult. Now you've got to simultaneously look for a black dog and one that is tri-color; one dog has short, perky ears, and another has long, droopy ones. To the extent that these dogs are unlike one another, your template becomes more and more broad, because you add features to it that, though necessary to the overall task, may be unhelpful with respect to the singular animal that is under your gaze at any given moment in time.

At the opposite end of the spectrum are situations in which you have only broad, categorical knowledge of your target, and thus cannot necessarily load specific features into memory when constructing your mental template. Perhaps the dog you are picking up belongs to a friend. You've never seen this animal before, but you have a loose, verbal description of it. Here, your template is quite imprecise, because you've had little or no direct experience with the target you seek out. You know the dog has "scraggly" fur, but are unsure if that description indicates a dog with bristly, wiry fur, or longer, wavy fur. It is worth noting that many of our day to day searches are conducted in an analogous manner. That is, we very frequently construct a mental template not by calling to mind the specific features of one exemplar, but rather by assembling the features of our target that are typical of its category. For instance, when looking for a rake in a friend's cluttered garden shed, calling to mind a mental picture of your own rake(s) may not be helpful. Rather, you search using the features you know are definitive of the category "rakes", and perhaps allow your attention to be also guided to those features that are likely to appear. You direct your attention to long, skinny tools, for instance, but will allow yourself to more closely examine those with either wooden or plastic handles.

Although template imprecision may increase the difficulty of completing a search, it is unlikely that any of these tasks are insurmountable. Like so many other daily activities we perform, we are often blissfully ignorant of the obstacles our cognitive systems overcome on a regular basis. But the changes in behavior across these situations is potentially great, and perhaps even lawful. Search speed likely slows down as a function of template imprecision, but why would that be so? The overarching question that we seek to answer in this investigation is, “how does the precision of your internal template affect your ability to locate an item of interest?” More specifically, we want to know the extent to which the similarity between the actual to-be-located item and your mental representation of that item influences (1) the efficiency with which attention can be directed to the target, and (2) the speed (and accuracy) with which one is able to identify the target once attention has fallen upon it.

Importantly, this work may have considerable implications, particularly for professional visual searchers, such as airport baggage (or medical) screeners. Consider the difficulty of their profession: Their searches are conducted in high-stress environments, with constant time pressure. The outcome of their search often lacks explicit feedback, and the cost of missing a target could be catastrophic (e.g., a weapon makes it aboard a plane, or a tumor goes undiagnosed). Moreover, they must look for many potential threats simultaneously. The searchers’ templates are often poorly specified because, for instance, real weapons are oriented bizarre ways, may be occluded by other items in the baggage, and so on. The way in which searchers mentally represent potential targets, therefore, must be flexible and highly tolerant to deviations. If they search for too many features, it takes too long, and the line backs up. If they search for too few features, or the wrong ones, they may miss a weapon altogether. Certainly, understanding the nature of a searcher's target representation may have important implications for everyday life, and for high-stakes searchers in particular.

Simulating template imprecision using template-to-target similarity.

Similarity is a pervasive theoretical construct across many fields of psychology, including memory (e.g., Gillund & Shiffrin, 1984; Hintzman, 1986), language (e.g., Goldinger, 1998; Goldinger & Azuma, 2004), categorization (e.g., Homa, Proulx, & Blair, 2008; Homa, Hout, Milliken & Milliken, 2011), and others (see Shepard, 1987, 2004 for a more thorough treatment). Often, theorists rely critically upon a notion of "sameness" between stimulus items (Goldstone & Medin, 1994; Medin, Goldstone, & Gentner, 1993). For instance, imagine you stumble across someone that seems familiar to you. You cannot remember where (or if) you have seen them before, so you engage in a process whereby their face is compared to imperfect "traces" that you have stored in long-term memory (e.g., Hintzman, 1984, 1988). If their face is highly similar to some of these stored experiences, you may conclude that you have encountered this person before; if the similarity is low, you may decide that this person is a stranger to you. Analogously, when conducting a visual search, you may compare the item you are gazing at to a stored representation of the target held in visual working memory. When the perceived item is similar to that stored in VWM, you conclude that you have found your target; otherwise, you continue searching (Hout & Goldinger, 2011).

Broadly, similarity is a concept that has been crucial to theories of visual search. Building from Treisman's *Feature Integration Theory* (Treisman & Gormican, 1988; Treisman & Sato, 1990), Duncan and Humphries (1989; 1992) proposed a new theory of visual search that showed how search efficiency was hindered with increasing similarity of targets to nontargets, and by decreased similarity between nontargets (relative to one another). When looking for your Scottie, for example, the presence of other small, black dogs will make it harder to find your own, and if the doggy-daycare has a wide assortment of dogs, it will be more difficult to locate your pet, relative to searching through a homogenous collection of bulldogs. More recently, models of visual search have been developed, many of which rely extensively on visual similarity relationships (e.g., Hwang, Higgins, & Pomplun, 2009; Wolfe, 1994; Zelinsky, 2008).

Although a vast literature exists which examines target-to-distractor similarity effects, and distractor-to-distractor similarity effects during search, there is a notable absence of work on template-to-target effects (i.e., the degree to which the to-be-located item matches your internal representation of it). Largely, this may be because we can simply never know with much certainty how a person’s guiding representation is constructed. This is a difficult problem to overcome. The closest approximation to this line of questioning are experiments that compare template-guided search (i.e., search using picture cues) to categorical search (i.e., search using word cues). In such experiments, an assumption is made that the searchers’ representations differ across the two situations: In template-guided search, the observer has a (nearly) veridical representation of the target in mind, and in categorical search, the observer has a much less precise guiding representation. The initial approach of this investigation, however, is different in that we employed two (new) distinct techniques for manipulating the precision of our searchers’ internal representations. More specifically, we aimed to induce a continuum of precision in our searchers’ templates by manipulating the similarity of the target cue (or cues), relative to the target that actually appeared in the search display.

Our first technique involved a search paradigm where on the majority of trials, the searcher was shown a target, and the item that showed up in the display matched that cue exactly. Instructions indicated that the target cue may not be entirely specific to the item they were to find: “Please search for this item or something very much like it.” On a minority of (randomly-dispersed) trials – the critical trials – the target that appeared in the display was a slight variation of that which was shown as a cue prior to search. For example, the participant may have been shown a target coffee mug and then, in the search display, encountered the same mug, but oriented differently. Or, their pre-search cue may have been a starfish, but the starfish encountered during search may have been a different color. In general, our goal was to encourage people to adopt a particular, item-specific template which was, on most trials, wholly accurate. Then, performance on these exact match trials could be contrasted with situations in which the actual target was a slight mismatch to the cue. By varying the similarity between these pairs of images (i.e., between the target cue and the target that actually appeared in the search display), inferences could then be made regarding the extent to which behavior was affected by precise or imprecise guiding representations. Importantly, although the to-be-located target was sometimes different in appearance from its cue, our targets were always unmistakable. When searching for a coffee mug, for example, the participant never encountered any other items (i.e., distractors) that could have been misconstrued as the target, therefore making our target cues imprecise, but unambiguous.

The second technique we employed involved manipulating the “width” of our searchers’ template “feature space”. For this method, multiple-target search was conducted (though, we included single-target trials for control conditions). Two target cues were shown prior to search, and the participants’ task was to locate one of the items (only one ever appeared in the search display). Here, we widened the feature space of the search template by manipulating the similarity of the two potential target cues to one another. When the two target cues were similar to one another (e.g., the same coffee mug, oriented in two different ways), the feature space was narrow, but to the extent that these two pictures were dissimilar to one another, the feature space became broader through the inclusion of new, potentially important feature dimensions (e.g., two different color dimensions, imposed by using a pair of distinct starfish cues). By examining behavior across these different conditions, inferences could then be made about how search behaviors are affected by the width of the searcher’s guiding template.

In essence, these two techniques allowed us to examine situations in which the searcher’s template is made imprecise either through the inclusion of inaccurate features, or the addition of unhelpful ones. It is also worth mentioning that we stand to gain much from the analysis of eye movements during visual search. It is well established that attention and eye movements are linked, at not just a functional (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Williams et al., 1997), but a neuroanatomical level (Corbetta et al., 1998). Importantly, for this investigation, we assume that eye movements can be used to directly infer cognitive processing during search. This is an assumption that is made across a wide variety of studies in psychology. Specifically, it has been shown that fixation durations are not arbitrary or random (Hooge & Erkelens, 1998), rather, that they reflect item processing and the properties of the stimulus being viewed (Rayner, 1998; Reingold et al., 2001, 2012; Henderson & Ferreira, 1990; Becker, 2011; Pannasch, Schulz, & Velichkovsky, 2011). This finding has been supported by research both within the field of visual search – for instance, in visual search through real-world scenes (Henderson & Pierce, 2006; Henderson & Smith, 2009; Henderson, 2012; Nuthmann, et al., 2010) – and elsewhere, such as in reading research, wherein fixation durations are thought to reflect the time it takes for lexical access to occur on the currently fixated word (Pollatsek, Reichle, & Rayner, 2006; Rayner et al., 2004; Reichle et al., 1998, 2006, 2009, 2012). Moreover, saccade metrics have been used to infer cognitive processing, particularly in the realm of learning and practice, where it has been found that saccadic metrics often account for a substantial amount of variability in the performance of observers over time (Phillips & Edelman, 2008; but see also McCarley et al., 2004; Kotowicz, Rutishauser & Koch, 2010). Clearly, much can be gained from the analysis of eye movements during our search tasks, particularly when combined with standard behavioral indexes.

Obtaining stimulus pairs along a continuum of similarity.

Empirically, manipulating the similarity of stimulus items is a challenging task. Early work accomplished this by employing simplistic stimuli, and varying a single feature of an item, such as the color or orientation of a rectangular bar (e.g., Treisman, 1991). But this is a suboptimal approach to higher-level vision. Real-world objects are composed of many features, some of which may be ill-defined or inconsistent with respect to the ability of a searcher to guide attention based on them. It is clear that properties like color, motion, and orientation are attributes by which one can guide attention, but it is less clear whether or not people can direct their attention to items based on luminance, shape, or shading (Wolfe & Horowitz, 2004). Importantly, these latter properties are indeed aspects of a stimulus that are subject to change over time. The shape of your favorite coffee mug changes dramatically if the handle were to break off, so your ability to direct attention to the mug in a cluttered cabinet may be hindered if your mental representation of it has a fixed shape (e.g., round, with a protruding handle). Moreover, it is impossible to know *a priori* which features a person is likely to favor during search, and the salience of various stimulus dimensions is likely to change across contexts (cf. Adams & Chambers, 2012). When looking for your Scottie, you would probably seek out dark-colored dogs if most of the animals have light fur, or instead focus your attention on small creatures if the daycare houses lots of larger breeds.

The approach taken in this investigation (with respect to manipulating item similarity) was twofold. First, we identified semantically matched pairs of images that were classified as *state*- or *exemplar-pairs*. State-pairs were two pictures of the same item in different perceptual states, and exemplar-pairs were two distinct items from the same semantic category. We argued that because the state-pairs involved two variations of the same object, they were more similar to one another than were the exemplar pairs. Moreover, in Experiment 1, we established limited empirical support for this claim, through the use of two perceptual discrimination tasks.

Our second approach was, in Experiment 2, to obtain large-scale, explicit estimates of similarity, and to use multidimensional scaling (MDS) to find combinations of stimuli with varying degrees of similarity to one another. The appeal of this technique is that it allows people to provide similarity estimates based on whatever featural dimensions they so choose. An MDS approach does not require *a priori* identification of feature dimensions, or arbitrary rating schemes (e.g., ranking the similarity of colored bars based on degree of rotation, or distance in RGB "color space"). An inescapable concern is that with real-world objects (which are the sole focus of this investigation), similarity estimates may be governed by visual and nonvisual features alike (cf. Wolfe et al., 1992; Võ & Henderson, 2009), and there is currently a debate in the literature regarding the extent to which attention can be guided by semantic attributes (e.g., Godwin, Hout, & Menneer, *under review*; Schwarz & Eisel, 2012). A curtain and a blanket, for example, are perceptually similar and conceptually dissimilar, whereas a curtain and window blinds serve a common purpose, despite being visually quite different. When searching for curtains then, is one more likely to direct their attention to the visual competitor (the blanket) or the semantic one (the blinds)? Some reports suggest that semantic information can guide even very early eye movements (Rayner, Castelhano, & Yang, 2009; Underwood et al., 2008; Yang & Zelinsky, 2009), but others have suggested that eye movements are not guided by semantic information at all (De Graef, Christiaens, & d'Ydewalle, 1990; Henderson, Weeks, & Hollingworth, 1999).

An interesting, analogous experimental technique was employed by Alexander and Zelinsky (2011), who investigated how similarity relationships affect search guidance during categorical (i.e., word-cued) search. They collected similarity rankings over the internet: Volunteers were shown five pictures (e.g., a lamppost, a desk, a stuffed rabbit), and were asked to rank-order the items from most to least similar, relative to the target category of interest (teddy bears or butterflies; target categories were always word-cued but were now shown pictorially). These similarity estimates were used to create visual search backgrounds that were composed of distracting nontargets with varying degrees of similarity to the target category. For instance, when searching for a teddy bear, observers sometimes encountered distractors with high similarity to the target category (e.g., other stuffed animals) and at other times, saw distractors with low similarity (e.g., furniture). They also implemented a computer vision model that performed the same rank-ordering task by analyzing the pictures for color, texture, and shape similarity. Often, but not universally, the human and model similarity estimates were in agreement. In the final, critical experiment, they compared search behavior (i.e., search RTs, fixation durations) on backgrounds that were derived from cases upon which the model and human rankings agreed, and ones wherein the rankings differed. Their findings revealed that similarity effects in visual search were best predicted by cases wherein the humans and model agreed, suggesting that this technique captured aspects of similarity that are important for guiding attention to categorically-defined targets.

Because we were exclusively concerned with the similarity among pairs of semantically-matched images, the possibility that conceptual factors would contaminate our participants' similarity ratings was a non-starter. Simply put, our similarity ratings were always obtained within-categories, rather than across-categories, effectively isolating the visual characteristics of our stimuli from the conceptual ones.

Experiment 1

In any investigation of visual search, it is important to consider the complexity of the stimuli that are to be employed. In our experiments, we used pictures of real-world objects. This is an appropriate level of complexity, because it strikes a balance by achieving a desirable degree of representativeness (or ecological validity), without sacrificing experimental control. Restricting stimulus complexity to simple, arbitrary objects (e.g., rotated, colored shapes) may permit a high degree of experimental control, in that visual similarity (across items) can easily be assessed and manipulated. But real-world searches are rarely guided by a few arbitrary features, such as “red, 45-degree rotation, rectangular”.

Instead, we used visually complex pictures of real-world objects; things that people are likely to encounter in everyday life, and that are defined by a conjunction of several features simultaneously. This afforded us flexibility in stimulus selection, in that we could vary the individual features of an item without compromising its categorical distinction. That is, we were able to vary the specific features of the target (from when it is shown as a cue prior to search, relative to its appearance in the search display itself), while maintaining its categorical label unambiguously. For instance, a participant may have been shown a picture of a coffee mug as a cue, and been asked to search for “this or something very much like it”. Then, in the search display, the mug may have changed orientation or color. The target category was still “coffee mug”, but the particular features of that item were altered, thus simulating a situation in which the searcher has an imprecise template.

Unlike real-world search, however, the backdrops of the displays in these experiments were simple (blank white screens). The reason for this is that the use of real-world scenes (e.g., searching for a coffee mug in a kitchen scene) introduces potential confounding factors related to issues of top-down attentional control. The primary goal of this investigation is to examine how someone's mental representation of a target affects their ability to find it. But when people look through real-world scenes, they bring all kinds of other top-down knowledge to bear on the task.

In nature, organisms and objects rarely appear as they do in an experiment: Isolated, scattered arbitrarily across a blank background. Rather, they appear in predictable locations, relative to the rest of the scene (Biederman, 1972). If you are asked to locate a person in a photograph of a college campus, you are unlikely to search for someone scaling the wall of a building, or hovering in space. Instead, you will look for people standing on the sidewalk, sitting on the benches, and so forth. The visual system is sensitive to these regularities. For example, scan paths to objects that appear in natural locations (e.g., a beverage located on a bar) are shorter, relative to objects that appear in arbitrary locations (e.g., a microscope located on a bar; Henderson, Weeks, & Hollingworth, 1999), and models that incorporate such top-down contextual information have been successful in extrapolating regions of interest when compared to data from human observers (Oliva, Torralba, Castelhano, & Henderson, 2003). By limiting the observers' useable top-down knowledge solely to the features of the target stimulus, and not the background, the cognitive processes of current interest are effectively isolated.

All of our stimuli came from the "Massive Memory" database (Brady et al., 2008; Konkle et al., 2010; cvcl.mit.edu/MM/stimuli.html). A distinction that is used in describing these stimuli is to identify pairs of items as “state-pairs” or “exemplar-pairs” (see Figure 1). State-pairs are two pictures of the same object, presented in different perceptual states. For instance, an old-fashioned phone, and the same phone oriented differently, with the receiver off-the-hook. Exemplar-pairs are two discrete items that are compiled from the same semantic category. For instance, two different cows. As shown in Figure 1, our stimuli had a range of variation in how similar the pairs of images are to one another. For the abacus state-pair, the difference is visually very subtle, just like the exemplar-pairing of bricks. By contrast, there is arguably greater visual dissimilarity for the piggy-bank state-pair, and the pair of starfish exemplars. Clearly, there is a wide range of similarity both within and across categorical labels. On average, it seems likely that the state-pair items will be more similar to one another, relative to the exemplar-pairs. It is intuitive that one should find greater variation across objects, relative to when a single object is manipulated in some way. Indeed, this is an assumption that is made in the current experiments.

The goal of Experiment 1 was to provide empirical support for the assumption that state-pairs are more similar to one another than are exemplar-pairs. A simple way to validate this notion could be to present people with pairs of stimuli, and have them rate (using a Likert-scale) how similar each pairing is. However, such a procedure is imperfect for two reasons: 1) Participants are likely to decipher the purpose of the experiment, which may influence their responses (e.g., “These are two different cows so I should rate them more dissimilar, relative to the phone I just saw in two orientations”). Secondly, 2) participants may allow nonvisual, semantic features to influence their ratings (e.g., “Those two pairs of bricks should be rated as dissimilar because I know they are discrete items, even though they are visually very alike"). Because the goal is to obtain estimates based solely on the perceptual characteristics of the stimuli, rather than semantic ones, we performed two perceptual experiments, wherein people had to quickly determine whether a pair of images belonged to the same semantic category, or different ones. In Experiment 1a, participants saw two pictures presented simultaneously, and in Experiment 1b, participants kept one item in memory, and then made a swift eye movement to the comparison item.

Experiment 1a Method

Participants.

Thirty-seven 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.

Design.

There were two trial types: 200 *same-category* trials, and 200 *different-category* trials, presented in random order. Of the 200 *same* trials, 100 contained an exemplar-pair of images, and 100 contained a state-pair (presented in random order). On the *different* trials, a single image was quasi-randomly selected from two distinct categories, with the stipulation that each image be used once (and only once). Thus, each image was shown twice in the entire experiment: Once on a *different* trial, and once on a *same* trial. There were four blocks of trials (with 100 trials per block), and a brief (2-minute) rest in between each block.

Stimuli.

The stimuli were photographs of real-world objects, resized (maintaining original proportions) to a maximum of 2.5° of visual angle (horizontal or vertical), from a viewing distance of 55 cm. Images were no smaller than 2.0° of visual angle along either dimension. 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).

Apparatus.

Data were collected on up to 12 computers simultaneously (all had identical hardware and software profiles). Dividing walls separated each subject station, and experimental sessions were monitored at all times by one or more research assistants. 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 (operated at a 60 Hz refresh rate). Displays were controlled by an Intel G41 Express chipset, and the operating systems were Windows XP. E-Prime v2.0 Professsional software (Schneider, Eschman, & Zuccolotto, 2002) was used to control stimulus presentation and collect responses.

Procedure.

Each trial began with a 500 msec fixation cross, followed by the presentation of a pair of images, side by side (placement of each image on the left or right was randomized). Participants rested their fingers on the keyboard at all times, and pressed "f" when the two images were members of the same semantic category, and "j" when the items were from different categories. The images remained on screen until the participant responded, or 2500 msec had elapsed (in which case the trial was marked as incorrect). Feedback (in the form of a centrally-presented, large red "X") was given only for incorrect responses, to encourage accurate responding. A 1000 msec inter-trial interval (ITI) preceded the start of the next trial. People were encouraged to respond as quickly and accurately as possible, and were constantly reminded of the keyboard response options (text indicating the keys for "same" and "different" was displayed at the bottom of the screen throughout the experiment).

Experiment 1a Results

All data were analyzed using 3 (Trial Type: different category, same category state-pair, same category exemplar-pair) x 4 (Block: 1-4) within-subjects, repeated measures ANOVAs. One participant was excluded from the analyses for producing reaction times that were more than 2.5 standard deviations above the mean of all participants.

Accuracy.

We found a main effect of Trial Type, with the best performance on different trials (97%), followed by state-pair same trials (95%), and exemplar-pair same trials (93%); $F(2, 34) = 20.83$, $p < .01$, $n^2_p = .55$. See Table 1 for all observed means. There was no main effect of Block, nor a Trial Type x Block interaction (*F*s < 2).

Reaction time.

Only RTs on correct response trials were analyzed. There was a main effect of Trial Type (see Figure 2), with fastest responding on state-pair same trials (684 msec), followed by exemplar-pair same trials (707 msec) and different trials (727 msec); $F(2, 34) = 11.07$, $p < .01$, $n^2_p = .39$. Planned-comparisons revealed a significant difference between state- and exemplar-pair same trials ($p < .01$). We also found a significant main effect of Block: RTs decreased over time (784, 696, 681, and 663 msec for Blocks 1-4, respectively); $F(3, 33) = 23.16$, $p < .01$, $n^2_p = .68$. The Trial Type x Block interaction was significant, $F(6, 30) = 3.76$, $p < .01$, $n^2_p = .43$, indicating that RTs quickened more dramatically on different trials (by 50 msec per block), relative to exemplar- and state-pair same trials (by 31 and 32 msec per block, respectively).

Experiment 1a Discussion

These results lend credence to the contention that our state-pair stimuli are more similar to one another than are exemplar-pairs. People more quickly indicated that state-pair images were members of the same category, relative to exemplar pairs. We conducted a second experiment in an attempt to replicate this finding, and to bring the procedure closer in spirit to the behaviors performed during a visual search task. In Experiment 1a participants saw both images at the same time. They were presented close enough to one another that a viewer could encode both of them simultaneously (if looking in between the pair) without moving their eyes. However, it seems more likely that people looked at one image, then the other (and perhaps back again), before indicating their decision. During visual search, however, one does not compare two items that are both visible. Rather, each observed item is compared to a target representation held in memory.

In Experiment 1b, participants were asked to do two things that are ubiquitous in visual search: 1) to hold the target image in memory, and 2) to rapidly shift attention to the location of the target before responding. Participants were first shown a single item, and then made a speeded saccadic eye movement to an unpredictable location, where the second image from the pair was shown.

Experiment 1b Method

The design, stimuli, and apparatus were identical to Experiment 1a.

Participants.

Seventeen new students from Arizona State University participated in Experiment 1b as partial fulfillment of a course requirement. All participants had normal or corrected-to-normal vision, and all reported normal color vision.

Procedure.

Participants first saw a central fixation cross (see Figure 3) for 500 msec, followed by a single image (centrally presented) for 500 msec. Then, a noisy backward-mask appeared (at the location of the first item) for 250 msec, followed by a 50 msec blank screen, and then the second image. This item was shown until the participant responded, or 2500 msec had elapsed (in which case the trial was marked as incorrect). The second item was shown in one of four locations: Equidistant from the center of the screen, at the top, bottom, left or right of fixation.

Experiment 1b Results

All data were analyzed in the same manner as Experiment 1a.

Accuracy.

There was a main effect of Trial Type, with the best performance on different trials (98%), followed by state-pair same trials (95%), and exemplar-pair same trials (93%); $F(2, 17) = 28.30$, $p < .01$, $n^2_p = .79$. See Table 1 for all observed means. There was no main effect of Block, nor a Trial Type x Block interaction (*F*s < 2).

Reaction time.

We found a main effect of Trial Type (see Figure 2), with fastest RTs on state-pair same trials (586 msec), followed by exemplar-pair same trials (595 msec) and different trials (620 msec); $F(2, 15) = 9.30$, $p < .01$, $n^2_p = .55$. However, planned-comparisons revealed no significant difference between state- and exemplar-pair same trials (p =.57). There was a significant main effect of Block, with RTs decreasing across blocks (682, 593, 575, and 552 msec, for Blocks 1-4, respectively); $F(3, 14) = 21.59$, $p < .01$, $n^2_p = .82$. The Trial Type x Block interaction was also significant, $F(6, 11) = 9.39$, $p < .01$, $n^2_p = .84$, indicating a sharper drop in RTs for different trials (by 56 msec per block), relative to exemplar- and state-pair same trials (by 34 and 32 msec per block, respectively).

Experiment 1b Discussion

The trends of the data in Experiment 1b appear to be a replication of Experiment 1a: Different trials elicited the longest RTs, and state-pair same trials the shortest. However, the difference between the two types of same trials was smaller in this experiment, and was not statistically reliable. The conclusion from these experiments is simple. The trends in accuracy and RT were all in the anticipated direction, if our contention that state-pairs are more similar (relative to exemplar-pairs) holds true. We must make this conclusion with some caution, however, due to the unreliable RT finding in Experiment 1b. Nevertheless, to preview the findings from forthcoming experiments, these small perceptual differences can elicit big changes in visual search behavior when they are used to simulate imprecise target representations in memory.

Experiment 2

In Experiment 1, we identified two classes of stimulus pairings: exemplar-pairs, and state-pairs. The latter are assumed to be more similar to one another than are the former, and beginning in Experiment 3, this assumption plays a critical role; one that allows us to manipulate the precision of a searcher's mental representation of the target. The manner in which this is accomplished relies critically on being able to faithfully estimate the similarity among pairs of images. However, the results from Experiment 1 indicate that the range of perceptual dissimilarity across these two stimulus classes is small, and potentially fragile. That is, the state-pair images seem (on average) to be more similar to one another than are the exemplar-pairs, but the perceptual evidence supporting this claim is not robust.

In Experiment 2, we obtained similarity ratings in a more direct manner, by employing multi-dimensional scaling (MDS). Our goal was to acquire a wider range of perceptual similarity in our stimulus pairings, in order to implement converging techniques for manipulating target template precision. MDS is a tool by which researchers can obtain quantitative estimates of the similarity among groups of items (Hout, Papesh & Goldinger, 2013); in this instance, pictures of various object categories. More specifically, MDS is a set of statistical techniques that takes as input item-to-item similarity ratings. It then uses data-reduction procedures to minimize the complexity of the similarity matrix, permitting (in many cases) a visual appreciation of the underlying relational structures that were used to govern the similarity ratings.

As in Experiment 1, our stimuli were drawn from the Massive Memory database. We acquired a large selection of semantically-matched pictures: 200 object categories that were represented by 17 individual exemplars, and another 40 object categories with 16 exemplars each. We then obtained MDS spaces for each of these 240 image categories (e.g., an MDS space of 17 coffee mugs, or 16 lamps). These "psychological spaces" were used to identify stimulus pairings with varying degrees of similarity. The logic is simple: By using the MDS spaces, pairs of images can be identified that are similar or dissimilar, indexed by their proximity in psychological space. Thus, rather than having a simple dichotomy of similarity (defined by the use of state- or exemplar-pairings), we were able to define pairwise similarity along a continuum of psychological distances. In addition to providing a wider range of similarity, this technique is appealing because it provides similarity ratings that are psychologically tractable (rather than defined somewhat arbitrarily), and because the large sets of category exemplars provide far more stimulus pairings to draw upon in our more substantive experiments (i.e., Experiment 3 onward).

Likely the most common method for obtaining similarity estimates is simply to ask people to numerically rate object pairs via Likert scales. In this technique, ratings are collected for every possible pairwise combination of stimuli (e.g., "respond '1' when the items are most similar and '9' when the pair is most dissimilar"). This *pairwise method* is useful and simple to implement. However, when the set of stimuli to be scaled is large, this technique is not ideal: The number of comparisons necessary to fill an item-to-item similarity (or *proximity*) matrix grows rapidly as a function of stimulus set size, leading to lengthy experimental protocols. Data collection becomes cumbersome, and concerns therefore arise regarding the vigilance of the raters (see Hout, Goldinger & Ferguson, 2013).

An alternative way to collect similarity estimates is the spatial arrangement method (*SpAM*), originally proposed by Goldstone (1994; see also Kriegeskorte & Marieke, 2012). This technique is faster and more efficient than the pairwise method, and produces output data of equal quality, relative to its more well-established counterpart (Hout, Goldinger & Ferguson, 2013). Here, many (or all) of the to-be-rated stimuli are presented at once, and participants move the items around on the computer screen, placing them at distances from one another that reflect the rater's subjective similarity estimates (items that are rated as similar are placed close to one another, and dissimilar items are placed proportionately farther away). The task can be conceptualized as having people project their own psychological spaces onto a two-dimensional plane (i.e., the computer screen). Once the participant has finished organizing the space, a proximity matrix is derived from the pairwise Euclidean distances (measured in pixels) between every pair of items. This technique is extremely well-suited for collecting a large quantity of MDS data in a relatively short period of time (a set of 17 stimuli would be scaled in roughly 20 minutes using the pairwise method, whereas SpAM could be completed in under 5 minutes).

Experiment 2 Method

The apparatus was identical to Experiment 1. Tables A1 and A2 (in the Appendix) display a full list of categories for which MDS data was collected. Also shown are the number of participants that contributed data to the scaling of each category.

Participants.

Two-hundred and forty new students from Arizona State University participated in Experiment 2 as partial fulfillment of a course requirement. Additionally, ten undergraduate research assistants (from the *Memory and Language Laboratory* at Arizona State University) participated voluntarily. All participants had normal or corrected-to-normal vision, and all reported normal color vision.

Design.

One-hundred and fifty participants completed an experiment wherein they were given 15 SpAM trials. On each trial, a new category of images was shown to the participant. Selection of image categories was counterbalanced, such that for every 16 participants, all 240 categories were scaled exactly once. We also collected a smaller amount of data from 90 participants. These people completed the same task, but were only administered 10 SpAM trials (with randomly selected categories). This shortened version of the task was given because the SpAM task was appended to an unrelated experiment. Finally, the 10 research assistants that participated in the experiment completed 16 sessions (with 15 SpAM trials per session), over the course of an entire semester. Thus, each of these participants scaled every one of the 240 image categories exactly once. For each image category, between 15 and 24 participants contributed similarity ratings (M = 19.48).

Stimuli.

All stimuli were photographs of real-world objects, resized as in Experiment 1. The pictures contained no background; a single object or entity was present in each image. Two-hundred of the categories of images contained 17 distinct exemplars; forty of the categories contained 16 exemplars.

Procedure.

On each trial, a new category of images was shown to the participant, arranged in discrete rows, with random item placement. People were instructed to drag and drop the images in order to organize the space such that the distance among items was proportional to each pair's similarity (with closer in space denoting greater similarity). Participants were given as much time as they needed to scale each category; typically, trials lasted between 2 and 5 minutes. Once participants finished arranging the items, a right mouse-click completed the trial. In order to avoid accidental termination of the trial, participants were then asked if they were satisfied with the space, or if they needed more time (responses were indicated via the keyboard).

Experiment 2 Results

MDS algorithm.

For each data set, we used the *PROXSCAL* scaling algorithm (Busing et al., 1997), with 100 random starts, via SPSS 20.0 (SPSS Inc., 2011). This algorithm uses a least-squares method of representation, and can accommodate multiple data sources (i.e., multiple participants).   The use of multiple random starts protects against the risk of the iterative scaling process falling into a local minimum (with respect to model fit). Each time the algorithms were implemented, the data were scaled 100 times, and the solution with the lowest *stress* value (i.e., a measure of the fit between the distances in space and the input proximities, with lower values indicating closer fits) was selected for output.

Choice of dimensionality.

The stress of the solutions were used to determine the appropriate dimensionality in which the data should be scaled. In order to choose the “right” number of dimensions, *scree plots* were created, which display stress as a function of dimensionality. Stress typically decreases with the addition of each dimension, but a useful heuristic is to look for the “elbow” in the plot; the value at which added dimensions cease to substantially improve fit (Jaworska & Chupetlovska-Anastasova, 2009; Lee, 2001). Typically, dimension selection is performed conservatively, because increasing the dimensionality of the MDS solution is not universally beneficial. Moreover, a common goal of MDS is to yield solutions in sufficiently low dimensionality to permit visual examination of the data (Rabinowitz, 1975). Choosing the correct dimensionality, therefore, depends on stress levels and on interpretability (Kruskal & Wish, 1978). In essence, one must strike a balance between finding a “good” solution, and one that is interpretable. For the vast majority of data sets, the elbow of the scree plot appeared at Dimension 2 or 3. To be consistent across stimulus sets, all data were scaled in three dimensions.

Sample analysis.

The quantity of the MDS data prevents a full report of the results in this document. However, below, we provide a sample analysis on two of the stimulus sets (teddy bears and butterflies), for demonstrative purposes. All data were analyzed in identical fashion. For the teddy bear category, 19 participants contributed data, and for the butterflies, there were 20 data matrices. Figure 4 shows scree plots for both stimulus sets. Clearly, stress is reduced to the largest degree moving from a single dimension to two dimensions, but the addition of a third dimension also appears to be meaningful (for both data sets). Higher dimensions (i.e., 4+) do not appear to substantially increase the quality of the solutions. Figures 5 and 6 show MDS plots of the stimuli, in three two-dimensional “maps”. Dimension 1 is the primary dimension (i.e., the one that explains the largest amount of variance), followed in order by Dimensions 2 and 3. See Figures A1 and A2, in the Appendix, for the MDS coordinates for these stimuli.

Identification of item pairings along a continuum of similarity.

With the coordinates obtained from each psychological space, it was possible to identify pairs of items that are more or less similar to one another (relative to other pairs). No basic unit of measurement is present in MDS, so the inter-item distance values are arbitrary. This makes it impossible to define “cutoff” values below which a pair is labeled as “very similar” (or conversely, above which a pair is labeled “very dissimilar”). In order to provide empirically-driven identification of item pairs that are tailored to each individual space, the followed procedure was employed.

First (for each MDS space), a vector of distances was obtained, corresponding to the three-dimensional Euclidean distance in psychological space between every pair of items. For each 17-item category, there were 136 inter-item distances; for each 16-item category, there were 120 inter-item distances. Next, the distances were rank-ordered, and categorized as “similar”, “moderate”, or “dissimilar”, based on a ternary-split. For example, the 45 pairs of teddy bears with the smallest inter-item distances were called “similar”. The 45 pairs with the largest inter-item distances were called “dissimilar”, and the 46 pairs in between were labeled “moderate” (in the forthcoming experiments, these labels were used to select pairs of stimuli for the manipulation of target template precision). See Figures A3 and A4, in the Appendix, for a sampling of item pairings:

Displayed are the three most similar item pairs, the three most dissimilar pairs, and three chosen from the center of the rankings. To appreciate this classification system, it may help to find these sample pairings in the MDS plots (shown in Figures 5 and 6).

When this system was applied to the teddy bear data, we found that similar pairs had distances ranging from .42 to .81 (*M* = .60). The moderate pairs had a range from .82 to 1.06 (*M* = .96). And the dissimilar pairs ranged from 1.07 to 1.34 (*M* = 1.21). For the butterflies, similar pairs had distances ranging from .30 to .82 (*M* = .61). Moderate pairs ranged from .83 to 1.10 (*M* = .96), and dissimilar pairs ranged from 1.11 to 1.38 (*M* = 1.22).

Experiment 2 Discussion

In Experiment 2, we collected a large amount of MDS data on a variety of item categories. Our goal was to acquire a set of psychological spaces, which could be used in selecting item pairs across a range of visual similarity. In the forthcoming experiments, this information was used in the service of selecting a pair of images that fit whatever degree of target-to-template similarity the experiment called for (on any given trial). All 240 image categories were used in each of the remaining experiments.

Experiment 3

The core theme of these experiments is that unlike typical laboratory searches, the items we locate in the real world are rarely, if ever, veridical representations of what we have in mind. When I look for my favorite coffee mug in a cluttered cabinet, it is impossible for me to know what orientation I will find it in, whether or not it will be occluded by other dishes, or whether its appearance will have changed in some way since the last time I saw it (e.g., a broken handle). In Experiment 3, we asked, "how does the inclusion of inaccurate features (in one's template) affect search behavior?" It is important to briefly reconsider the two specific hypothesized functions of the search template. First, the template may contribute to attentional guidance. That is, the template may affect the visual system's activation map (i.e., the mental "map" that is used to determine where next to direct one's eyes), by differentially weighting potential target locations as a function of their similarity to the stored representation in memory (Malcolm & Henderson, 2009). Second, the template may be used in target verification, as the criterion to which the current visual input is compared (Malcolm & Henderson, 2010).

As such, we entertained three possible hypotheses as to how peoples' behavior would change when searching with an imprecise target template. The first, dubbed the *attentional guidance* hypothesis, suggests that imprecise templates will hinder one's ability to quickly put their attention in the right place. Specifically, this idea states that the inclusion of incorrect features will cause people to direct their attention to regions of space that do not adequately match the target features, and thus their attentional guidance will be impeded. The second idea, which we refer to as the *decision-making* hypothesis, posits that an imprecise template will interfere with the comparison process whereby one decides if the current visual input matches their stored representation well enough to be identified as the target. Here, inaccurate features may lead to slowed perceptual decision-making, causing a delay in verifying the target. The final hypothesis, called the *dual-function* hypothesis, suggests simply that an imprecise target template will hinder both attentional guidance and decision-making during search. [2]

In Experiment 3a, we first established that template imprecision – induced through the use of imperfect target cues – would lead to decrements in search RTs. Then, in Experiment 3b, we honed in on this finding by using eye-tracking to deconstruct RTs into two functionally separate time periods. *Scanning*, which was the time from the start of search to the moment wherein the target was first fixated, and *decision-making*, which was the time that passed from first fixation on the target to the participant's overt response. Specifically, we chose two dependent measures to characterize searcher behavior during these time periods. *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 distance from central fixation, directly to the target. Thus, a SPR of one would indicate that the participant's eye moved straight from central fixation to the target, without deviating. Ratios greater than one indicate imperfect attentional guidance (i.e., that other locations were visited prior to the target being fixated). *Decision-times* (DTs) were obtained by calculating the time between when the target was first fixated to the time at which the participant pressed the spacebar to indicate that the target had been found.

In Experiments 3c and 3d, we provide converging evidence by manipulating the searchers' templates in analogous fashion, using a wider range of similarity (obtained through MDS ratings), rather than relying on a within- versus across-item classification system (i.e., state- and exemplar-pairs). Finally, in Experiments 3e and 3f, we supplement these findings with eye movement indexes.

Because these experiments were designed with several different factors (e.g., Set Size), a great many effects and interactions are possible. Accuracy is reported for diligence, but is not substantively interesting in these experiments, as it was quite high throughout. The specific findings we were interested in are any main effects of Template Precision (see below for an explanation), and any interactions of Template Precision with Block, which would indicate changes in behavior over time.

Experiment 3a Method

The apparatus and stimuli were identical to Experiments 1 and 2.

Participants.

Twenty new students from Arizona State University participated in Experiment 3a as partial fulfillment of a course requirement. All participants had normal or corrected-to-normal vision, and all reported normal color vision.

Design.

Three levels of Template Precision (precise, imprecise, inaccurate) were manipulated within-subjects. Within every condition, three levels of Set Size (12, 16, 20) were manipulated (in equal proportions). There were four blocks of 90 experimental trials (for a total of 360 trials). Additionally, four practice trials were administered during the instructions phase, to familiarize participants with the task. Two-hundred and forty experimental trials of the precise condition were administered, and sixty trials each of the imprecise and inaccurate conditions were given.

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" (see Figure 7). The participant was shown this cue for as long as they needed to encode it. When they were ready to begin, they pressed the spacebar to start the trial. Then, they saw a fixation cross for 500 msec, followed by the visual search display, which remained on screen until the participant responded or 10 seconds had elapsed (in which case the trial was marked as inaccurate). People rested their fingers on the spacebar during search. As soon as they found the target, they depressed the spacebar, which cleared the images from view (RTs were measured from the onset of the search display to the pressing of the spacebar). Each image was then replaced with a randomly generated number (between one and the set size) for two seconds. The numbers then disappeared and the participant was asked to indicate the number that appeared at the location of the target; they were shown the correct number and a randomly generated number (the correct number was presented equally often on the left or right of the screen). Using the keyboard, they then indicated the number corresponding to the target location ("f" for the number shown on the left; "j" for the number shown on the right).

This procedure is novel, and was used to obtain the cleanest measure of visual search completion time possible, while still ensuring that the participant actually located the correct image. Alternative methods, such has having participants click on the target image (using the mouse) add hand-movement time to the search RT; thus, "noise" would be included in the RT, that varies as a function of distance from the starting location of the mouse cursor to the position of the target. Feedback was provided on each trial in the form of a centrally presented green checkmark, or a large red “X”; feedback for correct trials was one second in duration, and was two seconds in duration for incorrect trials, to encourage accurate responding. Instructions asked participants to respond as quickly as possible while still retaining a high degree of accuracy.

Search array organization. A search array algorithm was used to create spatial configurations with pseudo-random organization (see Figure 8). An equal number of objects appeared in each quadrant of the display (three, four, or five, according to set size). Each quadrant was broken down into nine equally sized “cells” (effectively making the entire display into a 6x6 grid). On each trial, images were placed in random cells (per quadrant); specific locations were generated within these cells 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 that participants’ gaze would never immediately fall on a target at the start of a trial. The target appeared in each quadrant of the display an equal number of times.

Stimulus selection. On most trials (2/3 of the total), the target appeared in the search display exactly as it was cued (these are the “precise” Template Precision trials). On the minority of trials (1/3 of the total), the target that appeared in the display was a slight deviation from that which was shown as a cue prior to search. These are the “imprecise” and “inaccurate” Template Precision trials. The deviated target was either the state- or exemplar-partner of the picture shown as a cue (imprecise and inaccurate conditions, respectively). Participants were told that on the majority of trials, the target would appear exactly as it was cued to them, and that only on a minority of trials would they encounter a target item that deviated from the cue. They were not given explicit information about the proportion of precise to imprecise/inaccurate trials (but all three trial types were administered during the practice session), nor were they able to anticipate the condition of any given trial. Essentially, the goal was to encourage the participants to adopt the cue as a search template, by conducting an experiment wherein the cue was usually a "safe" template to adopt.

On each trial, the target was selected quasi-randomly from among the 100 exemplar- and 100 state-pairs (depending on which condition was being called for). Each of these pairs of stimuli were used 1 or 2 times throughout the entire experiment. On each trial, distractors were selected quasi-randomly from among the 240 categories (which were the focus of Experiment 2). Distractors were chosen such that no more than one exemplar per semantic category was represented on any given trial; across trials, no category was repeated until each had been used at least once. This set of images contained 4,040 distinct exemplars; each picture was used no more than twice over the entire experiment.

Experiment 3a Results

All data were analyzed using 3 (Template Precision: precise, imprecise, inaccurate) x 3 (Set Size: 12, 16, 20) x 4 Block (1-4) within-subjects, repeated measures ANOVAs. Only correct trial RTs were analyzed.

Accuracy.

Overall, accuracy was high (98%). We found a main effect of Template Precision, with the best performance on precise trials (99%), followed by imprecise (98%) and inaccurate trials (96%); $F(2, 18) = 7.94$, $p < .01$, $n^2_p = .45$. The main effects of Set Size and Block were not significant (Fs < 2). There was a Precision x Set Size interaction ($F(4, 16) = 5.00$, $p < .01$, $n^2_p = .55$), indicating that higher set sizes were detrimental to the imprecise condition, but not the other conditions (means are presented in Table 2). No other interactions were significant (Fs < 3).

Reaction time.

There was a main effect of Template Precision, with fastest RTs on precise trials (1006 msec), followed by imprecise (1321 msec) and inaccurate (1941 msec); $F(2, 18) = 77.75$, $p < .01$, $n^2_p = .90$. We also found main effects of Set Size (1242, 1424, and 1601 msec for 12, 16, and 20, respectively), $F(2, 18) = 24.05$, $p < .01$, $n^2_p = .73$, and Block (1579, 1473, 1360, and 1277 for Blocks 1-4), $F(3, 17) = 13.53$, $p < .01$, $n^2_p = .70$: RTs lengthened with increasing set size, and decreased across blocks. There was a Precision x Block interaction (see Figure 9; $F(6, 14) = 5.89$, $p < .01$, $n^2_p = .72$), indicating that searchers improved to the greatest degree when their template was less precise (improvements of 128, 245, and 533 msec across blocks for the precise, imprecise, and inaccurate conditions, respectively).

We also performed a secondary analysis to ensure that the peculiarities of the stimuli themselves did not lead to differential responding. For instance, perhaps the exemplar-pair stimuli are simply more difficult to process (e.g., due to visual complexity, ease of identification or naming), relative to the state-pairs. If so, this would taint our conclusion that template imprecision leads to increased RTs. On each precise trial, only a single image was shown (as the cue and as the target during search), but the individual pictures were a priori identified as members of the state- or exemplar-pair sets. Therefore, we conducted an analysis that looked only at the precise trials, to determine if RTs were different for state- and exemplar-pair stimuli. The analysis showed that RTs were equivalent across these groups of pictures (1025 and 987 msec for the state- and exemplar-pair stimuli, respectively), suggesting that there is nothing inherently more difficult about processing the exemplar-pair pictures, relative to the state-pairs.

Experiment 3a Discussion

The results from Experiment 3a were straightforward. We found that template imprecision caused a decrement in performance, such that people became slower searchers as their templates were made less precise. We also found an interaction of Template Precision and Block, suggesting that the main effect of Block was not simply a practice effect. That is, search RTs remained relatively flat in the precise condition, improved across trials in the imprecise condition, and improved to the greatest extent in the inaccurate condition. There are two possible explanations for this interaction. First, over time, participants may have become more resilient to the inaccurate features that were imposed upon them by imperfect target cues. They may have adopted a broader (perhaps categorical) template based on the cue, rather than sticking to the specific features shown to them prior to search. This could, for instance, lead to faster decision-making when they viewed the actual target in the search display. The second, less interesting possibility is simply that of a “ceiling effect”. Search RTs were very fast overall, dipping to approximately one second in the fourth block for the precise condition. Therefore, it seems that a more parsimonious explanation of this interaction is simply that performance in the precise condition reached its peak, whereas the comparatively harder conditions afforded greater room to improve.

In Experiment 3b, we followed up on this question in two ways. First, the introduction of eye-tracking allowed us greater insight into people's behavior, by allowing us to deconstruct RTs into scanning and decision-making behavior. If this interaction is not the result of a ceiling effect, we should expect to see an increase in decision-making efficiency without a concurrent increase in scanning efficiency. It is unlikely that a broader template would increase the ability of a searcher to guide their attention to target relevant features, but it seems reasonable that a more tolerant template would allow faster resolution of search decisions once the target has been viewed.

The second way in which we addressed this interaction was to introduce a new manipulation in Experiment 3b. Here, we manipulated (across participants) the proportion of trials in which the target was an exact match to the cue. Arguably, an ideal observer would construct a template that respected the cue only to the degree in which they tended to be a reliable representation of the to-be-located items. If the cues were perfectly reliable, a searcher should construct a template on each trial that strongly resembles the cue they are shown. However, if the cues are rarely a good depiction of the actual targets that appeared in the search displays, then an ideal searcher would only loosely base his or her template upon those items. Indeed, a recent study by Machizawa and colleagues (2012) suggests that the precision of working memory representations can be controlled at will (this study is discussed in further detail in the General Discussion of Experiment 3). As such, we introduced a *Match Frequency* manipulation in Experiment

3b (this condition is explained in greater detail in the Methods section). Across conditions, we varied the likelihood that the cue would perfectly match the to-be-located target, with the expectation that behavior in the various conditions would be qualified by the trustworthiness of the cues. Specifically, for the high Match Frequency group (i.e., the people who encountered very trustworthy cues), we expected faster search RTs in the precise Template Precision condition, relative to the low Match Frequency group (i.e., the people who encountered very *un*trustworthy cues). Then, in the inaccurate Template Precision condition, we expected a “flip”, whereby the high Match Frequency group would be slower than the low Match Frequency group, due to their strict adherence to the initial cue.

Experiment 3b Method

The stimuli were identical to Experiment 3a.

Participants.

Twenty-nine new students from Arizona State University participated in Experiment 3b as partial fulfillment of a course requirement. All participants had normal or corrected-to-normal vision, and all reported normal color vision. There were 10, 10, and 9 participants in the low, medium, and high *Match Frequency* groups, respectively.

Apparatus.

Data were collected using a Dell Optiplex 755 PC ((2.66 GHz, 3.25 GB RAM). Our display was a 21-inch NEC FE21111 CRT monitor, with resolution set to 1280x1024, and refresh rate of 60 Hz. E-Prime v2.0 Professional software (Schneider, Eschman, & Zuccolotto, 2002) was used to control stimulus presentation and collect responses. Eye-movements were recorded by 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°/s (or acceleration reached $8000°/s^2$). Viewing was binocular, but only the left eye was recorded.

Design.

The design was identical to Experiment 3a, with two exceptions. First, the frequency of precise Template Precision trials (i.e., trials in which the target cue and the to-be-located item were identical) was manipulated between-subjects. There were three levels of *Match Frequency* (low, medium, high). For the low Match Frequency group, 20% of the trials were of the precise Template Precision condition (the other 80% of trials were evenly split between imprecise and inaccurate Precision. For the medium Match Frequency group, 53% of trials were precise (the other 47% of trials were evenly split between imprecise and inaccurate), and for the high Match Frequency group, 80% of trials were precise (and the remaining 20% were evenly split between imprecise and inaccurate). The second change was that there were now only three blocks of 90 experimental trials (for a total of 270 trials).

Procedure.

The procedure was identical to Experiment 3a, with the exception of details pertaining to the addition of eye-tracking. Participants used a chin-rest during all search trials, and were initially calibrated to ensure accurate tracking. The chin rest was adjusted so that the position of the eyes was maintained centrally on the computer screen when the participant looked straight ahead. The calibration procedure establishes 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 fixated a 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 mec in order for the search display to appear. If they did not do this within 10 seconds, due to human error or problems with the calibration, the trial was marked as incorrect, and a recalibration was performed before the start of the next trial.

Experiment 3b Results

All data were analyzed using 3 (Template Precision: precise, imprecise, inaccurate) x 3 (Match Frequency: low, medium, high) x 3 (Set Size: 12, 16, 20) x 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 trial RTs, SPRs and DTs were analyzed. SPRs and DTs were not analyzed for trials in which the target was not directly fixated.

Accuracy.

Overall, accuracy was high (98%). There was a marginal effect of Template Precision, with the best performance on precise trials (99%), followed by imprecise (98%), and inaccurate trials (97%); $F(2, 25) = 3.05$, $p = .07$, $n^2_p = .20$. None of the other main effects (Match Frequency, Set Size, Block) were significant (*F*s < 2). There was a significant Precision x Match Frequency x Set Size interaction; $F(8, 46) = 2.87$, $p < .05$, $n^2_p = .33$, but no other interactions were significant (*F*s < 2). Means are presented in Table 3.

Reaction time.

We found a main effect of Template Precision, with fastest RTs on precise trials (1146 msec), followed by imprecise (1390 msec) and inaccurate (1889 msec); $F(2, 25) = 79.55$, $p < .01$, $n^2_p = .86$. There was no main effect of Match Frequency ($F < 1$). There were also main effects of Set Size (1339, 1449, and 1637 msec, for 12, 16, and 20, respectively), $F(2, 25) = 23.84$, $p < .01$, $n^2_p = .66$, and Block (1581, 1531, and 1313 msec for Blocks 1-3), $F(2, 25) = 28.49$, $p < .01$, $n^2_p = .70$: RTs lengthened with increasing set size and decreased across blocks. There was a Template Precision x Block interaction (see Figure 10; $F(4, 23) = 7.90$, $p < .01$, $n^2_p = .58$), indicating that searchers improved to the greatest degree when their template was less precise (improvements of 78, 294, and 433 msec across blocks for the precise, imprecise, and inaccurate conditions, respectively). See Figure 10 for the mean RTs, SPRs, and DTs. No other interactions were significant (*F*s < 2).

Scan-path ratios.

There was a main effect of Template Precision, with most optimal SPRs on precise trials (1.63), followed by imprecise (1.79) and inaccurate (2.50); $F(2, 25) = 66.98$, $p < .01$, $n^2_p = .84$. There was no main effect of Match Frequency ($F < 1$). There were also main effects of Set Size (1.70, 1.96, and 2.26, for 12, 16, and 20, respectively), $F(2, 25) = 25.73$, $p < .01$, $n^2_p = .67$, and Block (2.01, 2.03, and 1.88 for Blocks 1-3), *F*(2,

25) = 3.42, $p < .05$, $n^2_p = .22$: SPRs increased with larger set size and decreased across blocks. There was a Precision x Block interaction (see Figure 10; $F(4, 23) = 3.59$, $p < .05$, $n^2_p = .38$), indicating that searchers improved across blocks when their template was imprecise or inaccurate (improvements of .23 and .20, respectively), but not when their template was precise (decrement of .05). There was also a Set Size x Block interaction, $F(4, 23) = 4.45$, $p < .01$, $n^2_p = .44$, and a Template Precision x Match Frequency x Block interaction, $F(8, 46) = 2.21$, $p < .05$, $n^2_p = .28$. These interactions are plotted in the Appendix (Figures A5 and A6, respectively). No other interactions were significant (*F*s < 2).

Decision time.

We found a main effect of Template Precision, with fastest DTs on precise trials (443 msec), followed by imprecise (621 msec) and inaccurate (869 msec); $F(2, 25) = 47.04$, $p < .01$, $n^2_p = .79$. There was no main effect of Match Frequency or Set Size ($F < 1$). We found a main effect of Block, $F(2, 25) = 22.50$, $p < .01$, $n^2_p = .64$, with DTs decreasing over blocks (734, 665, and 534 msec, for Blocks 1-3). There was a Precision x Block interaction (see Figure 10; $F(4, 23) = 5.07$, $p < .01$, $n^2_p = .47$), indicating that searchers improved across blocks the most when their template was less precise (improvements of 98, 182, and 321 msec, for precise, imprecise, and inaccurate, respectively). No other interactions were significant (*F*s < 2).

Experiment 3b Discussion

The findings from Experiment 3b supported those of Experiment 3a. We replicated the finding that search RTs were slowed by the imposition of imprecise target templates, as well as the Template Precision by Block interaction. We hypothesized that this interaction may be due to an increased ability to accommodate imperfect target cues over time (rather than a generalized practice effect, combined with peak performance at the easier conditions). However, our eye-tracking analyses do not support this contention. If our searchers adopted a broader, more tolerant template, we would expect to find that their decision-making behavior improved over the course of the experiment, but that their scanning behavior remained constant, or worsened. In fact, both aspects of search behavior improved significantly more among the harder conditions. This suggests that the Template Precision by Block interactions we found are driven largely by our participants having reached peak levels of performance in the easiest condition. Moreover, we did not find any interactions of Block with Mismatch Frequency, which also suggests that participants were not willfully changing the nature of their search templates.

In fact, the only effect of Mismatch Frequency we found was a Template Precision by Match Frequency by Block interaction. As shown in the Appendix (Figure A6), however, this interaction was not systematic. This finding is surprising, and suggests that even with explicit instructions regarding the trustworthiness of the cue, our participants did not alter the manner in which they created a search template. Rather, this seems to indicate that they universally adopted the cue as their template and adapted to the challenges imposed by its inaccuracy as necessary. In Experiment 3d, we follow up on this question by introducing a condition wherein the cue was never a reliable indicator of the to-be-located target. For instance, if a participant saw a target cue of a particular coffee mug prior to search, that person knew that they would be searching for the category “coffee mug”, but that they would not find that exact picture.

To return to our attentional guidance, decision-making, and dual-function hypotheses: 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 on decision-times, with patterns that mirror those of the overall search RTs. We conclude that the presence of inaccurate features to one's target template hinders not just their ability to put their attention in the correct location, but to verify the identity of the target once it is viewed. It is also interesting to note that decision-making time accounts for nearly half of the overall RT, highlighting the importance of deconstructing RTs into the two processes that result in any successful search decision: Scanning, and deciding.

In Experiments 3c and 3d, we provide converging evidence for these findings through the use of a new method of defining template-to-target similarity. Here, rather than define similarity based on a within- versus across-item classification system, we employed numerous exemplars of each category, and varied pairwise similarity by assessing the distance between any two items in MDS space.

Experiments 3c and 3d Method

The stimuli and apparatus of Experiments 3c and 3d were identical to Experiment 3a.

Participants.

Thirty and sixty new students from Arizona State University participated in Experiments 3c and 3d, respectively, as partial fulfillment of a course requirement. All participants had normal or corrected-to-normal vision, and all reported normal color vision.

Design.

The design of Experiment 3c was identical to Experiment 3a, save the manner in which Template Precision was manipulated. Here, there were four levels of Template Precision (precise, similar, moderate, dissimilar). There were four blocks of 60 experimental trials, with 15 trials per Template Precision condition, presented in random order (for a total of 240 trials). The design of Experiment 3d was identical to Experiment 3c, except that the precise Template Precision condition was removed. There were four blocks of 60 experimental trials, with 20 trials per Template Precision condition.

Procedure.

The procedure of Experiments 3c and 3d was identical to Experiment 3a, with the exception of the manner in which stimuli were selected to manipulate Template Precision. As before, *precise* Template Precision trials involved presenting a target in the search display that was unaltered, relative to when it appeared as a cue prior to search. On other trials, the target that appeared in the search display was a slight deviation from that which was shown as a cue. In contrast to Experiment 3a (which relied on the state- and exemplar-pair designation to estimate cue-to-target similarity), pairwise similarity in Experiments 3c and 3d was manipulated by selecting item pairs from the MDS spaces obtained in Experiment 2.

On each trial, a pair of images was selected quasi-randomly from among the 240 image categories for which we acquired MDS data. When the similar Template Precision condition was called for, a pair of images was selected from the chosen category whose distance from one another fit the *similar* designation (i.e., from the 1/3 of item pairs which were closest to one another in MDS space). One of the images was used as the search cue, and its partner was presented in the visual search display (designation as the cue or actual search target was randomized). When the moderate Template Precision condition was implemented, a pair of images was selected from among those with a *moderate* distance designation, and when the dissimilar Template Precision condition was employed, a pair of images was selected from among the *dissimilar* pairs. Distractors were chosen randomly from the other 239 categories, such that no more than one exemplar per semantic category was represented on any given trial. Target pairs were used only once over the entire experiment.

In Experiment 3d, the precise Template Precision condition was removed, and participants were instructed that the exact target cue would never show up in the search display. They were told that the cue was a “guide” for which to direct their attention to the appropriate target category during search.

Experiments 3c and 3d Results

For Experiment 3c, all data were analyzed using 4 (Template Precision: precise, similar, moderate, dissimilar) x 3 (Set Size: 12, 16, 20) x 4 (Block: 1-4) within-subjects, repeated measures ANOVAs. For Experiment 3d, data were analyzed in similar fashion, but the design was now a 3x3x4, due to removal of a condition from the Template Precision group. Only correct RTs were analyzed.

Accuracy.

Experiment 3c. Accuracy was high overall (97%). There were no main effects of Template Precision, Set Size, or Block (*F*s < 3), and there were no interactions (*F*s < 2).

Experiment 3d. Accuracy remained high, overall (96%). There were no main effects of Template Precision or Block (*F*s < 2). We found a main effect of Set Size, *F*(2,

58) = 6.43, $p < .01$, $n^2_p = .18$, with poorer accuracy at the largest set size (97%, 97%, and 96%, for 12, 16, and 20, respectively). Table 4 presents the mean accuracy per condition, for Experiments 3c and 3d.

Reaction time.

Experiment 3c. We found a main effect of Template Precision, $F(3, 27) = 72.27$, $p < .01$, $n^2_p = .89$, with slower RTs as precision decreased (1487, 1985, 2128, and 2221 msec for precise, similar, moderate, and dissimilar conditions, respectively). See Figure 11. We performed a further analysis to ensure that this effect was not driven exclusively by the difference between trials wherein the cue was a perfect match (i.e., precise trials) relative to those with cue-to-target mismatch. These planned-comparisons revealed a significant difference between the similar and moderate conditions ($p < .05$) and between the similar and dissimilar conditions ($p < .01$), indicating that the main effect of Template Precision was not due solely to the precise condition. We also found a main effect of Set Size, $F(2, 28) = 42.29$, $p < .01$, $n^2_p = .75$, with slower RTs as set size increased (1753, 1946, and 2169 msec, for 12, 16, and 20, respectively). There was no main effect of Block ($F < 3$), nor any interactions (Fs < 2).

Experiment 3d. There was a main effect of Template Precision, $F(2, 58) = 5.97$, $p < .01$, $n^2_p = .17$, with slower RTs as precision decreased (2024, 2086, and 2153 msec for similar, moderate, and dissimilar conditions, respectively). We also found a main effect of Set Size, $F(2, 58) = 92.72$, $p < .01$, $n^2_p = .76$, with slower RTs as set size increased (1884, 2054, and 2326 msec for 12, 16, and 20, respectively). The main effect of Block was not significant ($F < 3$), nor were any interactions (Fs < 1).

Experiments 3c and 3d Discussion

In Experiments 3c and 3d, we again replicated the finding that imprecise templates hinder search times. Importantly, we now found that this effect was not limited to the manipulation of template precision through state- and exemplar-pairing systems, but rather that a psychologically valid metric of similarity (i.e., MDS ratings) produced a comparable continuum of performance. Moreover, the findings from Experiment 3d suggest that our participants were adopting the cues provided to them, and were not attempting to manipulate their target templates in the service of accommodating imperfect cues. That is, the people in Experiment 3d never encountered a perfectly accurate cue, and yet, their RTs still deviated as a function of the similarity between the cue and the to-be-located target (measured in MDS units). In Experiments 3e and 3f, we attempt to replicate this finding, and to examine our three hypotheses (attentional guidance, decision-making, and dual-function) by incorporating eye-tracking measures.

Experiments 3e and 3f Method

The stimuli were the same as in Experiments 3a-3d. The apparatus was the same as Experiment 3b. The procedure and design of Experiments 3e and 3f was identical to that of Experiments 3c and 3d, respectively, except that eye-tracking procedures were now added to the experiments.

Participants.

Eighteen new students (each) from Arizona State University participated in Experiments 3e and 3f as partial fulfillment of a course requirement. All participants had normal or corrected-to-normal vision, and all were subjected to the Ishihara colorblindness assessment (Ishihara, 1993), and were deemed to have normal color vision. [1]

Experiments 3e and 3f Results

For Experiments 3e and 3f, all data were analyzed exactly as in Experiments 3c and 3d, respectively. As before, only correct trial RTs, SPRs and DTs were analyzed, and SPRs and DTs were not analyzed for trials in which the target was not directly fixated.

Accuracy.

Experiment 3e. Accuracy was high overall (97%). We found a main effect of Template Precision, $F(3, 15) = 3.49$, $p < .05$, $n^2_p = .41$, with poorer accuracy as precision decreased (98%, 97%, 96%, and 96% for precise, similar, moderate, and dissimilar conditions, respectively). There was also a main effect of Set Size, $F(2, 16) = 5.31$, $p < .05$, $n^2_p = .40$, with poorer accuracy at the largest set size (97%, 97%, and 96% for 12, 16, and 20, respectively). There was no main effect of Block ($F < 1$), nor any interactions (Fs < 1).

Experiment 3f. Overall accuracy was high (97%). There was a main effect of Template Precision, $F(2, 15) = 3.98$, $p < .05$, $n^2_p = .35$, with highest accuracy for the moderate condition (98%), relative to the similar and dissimilar conditions (both 97%). The main effects of Set Size and Block were not significant (Fs < 4), and none of the interactions were significant (Fs < 3). Table 5 presents the mean accuracy per condition, for Experiments 3e and 3f.

Reaction time.

Experiment 3e. We found a main effect of Template Precision, $F(3, 15) = 73.82$, $p < .01$, $n^2_p = .94$, with slower RTs as precision decreased (1465, 2085, 2378, and 2447 msec for precise, similar, moderate, and dissimilar conditions, respectively). As before (in Experiment 3c), we performed planned-comparisons to ensure this main effect was not due exclusively to the precise condition. We found a significant difference between the similar and dissimilar conditions ($p < .01$), indicating that this was not the case. There was also a main effect of Set Size, $F(2, 16) = 17.14$, $p < .01$, $n^2_p = .68$, with slower RTs as set size increased (1885, 2104, and 2293 msec for 12, 16, and 20, respectively). The main effect of Block was not significant ($F < 1$), nor were any interactions (Fs < 1). See Figure 12 for the mean RTs, SPRs, and DTs from Experiments 3e and 3f.

Experiment 3f. There was no main effect of Template Precision ($F < 2$), but the trend was in the same direction as prior experiments. There was a main effect of Set Size, $F(2, 15) = 26.28$, $p < .01$, $n^2_p = .78$, with slower RTs at higher set sizes (2067, 2290, and 2476 msec for 12, 16, and 20, respectively). There was no main effect of Block ($F < 1$), nor any interactions (Fs < 4).

Scan-path ratios.

Experiment 3e. We found a main effect of Template Precision, $F(3, 15) = 68.63$, $p < .01$, $n^2_p = .93$, with higher SPRs as precision decreased (2.18, 2.97, 3.35, and 3.50 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; $p < .05$). There was a main effect of Set Size, *F*(2, 16) = 14.46, $p < .01$, $n^2_p = .64$, 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 significant ($F < 1$), nor were any interactions (*F*s < 1).

Experiment 3f. There was a marginal main effect of Template Precision, *F*(2, 15) = 3.57, $p = .05$, $n^2_p = .32$; however, the trend was not consistent with prior experiments, 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, $p < .01$, $n^2_p = .76$, 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 time.

Experiment 3e. There was a main effect of Template Precision, *F*(3, 15) = 22.49, $p < .01$, $n^2_p = .82$, with longer DTs as precision decreased (345, 644, 736, and 798 msec 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; $p < .05$) There was no main effect of Set Size or Block (*F*s < 1), nor were there any interactions (*F*s < 2).

Experiment 3f. The main effect of Template Precision was not significant ($F < 2$), although the trend was consistent with prior experiments (730, 746, and 819 msec 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, $p < .05$, $n^2_p = .55$, with the longest DTs in Block 3 (713, 711, 836, and 802 msec for Blocks 1-4). There was also a significant Set Size x Block interaction, *F*(6, 11) = 6.74, $p < .01$, $n^2_p = .79$, indicating that the effect of Set Size was inconsistent across blocks (see Figure A7, in the Appendix). No other interactions were significant (*F*s < 2).

Experiments 3e and 3f Discussion

What we found in Experiments 3e and 3f was not a perfect replication of our prior findings. In general, however, the data were largely consistent with the results of Experiments 3a – 3d. In Experiment 3e, we found that search RTs lengthened as a function of Template Precision, and that both scan-path ratios and decision-times followed suit. Moreover, we established (through planned-comparisons) that this effect was not driven only by the difference in performance from the precise Template Precision conditions (wherein the cue and the to-be-located item were the same picture), relative to the others (wherein the cue was different from the to-be-located target). Rather, this is a graded effect that increases in magnitude as the dissimilarity of the cue and target grows.

In Experiment 3f (wherein cues were never a perfectly representation of the target), we found consistent trends in the RT and decision-time data, whereby RTs and DTs slowed with less template precision, but neither of these findings were statistically reliable. For the scan-path ratio data, we found that SPRs did not change systematically as a function of template precision.

Experiment 3 General Discussion

The important findings from Experiment 3 can be summarized by the following three points: 1) Template imprecision, imposed through the use of imperfect target cues, causes a decrement in search performance that is inversely proportional to the similarity of the cue and the actual to-be-located target; 2) This finding holds true both when the actual target is the same exemplar as the cue but in a different perceptual state (i.e., our state-pairs), and when the cue and target are different exemplars altogether (i.e., our exemplar-pairs, and our MDS derived pairings); and 3) this decrement in search RT arises due to deficiencies in both attentional guidance and decision-making during search.

Target templates operate on low- and high-level aspects of visual search.

Our results are consistent with a large literature showing that the contents of VWM bias attention toward target-defining features (Olivers et al., 2011; Woodman et al., 2007; Dowd & Mitroff, *in press*; Soto et al., 2008). For instance, it has been shown that search template works at a functionally low level, acting in the service of gaze correction. Specifically, Hollingworth and colleagues (Hollingworth et al., 2009; 2010) found that small, corrective saccades tend to be directed towards features that match the searcher's template, often outside of the observer's conscious awareness. We make thousands of saccades each day, but these eye movements are ballistic in nature, and therefore are highly prone to error. When our eyes miss the intended location, we make rapid, corrective saccades to alight our eyes to the area of space for which we aimed (Becker, 1972; Deubel, Wolfe, & Hauske, 1982); these saccades are initiated quickly, almost reflexively (Kapoula & Robinson, 1986).

Using a *gaze correction* paradigm, Hollingworth and colleagues induced saccadic errors using gaze-contingent displays. Their participants fixated a central cross, and were shown a circular array of different colored patches. One of the color patches then grew in size (and shrank back rapidly), signaling to the searcher that they must make an eye movement to 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 intended saccadic target was moved while the searcher's eyes were "in flight". The arrays were rotated only a small amount, causing participants to fixate the middle region between the intended target and a distractor. At this point, the task can be envisioned as a small-scale visual search, whereby the searcher must make a corrective eye movement to the target, and avoid bringing their eye upon the distractor. The results demonstrated that gaze-correction was highly accurate and fast: Participants correctly moved their eyes to the target more than 90% of the time and often did so in under 200 msec (indicating that they did so without consciously realizing they made two distinct saccades; Hollingworth, Richard, & Luck, 2008). This suggests that the search template – in this instance, simply a target color – biases attention in such a way as to allow the visual system to more accurately inspect regions of interest.

At the other end of the complexity spectrum, it has been shown that the search template is used to direct attention in visually complex environments. Malcolm and Henderson (2010) showed people pictures of real-world scenes, and asked them to locate a target that was either shown to them prior to search (using a picture cue), or one that was categorically defined (using a word cue). They found that specific (i.e., picture-cued) templates allowed observers to more efficiently place their attention at the location of the target (indexed by scan-path ratios), and to more quickly verify the identity of the target once their attention was situated appropriately (indexed by post-fixation reaction times). These findings are entirely consistent with our own.

Template imprecision hinders attentional guidance and decision-making.

Our data are also consistent with a considerably smaller literature showing that imperfect target cues slow visual search. Vickery and colleagues (2005), for example, conducted several experiments wherein people were asked to search for polygon shapes or real-world objects. On some trials, the target cue was an imperfect representation of the to-be-located item. Specifically, they manipulated the size and the orientation of targets, relative to the cue, and found that deviations along either dimension lead to reduced search speed. Nevertheless, these imperfect pictorial cues still produced faster search performance than word cues (e.g., “Find a vertically symmetric polygon”, “Look for a tissue box”). This suggests that detailed visual information is incorporated into the mental representation of the target during search.

Similarly, Bravo and Farid (2009) had people search for a variety of tropical fish in scenes of coral reefs. Prior to conducting the search task, their participants were trained: They were shown a single exemplar from five different species, and learned to associate the name of the species with each particular fish. The visual search task involved simply determining whether a fish of any species was present in the picture or not (no more than one fish was ever present). Across experiments pictorial and word cues were used to indicate the type of fish that was likely to appear in the display. Importantly, the targets appeared in one of three conditions: 1) No variation – in which the target was identical to the studied image; 2) 2D viewpoint variation – in which the target was rotated, flipped, and scaled, relative to the studied image; and 3) Subordinate level variation – in which the target was a different fish from the same species. When picture cues were used, participants found the targets most quickly when it was untransformed, were slower when 2D viewpoint variations were implemented, and were slowest when another member of the species appeared. By contrast, when word cues were used (e.g., “Search for a rainbow fish”), search times were equally fast when the target was unvaried or given a 2D viewpoint variation. However, search times were significantly slower when a different member of the species appeared in the display. This strongly suggests that participants create search templates that are detailed and specific, yet tolerant to deviation.

In our investigation, we built upon these findings by using a new method of imposing template imprecision, and by tracking the eye movements of our participants during search. We entertained three hypotheses regarding the changes in behavior that were likely to appear when people conducted search using an imprecise guiding template. The first idea, which we referred to as the attentional guidance hypothesis, suggested that imprecise templates interfere with people’s ability to quickly place their attention in the region of space that contains the target. Our second idea, called the decision-making hypothesis, posited that imprecise templates hinder the comparison process whereby one decides that the current visual input is in fact the target. Our data strongly support the third idea, called the dual-function hypothesis, which conjectures that both aspects of visual search behavior will be impacted by template imprecision. By deconstructing the search RTs into two functionally distinct epochs – scanning time and decision-making time – we consistently found that the inclusion of inaccurate features in our searchers’ target templates resulted in suboptimal performance in attentional guidance and decision-making.

Importantly, our results go beyond the published literature because of the manner in which template-to-target similarity was manipulated in our investigations. The study by Vickery and colleagues (2005) used simple visual variations of the target (relative to the cue), and Bravo and Farid (2009) expanded this technique to involve new exemplars of the same category as the target. Our method goes one step further by classifying the similarity of cross-exemplar pairings through the use of MDS, showing that when the cue and the target are distinct exemplars, there exists a graded decrement in performance that is proportional to the psychological (dis)similarity of the pair.

Can templates be flexibly controlled?

The usefulness of precise target templates is uncontroversial. Clearly, specific information about the appearance of the target can be used to guide attention, and to verify that incoming visual input matches (or does not match) that which the searcher seeks out. But, as Bravo and Farid (2012) point out, an effective template must do two things: It must differentiate the target from potential distractors, but it must also be tolerant to variability in the appearance of the target. After all, objects in the real-world frequently change appearance over time, and we can almost never perfectly predict exactly how an item will appear when we find it. This certainty led us to ask whether or not the qualities of the search template are under voluntary control.

Indeed, it seems that people can willfully manipulate the precision with which information is stored in VWM. Machizawa et al. (2012) used an *orientation-discrimination* paradigm in which people were prompted to anticipate a fine- or coarse-level discrimination following a delay interval. Their participants were shown a sample display that consisted of several lines, oriented at different angles. Then, the display was removed for more than one second, and thereafter was replaced by a test probe. This test probe looked exactly like the sample display, but one of the bars changed color and was rotated 15 or 45 degrees (fine- and coarse-level discriminations, respectively). The task was to indicate if the bar rotated clockwise or counterclockwise. Unbeknownst to the 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 a fine-level discrimination, relative to trials in which they were prompted to anticipate a coarse-level discrimination. This suggests that participants used the prompt to adjust the precision with which information was stored in VWM. Simply put, it seems that the quality of information storage in VWM is not all-or-none, but can be manipulated flexibly, depending on task demands.

A more recent study by Bravo and Farid (2012), examined the extent to which observers could hone their target template to meet variable task demands. Participants again looked for tropical fish in underwater coral reef scenes. There were two groups of participants: Half of the people searched for the same target image over and over again, whereas the other group of participants searched for multiple exemplars of the same species. Everyone came back one or two days after the initial training session, and conducted a second round of testing that included new exemplars from the target category. Search times to find these previously unseen pictures were faster by those who were trained on multiple exemplars, relative to those who were trained on a single category image. This indicates that being trained on several instances of a category allowed the searchers to adopt a more general template that was tolerant to deviations from that which was studied, thereby allowing them to efficiently locate items with similar features (even though those pictures had not been seen before).

In Experiment 3, we addressed this same question in two ways. First, in Experiment 3b, we introduced a Match Frequency manipulation, whereby participants were told (accurately) how often the target would appear exactly as it was cued. Our hypothesis was that participants who received frequently accurate cues would adopt a template that closely matched that of the cue. By contrast, those who received infrequently accurate cues should adopt a template that was broader, and more tolerant to deviance from the cue on any given trial. Unfortunately, we found no evidence that these strategies were in fact implemented: Our manipulation had no substantive effect on RT or eye movement indexes. Our second way of approaching this question was to involve, in Experiments 3d and 3f, cues that were never an accurate representation of the target. Here, we hoped that our participants would be encouraged to adopt a generalized template, one that was representative of the category but that didn’t necessarily resemble the cue. The results of these experiments were less straightforward, but in general, people’s performance tended to fall off as the similarity of the target (relative to the cue) decreased.

It is tempting to suggest that our findings argue against the notion that people can flexibly control the quality of their search template. An ideal searcher in Experiment 3b would construct a template that matched the cue only to the extent that those cues tended to be reliable. In Experiments 3d and 3f, the cue only indicated the category that was to be searched for, and people never found that exact picture. Therefore, they should construct a template that resembles a prototypical category member, rather than the cue itself. Neither of these “model searcher” scenarios arose, which may suggest that people cannot behave in such an ideal way. However, upon closer consideration, it seems clear that our participants simply chose to adopt the cue as their template because that was the easiest thing to do.

Simply put, it is likely that our manipulations were not strong enough to induce a change in the way searchers constructed their templates. It seems reasonable to suggest that under both of the scenarios discussed above, creating a prototypical template was more effortful than simply calling to mind a representation of the picture cue that was shown. Savvy research participants likely adopted our cues as their templates, and chose simply to deal with changes in the appearance of the target as necessary. Thus, for now, our data cannot speak strongly to the question of whether or not target templates can be flexibly controlled in the service of ideal search behavior.

Experiment 4

In Experiment 4, we investigated a second way in which target templates may be imprecise. Previously (in Experiment 3), our participants were shown imperfect target cues that created in their minds a template that contained inaccurate details regarding the appearance of the target. Sometimes, however, our templates may be imprecise not because they are wrong, per se, but because they are too far-reaching. Imagine, for instance, that you are searching for a particular brand of cereal at the supermarket. You know that this brand just changed their logo, but you're unsure if this particular grocery store will have boxes with the old logo or the new one. If we define template precision as the similarity between your mental representation of the target and its actual appearance in the environment, then looking for both instances of the cereal box will make your target template imprecise, because you will only find one of them. The features of the new logo, for instance, are superfluous if the store carries only boxes that bear the old one.

It now becomes necessary to requantify precision as the similarity between multiple items stored in VWM: We'll call this the *feature width* of the searcher's template. Consider a scenario in which you are searching for one of two items that are highly similar to one another: For example, the new cereal logo may be identical to the old one, except that the brand name is in a different font. Contrast this with a situation in which the two potential targets are highly dissimilar. Now, perhaps, the new cereal logo includes a different font, a newly colored background, and a fresh mascot. In both of these situations, your target template contains all of the "right" features, but in the second scenario your template also contains several features that will not be helpful in locating the desired box.

Thus, in Experiment 4, we manipulated target precision by having people search for two targets at a time (only one of which ever appeared in the displays), with varying degrees of similarity to one another. In Experiment 4a, we first established that this technique for influencing template precision would lead to decrements in search RTs. Then, in Experiment 4b, we supplemented our data by tracking the eye movements of participants during search. In Experiment 4c we provided converging evidence by using MDS ratings to expand the range of pairwise similarity, and in Experiment 4d, we added eye tracking analyses to this technique. All of the data were analyzed in the same manner as Experiment 3, whereby search RTs were deconstructed into eye-tracking indexes of scanning and decision-making behavior. As in Experiment 3, the design of these experiments affords a great many effects and interactions. Accuracy is again reported for diligence, but is not centrally interesting. Specifically, in Experiment 4, we are particularly interested in any main effects of Feature Width and any interactions with that factor.

The central theme of Experiment 4 is that laboratory search tasks rarely provide the searcher with a cue that contains extraneous information about the appearance of the target. In the real-world, however, we often encounter conditions of uncertainty, whereby we are unable to predict the features of a stimulus that are most likely to be beneficial to us. Thus, we may be inclined to construct a search template with a wide range of possible features, even though some of those features may be unnecessary or unhelpful. Thus, in this experiment we asked simply, “how does the inclusion of inessential features (in one’s template) affect search behavior?” To answer this question, we return again to the two hypothesized functions of the target template: That it is used in attentional guidance, and in perceptual decision-making. Although the results of Experiment 3 argue strongly in favor of a dual-function hypothesis, there is no *a priori* reason to assume that this new manipulation will produce the same results. That is, it is possible that the inclusion of extra features in the searcher’s template will hinder their ability to direct their attention to regions of space that contain target-similar features, but that the comparison process (of deciding whether or not the current visual input matches the target) will be unimpeded. It seems likely that these extraneous factors will hinder attentional guidance because they may cause the searcher to be attracted to regions of space that do not actually match to to-be-located item. By contrast, if one assumes, for instance, that a Sternberg-like comparison process (Sternberg, 1966; 1969; 1975) is completed upon viewing of each new item, then the similarity of the two potential targets in VWM should have no effect on decision-making times.

Experiment 4a Method

The apparatus and stimuli were identical to Experiments 1-3 (specifically, the non-eyetracking experiments).

Participants.

Twenty-six new students from Arizona State University participated in Experiment 4a as partial fulfillment of a course requirement. All participants had normal or corrected-to-normal vision, and all reported normal color vision.

Design.

Three levels of Feature Width (precise, narrow, wider) were manipulated within-subjects. As before, within every condition, three levels of Set Size (12, 16, 20) were manipulated (in equal proportions). There were four blocks of 96 trials (for a total of 384 trials). There were four practice trials administered during the instructions phase. Half of the experimental trials (192) were single-target trials (*precise* Feature Width), and half were two-target trials. Of the two-target trials, half were of *narrow* Feature Width, and half were *wider* Template Precision.

Procedure.

Visual search procedure. The procedure for single-target trials was identical to prior experiments. On 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 of the images would appear in the search display, and that they could therefore stop searching after a single target had been found. The actual search target was presented (during the cue phase) on the left and right of the screen equally often. Visual search arrays were constructed of random spatial configurations, as in prior experiments.

Stimulus selection. On each trial, a pair of images was selected quasi-randomly from among the 100 exemplar- and 100 state-pair stimuli. In the *precise* Feature Width condition (i.e., single-target trials), only one item from the chosen pair was selected (randomly) to be used as the cue (prior to search) and the target that appeared in the visual search display. Within these single-target trials, stimuli were selected from the exemplar- and state-pair "pools" equally often. The *narrow* and *wider* Feature Width conditions were both two-target trials. For narrow trials, a pair of images was selected from the state-pair stimuli, and both pictures were shown as a cue prior to search; on wider trials, a pair of images was selected from the exemplar-pair stimuli for use as a cue. See Figure 13 for an example of target cues used across the different levels of Feature Width (note that the *widest* condition was not included in Experiment 4a, but was added to later experiments). As in prior experiments, distractors were chosen quasi-randomly from the 240 picture categories, such that no more than one exemplar per semantic category was represented on any given trial. Target stimuli were not used more than twice throughout the entire experiment, nor were any distractors.

Experiment 4a Results

All data were analyzed using 3 (Feature Width: precise, narrow, wider) x 3 (Set Size: 12, 16, 20) x Block (1-4) within-subjects, repeated measures ANOVAs. Only correct trial RTs were analyzed. One participant was excluded from data analysis for producing mean accuracy and RTs that were more than 2.5 standard deviations below the mean of the group.

Accuracy.

Accuracy was high overall (98%). We found a main effect of Feature Width, $F(2, 23) = 6.27$, $p < .01$, $n^2_p = .35$, with poorer accuracy at the wider condition, relative to the other conditions (98%, 98%, and 97% for precise, narrow, and wider, respectively). There was no main effect of Set Size or Block (Fs < 3), nor any interactions (Fs < 2). All means are presented in Table 6.

Reaction time.

There was a main effect of Feature Width, $F(2, 23) = 18.89$, $p < .01$, $n^2_p = .62$, with slower RTs as the feature space widened (909, 958, and 1053 msec for precise, narrow, and wider, respectively). We also found main effects of Set Size, $F(2, 23) = 82.63$, $p < .01$, $n^2_p = .88$, and Block, $F(3, 22) = 4.00$, $p < .05$, $n^2_p = .35$. RTs lengthened as set size increased (855, 975, and 1090 msec for 12, 16, and 20, respectively) and shortened over blocks (1037, 982, 958, and 917 msec for Blocks 1-4). See Figure 14. There were no interactions (*F*s < 2).

Experiment 4a Discussion

The results of Experiment 4a were straightforward: We found that template feature width caused a decrement in performance, whereby people became slower searchers when their templates included extraneous features. In Experiment 4b, we followed up on this finding by honing in on the two processes that contribute to the search RT, through the use of eye-tracking. We also added a fourth condition to the Feature Width manipulation. In the precise, narrow, and wider Feature Width conditions, our searchers were looking for a single category (that may or may not have been represented by a pair of exemplars). Now, in the *widest* condition, we had people search for two semantic categories simultaneously, in order to provide a further comparison: Search for a single category (of varying feature width) to search for two distinct categories.

Experiment 4b Method

The apparatus and stimuli were identical to Experiments 1-3 (specifically, the eyetracking experiments).

Participants.

Sixteen new students from Arizona State University participated in Experiment 4b as partial fulfillment of a course requirement. All participants had normal or corrected-to-normal vision, and all reported normal color vision.

Design.

The design was identical to Experiment 4a, with two exceptions. First, a fourth condition was added to the Feature Width manipulation. In this *widest* condition, two images were shown as cues prior to search, but these items were not semantically matched (see Figure 13). The second change was that there were now only three blocks of 72 trials (for a total of 216 trials).

Procedure.

The procedure was identical to Experiment 4a, except that eye-tracking procedures were now added to the experiment, in addition to the fourth Feature Width condition. In the *widest* Feature Width condition, two images were selected quasi-randomly, to be used as cues prior to search (one item from the pair was randomly selected to be the actual search target). Stimuli for this condition were pulled from the state- and exemplar-pairs in equal proportions. Distractors were quasi-randomly selected from the 240 item categories, as in prior experiments; target and distractor stimuli were not used more than twice throughout the entire experiment.

Experiment 4b Results

All data were analyzed using 4 (Feature Width: precise, narrow, wider, widest) x 3 (Set Size: 12, 16, 20) x 3 (Block: 1-3) within-subjects, repeated measures ANOVAs.

Only correct trial RTs, SPRs, and DTs were analyzed; SPRs and DTs were only analyzed on trials in which the target was directly fixated.

Accuracy.

Overall, accuracy was high (99%). There was a marginal main effect of Feature Width, $F(3, 13) = 3.33$, $p = .05$, $n^2_p = .43$, with poorer accuracy in the widest condition (97% for the widest condition, and 99% for each of the other conditions). There was no main effect of Set Size or Block (*F*s < 3), and no interactions (*F*s < 3). Table 7 presents the means for each condition.

Reaction time.

We found a main effect of Feature Width, $F(3, 13) = 51.76$, $p < .01$, $n^2_p = .92$, with longer RTs as feature space widened (958, 1016, 1119, and 1563 msec for precise, narrow, wider, and widest, respectively). We also performed a further analysis to ensure that this main effect was not driven exclusively by the condition wherein people searched for two items that were not semantically matched (i.e., the widest condition). Planned-comparisons revealed a significant difference between the precise and wider conditions ($p < .05$), indicating that this was not the case. See Figure 15 for mean RTs, SPRs, and DTs. There were also main effects of Set Size, $F(2, 14) = 46.32$, $p < .01$, $n^2_p = .87$, and Block, $F(2, 14) = 3.90$, $p < .05$, $n^2_p = .36$: RTs lengthened at higher set sizes (1009, 1154, and 1329 msec for 12, 16, and 20, respectively), and were unequal across blocks (1215, 1134, and 1143 msec for Blocks 1-3). We found a significant Feature Width x Set Size interaction, $F(6, 10) = 4.64$, $p < .05$, $n^2_p = .74$ indicating that increasing set size had a larger effect when people's feature space was wider. The slope of the best-fitting line relating RT to Set Size is a measure of the cost associated with adding each new item to the display: Slopes were 15, 29, 44, and 72 msec/item for precise, narrow, wider, and widest conditions, respectively (see Figure A8 in the Appendix).

Scan-path ratios.

There was a main effect of Feature Width, $F(3, 13) = 68.67$, $p < .01$, $n^2_p = .94$, 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 revealed that this effect was likely not driven exclusively by the widest condition; there was a marginally significant difference between the precise and wider conditions ($p = .07$). We also found main effects of Set Size, $F(2, 14) = 30.63$, $p < .01$, $n^2_p = .81$, and Block, $F(2, 14) = 8.28$, $p < .01$, $n^2_p = .54$. 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 (Fs < 2).

Decision time.

We found a main effect of Feature Width, $F(3, 13) = 11.78$, $p < .01$, $n^2_p = .73$, with longer DTs as feature space was widened (362, 415, 413, and 521 msec 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 the other conditions; no other pairwise comparisons revealed significant differences. There was also a main effect of Block, $F(2, 14) = 10.82$, $p < .01$, $n^2_p = .61$, indicating that people became faster decision-makers over the course of the experiment (478, 414, and 391 msec for Blocks 1-3). There was no main effect of Set Size ($F < 2$). There was a significant Feature Width x Set Size x Block interaction, $F(12, 4) = 27.72$, $p < .01$, $n^2_p =$

.99 (this interaction is plotted in Figure A9 in the Appendix). No other interactions were significant (*F*s < 2).

Experiment 4b Discussion

The findings from Experiment 4b again showed that search RTs were slowed by target templates that were unnecessarily wide, and that this effect was not simply driven by the inclusion of the condition whereby people looked for two semantically different targets. That is, planned-comparisons revealed a significant difference in performance within the single-category conditions alone. The eye-tracking analyses revealed that widened target templates lead to a decrease in efficient attentional guidance and a decrement in decision-making behavior. However, the scan-path ratio data showed a significant effect of Feature Width within the single-category conditions alone, whereas the decision-time data only revealed an effect of Feature Width when the two-category condition was included in the analysis. This suggests that attentional guidance is affected more strongly by widening the feature space of the template than is decision-making ability.

In Experiment 4c, we provide converging evidence by manipulating the feature width of the template through the use of MDS-derived item pairs. Here, participants always searched for a single semantic category, but the width of the features of that category were varied: We compared search for a single category exemplar to search for a pair of exemplars with varying degrees of similarity (relative to one another).

Experiment 4c Method

The apparatus and stimuli were identical to Experiments 1-3 (specifically, the non-eyetracking experiments).

Participants.

Sixty-five new students from Arizona State University participated in Experiment 4c as partial fulfillment of a course requirement. All participants had normal or corrected-to-normal vision, and all reported normal color vision.

Design.

The design of Experiment 4c was identical to Experiment 4a, with two exceptions. First, Feature Width was manipulated using the MDS spaces from Experiment 2. Now, there were four levels of Feature Width: *precise, similar, moderate,* and *dissimilar*. The second change was to the number of trials that were administered. There were four blocks of 80 experiment trials (for a total of 320 trials), with 20 trials per Feature Width condition.

Procedure.

The procedure of Experiment 4c was identical to Experiment 4a, save the manner in which stimuli were selected to manipulate Feature Width. As before, *precise* Feature Width trials involved only a single image that was used as a cue and the actual search target. On other trials, two target cues were shown prior to search, but only one was present in the search display. Pairwise similarity was used to manipulate Feature Width; image pairs were selected from the MDS spaces obtained in Experiment 2. A pair of images was selected quasi-randomly from the 240 image categories. When the *similar* Feature Width condition was called for, a pair of images was selected from the chosen category whose distance from one another fit the *similar* designation (and so on for the *moderate* and *dissimilar* conditions).

Experiment 4c Results

All data were analyzed using 4 (Feature Width: precise, similar, moderate, dissimilar) x 3 (Set Size: 12, 16, 20) x 4 (Block: 1-4) within-subjects, repeated measures ANOVAs. Only correct trial RTs were analyzed. Two participants were excluded from analysis for performing at a level of accuracy that was more than 2.5 standard deviations below the mean of the group.

Accuracy.

Overall, accuracy was high (98%). There was a main effect of Feature Width, $F(3, 60) = 3.34$, $p < .05$, $n^2_p = .15$, with poorer accuracy at the dissimilar condition (97%), relative to the other conditions (98% each). There was also a main effect of Block, $F(3, 60) = 7.68$, $p < .01$, $n^2_p = .28$, indicating slightly poorer accuracy over blocks (99%, 98%, 97%, and 97% for Blocks 1-4). There was no main effect of Set Size ($F < 1$), nor any interactions (Fs < 2). Table 6 presents all means.

Reaction time.

We found a main effect of Feature Width, $F(3, 60) = 81.89$, $p < .01$, $n^2_p = .80$, with slower RTs as feature space widened (1111, 1366, 1374, and 1427 msec for precise, similar, moderate, and dissimilar, respectively). We performed a follow-up analysis to ensure that this main effect was not driven exclusively by the difference between single-target (i.e., *precise* Feature Width) trials and two-target trials (i.e., each of the other conditions). Planned comparisons showed a significant difference between similar and dissimilar Feature Width conditions ($p < .05$), indicating that this was not the case. We also found main effects of Set Size, $F(2, 61) = 151.33$, $p < .01$, $n^2_p = .83$, and Block, $F(3, 60) = 3.24$, $p < .05$, $n^2_p = .14$. RTs slowed as set size grew (1178, 1326, and 1455 msec for 12, 16, and 20, respectively), and quickened over blocks (1401, 1320, 1277, and 1282 msec for Blocks 1-4). There was a Feature Width x Set Size interaction, $F(6, 57) = 2.50$, $p < .05$, $n^2_p = .21$, indicating steeper RT by Set Size slopes as feature space was widened: Slopes were 23, 36, 39, and 42 msec/item for precise, similar, moderate, and dissimilar conditions, respectively (see Appendix Figure A10). No other interactions were significant (*F*s < 2).

Experiment 4c Discussion

In Experiment 4c, we found converging evidence for the notion that unnecessary features in one's template inhibit the ability to quickly perform a visual search. Participants always searched for a single semantic category, but the width of their template was widened by having people search in conditions wherein the category was represented by a single exemplar, or a pair of exemplars with variable similarity to each other. Importantly, we found that the effect of Feature Width was not driven exclusively by the difference between single-exemplar and two-exemplar trials; rather, there was a significant, graded effect within the two-exemplar conditions alone. In Experiment 4d, we attempted to replicate this finding, and to elucidate the cause of the search RT decrement by tracking the eye movements of our participants.

Experiment 4d Method

The apparatus and stimuli were identical to Experiments 1-3 (specifically, the eyetracking experiments). The procedure was identical to Experiment 4c, except that eye-tracking procedures were now added to the experiment.

Participants.

Seventeen new students from Arizona State University participated in Experiment 4d as partial fulfillment of a course requirement. All participants had normal or corrected-to-normal vision, and all passed the Ishihara color-blindness assessment.

Design.

The design was identical to Experiment 4c, with one exception. Again, there were four blocks of experimental trials, but each block now only contained 60 trials (for a total of 240 trials), evenly split across the four Feature Width conditions.

Experiment 4d Results

All data were analyzed using 4 (Feature Width: precise, similar, moderate, dissimilar) x 3 (Set Size: 12, 16, 20) x 4 (Block: 1-4) within-subjects, repeated measures ANOVAs. Only correct trial RTs, SPRs, and DTs were analyzed, and SPRs and DTs were only calculated for trials in which the target was fixated directly.

Accuracy.

Accuracy was high overall (98%). There was no main effect of Feature Width or Set Size (Fs < 2). We found a main effect of Block, $F(3, 14) = 3.63$, $p < .05$, $n^2_p = .44$, with poorer accuracy in Block 3 relative to the other blocks (99%, 99%, 97%, and 98% for Blocks 1-4). There were no interactions (Fs < 3). Table 9 presents the means for each condition.

Reaction time.

There was a main effect of Feature Width, $F(3, 14) = 24.85$, $p < .01$, $n^2_p = .84$, with longer RTs as feature space widened (1193, 1507, 1544, and 1567 msec for precise, similar, moderate, and narrow, respectively). Planned-comparisons revealed that this effect was driven exclusively by the difference between single-target trials (i.e., *precise* Feature Width) relative to two-target trials (none of the other conditions were significantly different from one another). See Figure 17 for mean RTs, SPRs, and DTs. There was a main effect of Set Size, $F(2, 15) = 54.78$, $p < .01$, $n^2_p = .88$, with slower RTs at larger set sizes (1267, 1443, and 1649 msec for 12, 16, and 20, respectively). The main effect of Block was not significant ($F < 2$). We found a significant Feature Width x Block interaction, $F(9, 8) = 4.89$, $p < .05$, $n^2_p = .85$, indicating an unstable effect of Block across levels of Feature Width (see Appendix Figure A11 for a plot of this interaction). No other interactions were significant (*F*s < 1).

Scan-path ratios.

We found a main effect of Feature Width, $F(3, 14) = 8.31$, $p < .01$, $n^2_p = .64$, with longer SPRs for two-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 supported the contention that the main effect was driven exclusively by the difference between single-target and two-target trials (none of the two-target trials were significantly different from one another). There was also a main effect of Set Size, $F(2, 15) = 12.18$, $p < .01$, $n^2_p = .62$, with longer 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 significant ($F < 1$), nor were any interactions (*F*s < 3).

Decision time.

There was a main effect of Feature Width, $F(3, 14) = 9.81$, $p < .01$, $n^2_p = .68$, again indicating a difference in performance during single-target trials, relative to two-target trials (378, 499, 489, and 467 msec for precise, similar, moderate, and dissimilar conditions, respectively). Planned-comparisons showed that none of the two-target conditions were significantly different from one another. There was a main effect of Set Size, $F(2, 15) = 3.84$, $p < .05$, $n^2_p = .34$, with longer DTs at higher set sizes (393, 452, and 530 msec for 12, 16, and 20, respectively). The main effect of Block was not significant ($F < 3$), nor were any interactions (Fs < 2).

Experiment 4d Discussion

Our findings in Experiment 4d were an imperfect replication of the prior results. Specifically, in Experiment 4c we found a main effect of Feature Width that was not exclusively reliant on the difference in RT between single-target and two-target trials. By contrast, in Experiment 4d, the main effect of Feature Width was in fact driven exclusively by the difference between the single-target condition relative to the others. The eye-tracking analyses revealed the same pattern of results: For both the scan-path ratios and decision-times, we found main effects of Feature Width that were not reliable when the single-target condition was excluded from the analyses.

Experiment 4 General Discussion

The results of Experiment 4 can be summed up by the following three points: 1) Template imprecision, imposed by widening the “feature space” of our searchers’ templates, causes a slowing of search performance that seems to be inversely proportional to the similarity of the potential target images; 2) A clear performance change arises between single-target and two-target search conditions, and (within the two-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 slowed search

RTs are caused by decrements in both attentional guidance and decision-making behavior.

The cost of searching for more than one item.

We are adept visual searchers, probably stemming from the fact that we look for things in our environment so frequently. In fact, under most circumstances, search is conducted so effortlessly that we can actively search for more than one thing at a time. Before departing for work at the start of the day, you often must search for several belongings to take with you: Your wallet, keys, and backpack, for instance. Most people will stroll through their home and look for these items simultaneously. It seems almost silly to suggest that a person would first locate their keys, then allow themselves to search for their wallet, and upon finding the wallet, begin looking for their backpack. However, multiple-target search incurs performance costs, relative to single-target search. For example, in prior work, we had people search for one, two, or three potential targets (a maximum of one target appeared on any given trial and participants were informed of this regularity) in cluttered search displays and in rapid-serial visual presentation “streams” (wherein single items were centrally presented for a brief duration; Hout & Goldinger, 2010). We found that multiple-target search affects both search accuracy and search speed: Participants are more likely to miss and to false-alarm during multiple-target search, and both locating the target (i.e., target-present trials) and determining its absence (i.e., target-absent trials) are performed more slowly, relative to single-target search. Moreover, these costs are revealed by the eye movement record as well. In a second study, we found that when people look for multiple targets, they require more fixations to complete the task (Hout & Goldinger, 2012).

Arguably, it feels more natural to conduct one multiple-target search than to, for instance, conduct several consecutive searches for each item one must collect before departing their home. Menneer and colleagues (2007), however, found that this was not the most efficient strategy to adopt (see also Menneer et al., 2009; 2010). They compared search performance when people looked for two targets simultaneously to conditions wherein people conducted two single-target searches back-to-back. They found that dual-target search incurred accuracy and reaction time costs, whereby search for two targets took longer than the summed search times for two consecutive single-target searches. In our everyday lives, this may seem like a trivial fact. If you do not locate your keys or wallet on the first pass through your living room, you can simply return to reinspect the room a few moments later (or choose not to leave home). However, for professional searchers, such as airport baggage screeners, this dual-target cost may have more dire consequences, as the opportunity for reinspection is lost once the passenger's bag has cleared the security station (and the cost of missing a target weapon is not merely an inconvenience). Menneer and colleagues' work suggests that when people look for multiple items at once, the fidelity of the target representations cannot be maintained as accurately; thus, high-stakes searchers may be better served by a divided search strategy.

Our findings from Experiment 4 strongly support the notion of a dual-target cost in visual search. We consistently found that search performance was worse when people constructed a template from a pair of target cues, relative to when they were only given a single cue. Specifically, in Experiments 4a and 4b, performance was better in the precise Feature Width condition, relative to the narrow, wider and widest conditions, and in Experiments 4c and 4d, performance was better in the precise condition, relative to the similar, moderate, and dissimilar conditions. Furthermore, we found that when the two potential targets were from different semantic categories, performance declined, relative to when the two pictures were categorically alike: In Experiment 4b, performance was markedly worse in the widest Feature Space condition, relative to the others.

What is the nature of the search template in multiple-target search?

An open question remains regarding the nature of mental representations during multiple-target search: Do people use multiple, discrete representations, or do they merge a pair of cues into a single, broad target template? The answer, it seems, may depend on the demands of the task at hand. Visual working memory is not an unlimited capacity system (Cowan, 2001; Vogel et al., 2005), and several different theoretical accounts exist that attempt to explain the organization of VWM in different ways. Some theories propose that VWM contains a limited set of "slots" (Anderson et al., 2011; Awh et al., 2007) in which information can be stored. Others envision VWM as a dynamic resource that is limited by overall precision, rather than by number (Bays & Husain, 2008; Gorgoraptis et al., 2011). And yet other theories suggest that the answer lies 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 and colleagues (2011; see 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. The authors entertained two hypotheses regarding the manner in which dual-target templates could be created. First, it is possible that searchers maintained two target templates simultaneously (or alternated between them over time; see Moore & Osman, 1993). The alternative hypothesis was that a single target template was constructed that included a range of feature values from both targets, and possibly the feature values in between (i.e., those colors that occupied the “color space” between the two target colors). Stroud systematically manipulated the similarity of the two potential targets by varying how far apart these items were in “color-steps” (defined using CIExyY color space). When people searched for a single target, they exhibited a high degree of 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 the targets. However, the data were more nuanced: When the target colors were two steps away from one another, people tended to fixate the intervening colors more often than when the two targets were four steps away from one another. This suggests that when the targets are similar, they are encoded as a single, unitary representation that meshes together the colors of the targets, as well as those in between. When the targets are dissimilar, however, they seem to be encoded as separate and discrete representations that do not “absorb” the feature space that lies between them (but see Houtkamp & Roelfsema, 2009).

Can our *attentional-guidance* and *decision-making* hypotheses answer this query?

Following the findings of Stroud and colleagues, we may be inclined to suggest that when people search for two targets that are highly similar (as when searching for two boxes of cereal who differ only in font), they are represented by a meshed, unitary representation, but when people search for dissimilar targets (as when searching for two boxes of cereal that differ in font, color, and mascot), they use detached representations. In Experiment 4b, the scan-path ratio data showed a marginally significant difference between single-target and two-target search (i.e., between the precise and wider conditions), and a clear difference between the single-category and two-category conditions (i.e., between the widest condition, relative to all others). In short, attentional guidance got worse as searchers moved from a single-target template to a two-target (but single-category) template, and from there to a two-category template. In Experiment 4d, however, we found a main effect of Feature Width that was solely dependent on the difference between single-target and two-target search (i.e., between the precise condition, relative to all others); none of the two-target conditions (that varied in similarity defined by MDS space) were different from one another. The decision-time data showed a similar pattern of results. In Experiment 4b, the effect of Feature Width was driven exclusively by the difference between two-category trials relative to all others, and in Experiment 4d, the effect was driven exclusively by the difference between single-target trials relative to all others.

In short, we did not observe consistent, graded effects of Feature Width on attentional guidance or decision-making. The overall main effects of Feature Width on scan-path ratios and decision-times were consistent, again strongly supporting the dual-function hypothesis regarding the role of the search template. Clearly, the addition of extraneous features in one’s template causes problems for both attentional guidance and decision-making. But at present, the data are not strong enough to speak to the question of whether or not two-target templates are constructed in such a way that can be considered meshed or discrete. It seems likely that when the two potential targets are highly similar (e.g., in our narrow Feature Width conditions), they may be fused into a single, slightly broader template, and that highly dissimilar target must be represented by discrete templates. But future studies will be needed to further clarify this issue.

Experiment 5

In the final experiment of this investigation, we turned our attention to another way in which target templates may become imprecise. Sometimes our templates are imperfect not because they are wrong or unnecessarily broad, but because their quality becomes unstable over time. Visual working-memory cannot retain information indefinitely (Baddeley, 2003; Vogel et al., 2001), though current theories of VWM disagree on whether individual features are lost gradually (e.g., Fougnie & Alvarez, 2009) or if objects are lost all at once (e.g., Zhang & Luck, 2011).

Imagine, for instance, that you are traveling to your hometown for a visit. Your best friend is busy and so cannot pick you up, but he has graciously left his car for you at the airport, and mailed you a spare key in advance. You know the general vicinity of his car in the parking lot, but it has been months since you've seen the car in person, and it is a very common make and model. Being a savvy searcher (and wishing to avoid having to try your key on every Toyota Corolla in the parking garage), you requested that your friend send a picture of his car to your cell phone. You open the phone and use the picture to refresh your mental template, before being confronted what seems like dozens of Toyota sedans. As you scan the parking garage, you begin to lose the minute details that distinguish your friend's car from other possible targets. Does his car have hubcaps or rims? Is there a spoiler on the back? Was it silver or gray? Gradually you become unconfident about the various details of your target vehicle, so you take your phone out once more to regenerate that pictorial template in your mind.

In Experiment 5, our aim was to examine how the target template changes over time, and how the dynamics of template decay affect search performance. A key comparison in these experiments is to distinguish between template-guided search (i.e., search using picture cues) and categorically-guided search (i.e., search using word cues). In particular, it was interesting to examine the extent to which internal representations generated from these different kinds of cues change over time. In the vast majority of laboratory visual search experiments, each search is completed within a matter of several seconds. This thrift is sub-optimal if we are interested in the extent to which a representation fades from VWM. Accordingly, the procedure we adopted here was to prolong search by having people scan through several displays sequentially, sometimes encountering a target early, and sometimes having to perform up to four consecutive searches before the target was encountered. We compared template-guided search to categorically-guided search, and investigated template deterioration by periodically "refreshing" some participants with the cue (or cues) after each search, and by showing other participants the cue only once (at the start of each trial).

Specifically, in Experiment 5, we were interested in both search accuracy and search RTs, predominantly during target-present trials. With respect to search accuracy, we were interested in how the modality of the cue would interact with the absence of cue refreshing (i.e., the Cue Type by Cue Reminders interaction). And for search RTs, we were specifically interested in how prolonged search would interact with these factors (i.e., the Cue Type by Cue Reminders by Search Length interaction).

Experiments 5a Method

The apparatus and stimuli were identical to Experiments 1-4 (specifically, the non-eyetracking experiments).

Participants.

Sixty-four new students from Arizona State University participated in Experiment 5a, as partial fulfillment of a course requirement. All participants had normal or corrected-to-normal vision, and only one participant reported partial color-blindness (all others reported normal color vision).

Design.

Two levels of Cue Type (pictures, words) and two levels of Cue Reminders (present, absent) were manipulated between-subjects. All other manipulations were within-subjects. Two levels of Target Load (1 target, 2 targets), two levels of Target Presence (present, absent), three levels of Set Size (12, 16, 20), and four levels of Search Length (1-4) were manipulated in equal proportions. There were two blocks of 60 experimental trials (for a total of 120 trials). Four practice trials were administered during the instructions phase.

Procedure.

The trial procedure (see Figure 18) was nearly identical to the previous experiments, except for two primary details. First, when picture cues were used (in contrast to word cues), the search target was unambiguous. That is, the picture shown prior to search was always a veridical representation of the target that appeared in the search display. Second, participants in these experiments had (often) to complete several consecutive searches before they found the target (and sometimes they did not ever find a target). Trials began by showing the participant the picture or word cue (or cues). When people were ready to begin the trial, they pressed the spacebar, at which point they saw a central fixation cross for 1 second. Then they saw the first search display. Participants rested their fingers on the spacebar during search, and depressed it when they found a target, or when they determined that no target was present. The search array was then replaced by a prompt screen, wherein they indicated their search decision (using the “f” and “j” keys to indicate target presence or absence, respectively).

If they correctly located a target, the trial came to an end. If the target was not present (and they indicated target absence), the next screen they encountered was either a reminder prompt wherein they saw the target cue(s) again (Cue Reminder present condition), or a prompt telling them simply to get ready to continue searching (Cue Reminder absent condition). Next, they pressed the spacebar to begin searching anew. They saw another fixation cross, followed by a new visual search display, and so on. This sequence was repeated for up to 4 consecutive searches. Eighty percent of the trials contained a target, which appeared with equal frequency across searches 1, 2, 3 or 4. The remaining 20% were target-absent trials (i.e., participants searched four times but never encountered a target); these trials were included so people would not learn to expect a target on the penultimate search (i.e., search #4). Incorrect responses (misses or false-alarms) on any of the displays caused the trial to terminate. Feedback was provided after each trial, and instructions asked participants to respond as quickly as possible while still retaining a high degree of accuracy. Spatial configurations were randomized on each trial and each search, consistent with prior experiments.

Stimulus selection proceeded much like the previous experiments. Here again, we used the state- and exemplar-pair stimuli for our target pool (however, we were no longer looking for pairs of stimuli with a set degree of similarity). On each single-target trial, one image was quasi-randomly selected from among the 100 state- and 100 state-pair stimuli. On two-target trials, images were selected from two different semantic categories. Each target category was used only once in the experiment. Distractors were chosen (from the pool of 240 semantic categories) for each search display such that no more than one exemplar per category was represented on any given display. Each distractor was used no more than twice over the entire experiment.

Experiment 5a Results

All data were analyzed using 2 (Cue Type: pictures, words) x 2 (Cue Reminders: present, absent) x 2 (Target Load: 1 target, 2 targets) x 3 (Set Size: 12, 16, 20) x 4 (Search Length: 1-4) mixed-model, repeated measures ANOVAs. Target-present and target-absent trials were analyzed separately, for simplicity, and because exhaustive searches were not of substantive interest in this experiment. One and seven participants (from the picture- and word-cues groups, respectively) were excluded from analysis for producing accuracy more than 2.5 standard deviations below the mean of their group. One participant from the word-cue group was excluded from analysis for self-reporting partial color-blindness. Only correct trial RTs were analyzed.

Accuracy.

Target-present trials. Accuracy was reasonably high (85%). There was a main effect of Cue Type, $F(1, 51) = 25.81$, $p < .01$, $n^2_p = .34$, with better performance for those who had seen picture-cues (89%), relative to word-cues (80%). The main effect of Cue Reminders was not significant ($F < 1$). We found a main effect of Target Load, $F(1, 51) = 59.13$, $p < .01$, $n^2_p = .54$, with better performance on 1-target trials (89%), relative to 2-target trials (81%). The main effect of Set Size was significant, $F(2, 50) = 11.77$, $p < .01$, $n^2_p = .32$, showing that participants committed an unequal proportion of errors across set sizes (87%, 82%, and 84% for 12, 16, and 20, respectively). The main effect of Search Length was not significant ($F < 2$).

Of particular interest, we found a Cue Type x Cue Reminders x Target Load interaction, $F(1, 51) = 10.93$, $p < .01$, $n^2_p = .18$ (see Figure 19). To more closely examine this three-way interaction, we performed a post-hoc analysis that examined accuracy as a function of each factor, but separately for the picture- and word-cue groups. For the picture-cue group, we found a marginal effect of Cue Reminders, $F(1, 30) = 3.86$, $p = .06$, $n^2_p = .11$, a main effect of Target Load, $F(1, 30) = 55.21$, $p < .01$, $n^2_p = .65$, and a Target Load x Cue Reminders interaction, $F(1, 30) = 5.82$, $p < .05$, $n^2_p = .16$. Simply, for picture cues, performance was worse under load, was worse without cue reminders, and these two factors were additive. For the word-cue group, there was no main effect of Cue Reminders ($F < 1$). There was a main effect of Target Load, $F(1, 21) = 15.52$, $p < .01$, $n^2_p = .41$, and a Cue Reminders x Target Load interaction, $F(1, 21) = 5.09$, $p < .05$, $n^2_p = .20$. For word cues, performance was worse under load, as with picture cues. However, cue reminders were ineffectual, overall.

Returning to the original omnibus ANOVA, there was also a Cue Reminders x Target Load x Set Size x Search Length interaction, $F(6, 46) = 2.65$, $p < .05$, $n^2_p = .26$. No other interactions were significant (Fs < 3). All means are presented in Table 10.

Target-absent trials. Accuracy was high, overall (98%). There was a main effect of Cue Type, $F(1, 51) = 52.02$, $p < .01$, $n^2_p = .51$, with better performance by those who saw picture cues (99%), relative to those who saw word cues (97%). The main effects of Cue Reminders, Target Load, Set Size, and Search Length were not significant (Fs < 4);

see Figure 20. There was a Set Size x Search Length interaction, $F(6, 46) = 2.66$, $p < .05$, $n^2_p = .26$ (see Appendix Figure A12). All means are presented in Table 11.

Reaction time.

Target-present trials. We found a main effect of Cue Type, $F(1, 51) = 62.48$, $p < .01$, $n^2_p = .55$, with shorter RTs among those who saw picture cues (1453 msec) relative to those who saw word cues (2303 msec). There was a main effect of Cue Reminders, $F(1, 51) = 4.57$, $p < .05$, $n^2_p = .08$, with faster RTs by those who saw cue reminders (1763 msec) than those who did not (1992 msec). There were main effects of Target Load, $F(1, 51) = 193.64$, $p < .01$, $n^2_p = .79$, and Set Size, $F(2, 50) = 48.38$, $p < .01$, $n^2_p = .66$; RTs slowed when searching for multiple targets (1576 and 2179 msec for 1-target and 2-target search, respectively), and when searching at higher set sizes (1677, 1923, and 2032 msec for 12, 16, and 20, respectively). The main effect of Search Length was not significant ($F < 2$).

We found several interactions: Cue Type x Length: $F(3, 49) = 3.28$, $p < .05$, $n^2_p = .17$; Cue Type x Set Size: $F(2, 50) = 8.18$, $p < .01$, $n^2_p = .25$; Cue Reminders x Set Size: $F(2, 50) = 6.64$, $p < .01$, $n^2_p = .21$; and Cue Type x Target Load x Search Length: $F(3, 49) = 4.45$, $p < .01$, $n^2_p = .21$. All group means are presented in Table 12.

To more closely examine our data, we performed a post-hoc analysis that examined RTs as a function of each factor, but separately for the picture- and word-cue groups. In particular, we were interested in the interaction of Cue Reminders and Search Length within each group (see Figure 21). For the picture cues group, there was a main effect of Cue Reminders, $F(1, 30) = 5.58$, $p < .05$, $n^2_p = .16$, and a marginally significant Cue Reminders x Length interaction, $F(3, 28) = 2.74$, $p = .06$, $n^2_p = .23$. For the word cues group, neither the main effect of Cue Reminders, nor the Cue Reminders x Search Length interaction were significant (Fs < 1). Put simply, for picture cues, cue reminders were effective in keeping search RTs short, and consistent across searches. Without them, search RTs were worse overall, and got systematically worse as search wore on. By contrast, for word cues, cue reminders were ineffective.

Target-absent trials. There was a main effect of Cue Type, $F(1, 51) = 33.34$, $p < .01$, $n^2_p = .40$, with faster RTs when people were shown picture cues (2491 msec) relative to those shown word cues (3751 msec). We found a marginal main effect of Cue Reminders, $F(1, 51) = 3.39$, $p = .07$, $n^2_p = .06$, with faster RTs when people were shown reminders (2920 msec), relative to when they were not (3322 msec). There were main effects of Target Load, $F(1, 51) = 235.09$, $p < .01$, $n^2_p = .82$, Set Size, $F(2, 50) = 147.68$, $p < .01$, $n^2_p = .86$, and Search Length, $F(3, 49) = 31.76$, $p < .01$, $n^2_p = .66$. Search was slower under load (2609 and 3633 msec for 1-target and 2-target search, respectively), slower at larger set sizes (2620, 3151, and 3591 msec for 12, 16, and 20, respectively), and people searched more carefully as time progressed (2923, 3033, 3200, and 3328 msec for Searches 1-4).

We found several interactions: Cue Reminders x Search Length: $F(3, 49) = 8.00$, $p < .01$, $n^2_p = .33$; Target Load x Set Size: $F(2, 50) = 11.08$, $p < .01$, $n^2_p = .31$; Cue Type x Target Load x Search Length: $F(3, 49) = 3.16$, $p < .05$, $n^2_p = .16$; Cue Reminders x Target Load x Search Length: $F(3, 49) = 3.15$, $p < .05$, $n^2_p = .16$; and Cue Type x Set Size x Search Length: $F(6, 46) = 3.47$, $p < .01$, $n^2_p = .31$. All group means are presented in Table 13.

As with the target-present RTs, we performed a post-hoc analysis that isolated the effects of Cue Reminders and Search Length for the picture- and word-cues groups separately (see Figure 22). For the picture cues group, there was no main effect of Cue Reminders (*F* < 3), but there was a Cue Reminders x Search Length interaction, *F*(3, 28) = 16.92, $p < .01$, $n^2_p = .64$. For those who saw word cues, there was no main effect of Cue Reminders (*F* < 2), nor was there a Cue Reminder x Search Length interaction (*F* < 2). As with target-present RTs, picture-cued RTs were affected by prolonged search without cue reminders, but with word-cues, reminders were ineffective overall.

Experiment 5a Discussion

The important findings from Experiment 5a can be easily summarized. First, when target templates were constructed from word cues, people committed more errors, relative to when the templates were created from picture cues. Second, when picture-cued searchers were not given target reminders, they committed more misses, but for word-cued searchers, cue reminders had no effect. Third, picture-cued search allows observers to complete the task more quickly than word-cued search. And fourth, without cue reminders, picture-derived templates deteriorate, resulting in slower search RTs to late arriving targets, but for word-derived templates, cue reminders had no effect.

These results are uncontroversial in showing that picture-cued search is faster and more accurate than word-cued search. However, the novel finding here is that unless these picture-derived templates are refreshed, they result in decrements in performance as the quality of the mental representation decays over time. One concern with the present method, however, is that the self-paced nature of the inter-search interval may have allowed people to refresh their own template, by attempting to call to mind the original picture without any time pressure. That is, in between searches, the participant may have taken a moment to reflect upon the original cue, creating a sort of self-imposed refreshing. Indeed, Wilschut and colleagues (2013) showed that a visual memory representation (rather than a percept of the cue) can be turned into an attentional set in under one second.

Therefore, in Experiment 5b, we attempted to replicate these findings with one modification to the procedure. Here, rather than a self-paced inter-search interval, we set the interval to one second in duration for all conditions, in order to limit the amount of time our participants would have to conduct self-refreshing. We hoped that this would exacerbate the effects we found in Experiment 5a.

Experiments 5b Method

The apparatus, stimuli, and design were identical to Experiment 5a.

Participants.

Eighty new students from Arizona State University participated in Experiment 5b, as partial fulfillment of a course requirement. All participants had normal or corrected-to-normal vision, and all others reported normal color vision.

Procedure.

The procedure was identical to Experiment 5a with one exception. In Experiment 5a, the cue-reminder prompt (presented after search search) was self-paced. That is, participants were shown a reminder of the cue(s) or the “Get ready!” prompt for an indeterminate period of time. When they were ready to begin searching anew, they pressed the spacebar to continue the trial. In Experiment 5b, by contrast, this prompt was presented for one second, and then the trial continued without user input.

Experiment 5b Results

The data were analyzed in identical fashion to Experiment 5a. Three participants from the picture-cues group were excluded from analysis for performance that was more than 2.5 standard deviations away from the mean of their group (one for abnormally low accuracy and two for abnormally high RTs). Two participants from the word-cues group were excluded from analysis for producing accuracy that was more than 2.5 standard deviations below the mean of their group. Only correct trial RTs were analyzed.

Accuracy.

Target-present trials. Accuracy was reasonably high (80%). There was a main effect of Cue Type, $F(1, 71) = 35.56$, $p < .01$, $n^2_p = .33$, with better performance by those who were given picture cues (85%), relative to those who saw word cues (76%). The main effect of Cue Reminders was marginally significant, $F(1, 71) = 3.66$, $p = .06$, $n^2_p = .05$, with better performance by those who saw cue reminders (82%), relative to those who did not (79%). There were main effects of Target Load, $F(1, 71) = 68.32$, $p < .01$, $n^2_p = .49$, Set Size, $F(2, 70) = 4.00$, $p < .05$, $n^2_p = .10$, and Search Length, $F(3, 69) = 5.94$, $p < .01$, $n^2_p = .21$. Participants committed more errors under load (85% and 77% for 1-target and 2-target search, respectively), more errors at higher set sizes (82%, 81%, and 79% for 12, 16, and 20, respectively), and an unequal amount of errors as the trials progressed (80%, 79%, 80% and 84% for Searches 1-4). There was a Set Size x Search Length interaction, $F(6, 66) = 3.27$, $p < .01$, $n^2_p = .23$ (see this interaction plotted in Appendix Figure A13). No other interactions were significant (*F*s < 3). All means are presented in Table 14.

Once again, we were particularly interested in the Cue Type x Cue Reminders x Target Load interaction (see Figure 23). This interaction was not significant here ($F < 1$). However, to be consistent with Experiment 5a, we performed the same post-hoc analysis which examined accuracy separately for the picture- and word-cue groups. For the picture-cue group, the main effect of Cue Reminders was not significant ($F < 2$). There was a main effect of Target Load, $F(1, 34) = 33.33$, $p < .01$, $n^2_p = .50$, but no Cue Reminders x Target Load interaction ($F < 1$). For the word-cues group, there was no main effect of Cue Reminders ($F < 3$), but there was a main effect of Target Load, $F(1, 37) = 36.91$, $p < .01$, $n^2_p = .50$. The Cue Reminders x Target Load interaction was not significant ($F < 1$). For picture cues, performance was worse under load, and there was a trend for poorer performance without cue reminders (87% and 84% with and without cue reminders, respectively), but this effect was not statistically reliable. Unlike in Experiment 5a, however, this factors did not exacerbate one another. With word-cues also, performance was worse under load, performance was worse under load, but was unaffected by cue reminders.

Target-absent trials. Accuracy was high, overall (98%). We found a main effect of Cue Type, $F(1, 71) = 32.57$, $p < .01$, $n^2_p = .31$, with better performance by those who saw picture cues (99%), relative to those who saw word cues (98%). See Figure 24. There was a main effect of Cue Reminders, $F(1, 71) = 4.19$, $p < .05$, $n^2_p = .06$, with better performance by those who saw reminders (99%) than by those who did not (98%). We found a main effect of Target Load, $F(1, 71) = 4.25$, $p < .05$, $n^2_p = .06$, with fewer errors when searching for one target (99%), relative to searching for two targets (98%). The main effects of Set Size and Search Length were not significant (Fs < 2). There was a

Cue Type x Target Load x Search Length interaction, $F(3, 69) = 2.81$, $p < .05$, $n^2_p = .11$ (see Appendix Figure A14). All means are presented in Table 15.

Reaction time.

Target-present trials. There was a main effect of Cue Type, $F(1, 71) = 149.72$, $p < .01$, $n^2_p = .68$, with faster RTs by those who saw picture cues (1292 msec), relative to those who saw word cues (2425 msec). There was no main effect of Cue Reminders ($F < 1$). We found main effects of Target Load, $F(1, 71) = 296.69$, $p < .01$, $n^2_p = .81$, Set Size, $F(2, 70) = 56.13$, $p < .01$, $n^2_p = .62$, and Search Length, $F(3, 69) = 7.27$, $p < .01$, $n^2_p = .24$. Search for a single target was faster than search for two (1578 and 2137 msec for 1-target and 2-target search, respectively), was slower at higher set sizes (1641, 1891, and 2040 msec for 12, 16, and 20, respectively), and was inconsistent across searches (1720, 1899, 1959, and 1849 msec for Searches 1-4).

We found several interactions: Cue Type x Target Load: $F(1, 71) = 15.65$, $p < .01$, $n^2_p = .18$; Cue Type x Set Size: $F(2, 70) = 16.30$, $p < .01$, $n^2_p = .32$; Target Load x Search Length: $F(3, 69) = 3.96$, $p < .05$, $n^2_p = .15$; Set Size x Search Length, $F(6, 66) = 3.12$, $p < .01$, $n^2_p = .22$; Cue Type x Cue Reminders x Set Size x Search Length: $F(6, 66) = 2.26$, $p < .05$, $n^2_p = .17$; and Cue Type x Cue Reminders x Target Load x Set Size x Search Length: $F(6, 66) = 2.57$, $p < .05$, $n^2_p = .19$. All group means are presented in Table 16.

Of key interest was the post-hoc analysis that we performed (as in Experiment 5a) which looked at the picture- and word-cue groups separately. Specifically, we are interested in the interaction of Cue Reminders and Search Length within each group (see Figure 25). For the picture cues group, there was a marginally significant effect of Cue

Reminders, *F*(1, 34) = 3.40, $p = .07$, $n^2_p = .09$, but no Cue Reminders x Search Length interaction (*F* < 2). For the word-cues group, neither the main effect of Cue Reminders, nor the Cue Reminders x Search Length interaction were significant (*F*s < 1). Simply put, for those who saw picture cues, cue reminders kept RTs short, but were ineffective for those who saw word cues.

Target-absent trials. We found a main effect of Cue Type, *F*(1, 71) = 93.86, $p < .01$, $n^2_p = .57$, with faster RTs by those who saw picture cues (2207 msec), relative to those who saw word cues (4023 msec). There was no effect of Cue Reminders (*F* < 1). We found main effects of Target Load, *F*(1, 71) = 559.87, $p < .01$, $n^2_p = .89$, Set Size, *F*(2, 70) = 187.27, $p < .01$, $n^2_p = .84$, and Search Length, *F*(3, 69) = 34.26, $p < .01$, $n^2_p = .60$. RTs were longer when searching for two targets (2622 and 3607 msec for 1-target and 2-target search, respectively), longer at higher set sizes (2644, 3115, and 3584 msec for 12, 16, and 20, respectively), and were longer as the trials wore on (2983, 3008, 3196, and 3273 msec for Searches 1-4).

There were several interactions: Cue Type x Target Load: *F*(1, 71) = 9.33, $p < .12$, $n^2_p = .33$; Cue Type x Set Size: *F*(2, 70) = 21.87, $p < .01$, $n^2_p = .39$; Cue Type x Search Length: *F*(3, 69) = 3.18, $p < .05$, $n^2_p = .12$; Cue Reminders x Search Length: *F*(3, 69) = 3.16, $p < .05$, $n^2_p = .12$; Target Load x Set Size: *F*(2, 70) = 23.73, $p < .01$, $n^2_p = .40$; and Target Load x Search Length: *F*(3, 69) = 5.67, $p < .01$, $n^2_p = .20$. All group means are presented in Table 17.

As before, we performed a post-hoc analysis that isolated the effects of Cue Reminders and Search Length for the picture- and word-cue groups separately (see Figure 26). For the picture-cues group, there was no main effect of Cue Reminders (*F* <

2), but there was a marginally significant Cue Reminders x Search Length interaction, $F(3, 32) = 2.82$, $p = .06$, $n^2_p = .21$. For the word-cues group, there was no main effect of Cue Reminders ($F < 1$), nor was there a Cue Reminders x Search Length interaction ($F < 2$). Thus, picture-cued RTs were affected by prolonged search when cue reminders were absent, but for those who saw word cues, reminders were ineffective.

Experiment 5b Discussion

The results of Experiment 5b partially replicated those of Experiment 5a. As before, picture-cued search was faster and more accurate, relative to word-cued search. Again there was a trend in picture-cued search to commit more misses without cue reminders, but this effect was not statistically reliable. On target-present trials, RTs were slower for picture-cued searchers that did not receive cue reminders, but this effect did not interact with the duration of search as in Experiment 5a. And for word-cued search, we replicated the finding that cue reminders did not benefit search in any measurable way.

Our initial rationale for changing the procedure of Experiment 5a was to reduce the amount of time that our participants had to self-refresh their template (when not given cue reminders). It is possible, however, that the imperfect replication of effects in Experiment 5b was a side effect of this change in procedure. Namely, by specifically controlling the duration of the inter-search interval, we also shortened the overall duration of the trials (on average), and thus may have afforded less time for the searchers' templates to deteriorate naturally.

Experiment 5 General Discussion

The findings from Experiment 5 can be summarized by the following three points: 1) picture-cued search is faster and more accurate than word-cued search; 2) When people are not given cue reminders, they tend to miss more targets when their templates are picture-derived, but cue reminders are unhelpful when templates are word-derived; and 3) Finding the target when it is present takes longer when a picture-cued template has deteriorated (over time), but this finding does not hold true for word-cued templates.

Increasing the specificity of the template aids search accuracy and speed.

One of the earliest studies to directly compare template- and categorical-search (picture cues and word cues, respectively) was conducted by Wolfe and colleagues (2004). In their study, people searched for conjunctive targets (e.g., a black vertical bar) in blocks wherein the target was new on each trial, or where the target was consistent throughout a block of trials. Target cues were either words or pictures, and were displayed on-screen briefly, followed by a variable stimulus onset asynchrony (SOA; between 50 and 800 msec). When an exact picture cue was given, people could “set up” their template within 200 msec, such that performance to varied targets was just as efficient as performance to blocked targets (with SOAs under 200 msec, blocked performance was superior). Importantly, word cues took longer to set up, and never quite reached the effectiveness of the picture cues.

Using real-world scene stimuli, Malcolm and Henderson (2009, 2010) tracked the eye movements of participants as they searched for targets with picture or word cues. They divided search time into distinct behavioral epochs: Search initiation time (the time from the offset of the cue to the first fixation in the scene), scanning time (search time up to first fixation on the target), and verification time (time from first fixation on the target to response). They found that picture cues elicited faster scanning times, and faster verification times (but not faster initiation times). Moreover, when picture cues were used, fewer scene regions (the displays were broken up into rectangular areas of interest) were visited, suggesting that knowledge of the target properties better allows the visual system to ignore "noisy" activity from regions of the scene that do not contain target relevant features.

It is uncontroversial that picture cues produce better search performance, relative to word cues. However, what is less certain is the extent to which categorical cues are able to *guide* attention. Some researchers have argued that semantic factors either cannot guide eye movements, or can only do so very weakly (e.g., Henderson, Weeks, & Hollingworth, 1999; Võ & Henderson, 2009). Others, however, have argued differently, for instance, after observing strong effects of categorical or semantic (rather than visual) similarity during visual search (e.g., Meyer et al., 2007; Belke et al., 2008). For instance, Dahan and Tanenhaus (2005) found that when participants searched for an outstretched snake, they were more likely to fixate distractors that looked like typical snakes (e.g., a curled rope), relative to non-snakelike distractors (see also work on "language-mediated eye movements": e.g., Huetigg & Altmann, 2005, 2011; Altmann & Kamide, 2007).

Zelinsky and colleagues are likely the biggest proponents of categorical guidance in visual search. 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 made more fixations. However, they also found that categorically-defined targets were fixated far sooner than would be expected by chance, suggesting some level of attentional guidance with word cues. Importantly, Schmidt and Zelinsky (2009) argued that the amount of guidance that a cue elicits is directly related to the cue's categorical specificity. Their 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 information used in guiding attention need not be limited to pictures of the target, and that guidance improves to the extent that visual information can be extracted from the cue, and loaded into VWM (these assumptions are also built into two recent search models: See Zhang et al., 2006; Navalpakkam & Itti, 2005).

Constructing a template from word cues is effortful.

Deriving a mental template from a picture cue requires little more than simply remembering that picture for a short period of time. By contrast, creating a template from a word cue is an effortful undertaking. One must call upon target-defining features stored in long-term memory to construct a prototypical template that is suitable to the task at hand. In a different investigation, we have been exploring the effortful creation of word-derived templates using pupillometry (Hout, Papesh, & Goldinger, 2012). Our participants searched for one or three potential targets in a rapid-serial visual presentation (RSVP) stream. Cues were either pictures or words. Participants encoded the search targets and then pressed the spacebar to begin a trial. They then saw a fixation cross for 2 seconds, followed by 24 different pictures, centrally presented, for 250 msec each (followed by a 50 msec blank interval in between picture presentation); at the end of the trial, they indicated if a single target was present, or if all targets were absent. Importantly, we used pupil dilation to index cognitive demand across conditions, and throughout the duration of a trial. Pupil dilation is directly influenced by parasympathetic activity in the nervous system (Aston-Jones & Cohen, 2005), and reflects increases in mental effort (Steinhauer et al., 2004; Beatty, 1982; Beatty & Kahneman, 1966; Kahneman, 1973; Papesh & Goldinger, 2012) as well as the creation (and retrieval) of memories (Goldinger & Papesh, 2012; Papesh, Goldinger & Hout, 2011; Võ et al., 2008). Specifically, we found that pupil dilation 1) can be used to index the challenges of difficult searches (Porter, Troscianko & Gilchrist, 2007), 2) reflects the difficulty of setting up a categorical target representation from memory, and 3) is a reliable indicator of target identification (Privatera et al., 2010).

Specifically, we found a number of interesting effects that speak to the mental effort devoted to visual search under various conditions. Recall that an assumption made in Experiment 4 was that by having people search for multiple targets, we increase the width of their feature space, and thereby make search more difficult. Our pupil data corroborate this assumption by showing that pupils are more dilated under high load (i.e., search for three targets), relative to low load (i.e., search 1 target); see Figure 27. More germane to Experiment 5 of the current investigation, we found that pupil dilation also increases as a function of cue type. Figure 28 shows pupil dilation as a function of cue type (template vs. categorical) for target-present trials. Most striking is the difference in pupil dilation during the *fixation* interest period (higher for word cues). Keep in mind, nothing of interest was on the screen during this time (just a simple, central cross), and there were no changes in luminance or anything to cause a reflexive pupil change. Here, the participants were either "setting up" their target template(s) from long-term memory, or were simply trying to maintain the identity of the target(s). Thus, we have evidence that search using word cues requires more cognitive effort, relative to word cues.

Finally, and perhaps most importantly, Figure 29 shows target-present trials, collapsed across all other factors. Four separate lines are plotted, one for trials in which the target showed up in the first, second, third, or fourth quarter of the stream. Two things are important to note about this figure. First, in each case, the targets begin to peak shortly after presentation of the target, indicating that the target has been identified. And second, when the target shows up later in the stream, the height of the peak is reduced. This suggests that as time elapses, the searcher's target template begins to fade. When the target finally appears, it resonates less strongly with what is left of the representation in VWM. This would not be an unprecedented finding. For instance, recently (Papesh, Goldinger & Hout, 2011) we found that pupils were sensitive to the cue match between the first and second presentation of a spoken word. When words were repeated in the same voice, pupils dilated more, relative to when voice was changed from encoding to test. Moreover, strong memory responses (i.e., trials in which participants reported highly confident remembering) elicited higher pupil dilation, relative to weaker responses.

We suggest that an analogous process is occurring here, with respect to the match between what is seen (i.e., the target in the stream), and the representation of that item in VWM. If target templates are degrading over time, this is the exact pattern of findings we would expect, as the cue match becomes weaker with more fading from VWM.

Future studies are being carried out to delineate this account of our findings from alternative explanations. Moreover, we are continuing to model this data using our multiple-trace memory model, *Scanning QUiet Echoes And Learning* (SQUEAL; Hout & Goldinger, 2011).

The stability of picture- and word-derived templates.

Our current findings fit nicely with the established literature comparing picture-cued to word-cued search. Like others, we found that picture cues allow searchers to be fast and accurate, but there is a caveat. Highly specific templates are only useful insofar as that specificity is able to be retained over time. Unlike laboratory search experiments, search in the real-world can often be effortful and time-consuming. Our data suggests that highly specific templates are prone to decrements in performance during prolonged search, because the features of such templates may be lost over time. Whether this deterioration is due to decay from VWM, or interference with other information that is being viewed is unclear. It is possible that feature values are simply lost with the passage of time, but it seems equally likely that incoming visual input – which makes its way into VWM in order to be compared to the search template – may be “pushing” the target features out of VWM over time. Future work is needed to elucidate between these two possibilities.

What does seems clear is that templates derived from word cues, though slower and comparatively less accurate, are immune (or, at least, are resilient) to the hazardous effects of the passage of time. It is important to note here, that we are arguing from a set of null effects: Specifically, the absence of Cue Reminders or Search Length effects on word-cued participants’ behavior. It could therefore be argued that what we found was not a resiliency to the passage of time, but rather “floor-level” performance. Perhaps word-cued searchers were so slow and inaccurate that they could not perform any worse. Although we cannot use the observed pattern of effects to actively argue against this notion, we think this is a highly unlikely hypothesis. Although word-cued search was inferior to picture-cued search, overall performance was still very high: Overall accuracy was in the 90% range, and target location infrequently took longer than three seconds. It seems unlikely that a floor effect would be present in performance of such high caliber.

Undoubtedly, it is beneficial to have a target template that is highly precise and detailed. And in some real-life situations, we are afforded such a luxury. When we search for things we are highly familiar with, we often carry a mental picture of those items that is thorough, and perhaps robust to the passage of time. Looking for one’s backpack, for instance, may be easy because the specific details of the item are known, and the backpack has been seen so many times that its features are unlikely to escape one’s mental representation easily. But in other situations, such as looking for a friend’s car at the airport, our highly detailed representations may be more elusive, or prone to decay. More challenging yet, is the fact that a great many of our everyday searches are conducted with no explicit knowledge regarding the appearance of the target. Searching someone’s toolshed for a rake, or looking for a stuffed bunny at a toystore: These situations involve searching for an object that is categorically defined, and for which we have no specific prior experience. Yet, we can conduct such tasks with little more effort than when we have seen our target previously. Although a great many things have been learned about visual attention by studying picture-cued search, it seems high time that the field turn its focus to situations that real searchers are more likely to encounter: Namely, searching for items for which we have limited, or imprecise knowledge.

Future Directions

The experiments reported herein represent a substantial empirical contribution to the investigation of target templates in visual search, but they leave open several theoretical questions that may prove fruitful for future investigations. In Experiment 3, we provided convincing evidence that when inaccurate features are included in one's search template, both the attentional guidance and decision-making behavior of the searcher will be impaired. One question that remains from this experiment regards the extent to which searchers have flexible control over the quality of their templates. We attempted to answer this question by manipulating the trustworthiness of the cues, and by asking people to search using cues that were never fully accurate. But our results suggest that the participants simply adopted the cue as their template, out of convenience. In future experiments, it would be interesting to investigate the extent to which participants can create a broad, categorical template, or one that is dissimilar to the cue that is provided to them. For instance, if participants know that a cue is categorically accurate, but dissimilar to the actual target they will locate, can they hone their templates to match the category and not the specific features of the cue? If we show people a tall, red coffee mug, for example, are they able to create from memory an antithetical template that would allow them to quickly spot a short, blue mug? If so, this would suggest that memory derived templates are under the flexible control of the observer, and can be honed with a high degree of expediency.

In Experiment 4, we provided additional evidence that imprecise templates hinder both aspects of search, finding performance decrements when extraneous features were included in our searchers' templates. Taken together, these experiments provide strong support for a dual-function theory of target templates, whereby they are used both to select appropriate regions of space in which to place one's attention, and as the comparison material against which incoming visual input is matched. An unresolved issue from this experiment concerns the nature of mental 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 the eye-movement record between single- and two-target searches, and (within the two-target search conditions) between single-category and two-category search. Our findings were suggestive that highly similar targets could be meshed into a single representation, and that highly dissimilar targets will be maintained separately. The data, however, were inconclusive and so we cannot make a strong case for this at present.

Future experiments should be aimed at further delineating this issue by, perhaps, attempting to model the decision-making process in analogous fashion to the Sternberg memory search experiments (e.g., Sternberg, 1966). If we stretch out search to three, four or even five potential targets, and if we varied the similarity of items within those sets (relative to one another), it would be possible to fit a model to decision-times (and even individual fixation durations) in such a way as to make a case for the number of templates being maintained in tandem. Imagine, for instance, that you are asked to search for three targets. They could be three categorically different targets (e.g., a mug, a lamp, and a phone), or some combination wherein multiple targets are categorically alike (e.g., a mug and two similar lamps, or a mug and two dissimilar lamps). By contrasting the data across these various conditions, we may come to a closer understanding of the situations in which target templates can be collapsed, relative to when they must be maintained separately. If, for example, the data when people search for one mug and two similar lamps resembles that of when they search for one mug and one lamp, we may be inclined to argue that the searcher collapsed the similar lamps into a single representation. This again would suggest that people have flexible control over their templates, and can constrain the peculiarities of their template to match changing task demands.

Finally, in the penultimate experiment, we provided evidence showing that the specificity of the template affects both search accuracy and search time. When people have highly specific, picture-derived templates, their performance is superior to when they implement less detailed, word-derived templates. However, the specificity advantage is fragile, in that highly tuned templates may degrade over time, and become less useful once they have deteriorated. The next logical question to ask then, is whether or not there are situations in which categorical templates afford equal (or even superior) search performance to pictorial ones. In essence, one might imagine that a picture-derived template, given sufficient time to decay, would transform into something that resembles a categorical template. If remembering the search target becomes particularly difficult, the searcher may in fact bring to mind incorrect features in an attempt to refresh their template. When searching for a friend’s car in the airport parking garage, for example, you may incorrectly recall that the vehicle was silver, when in fact it was gray. As such, it would be interesting to elongate search to such an extent that picture-cued performance matches or is surpassed by word-cued search. Finding this would suggest that although template precision is important – and clearly all of the current results speak to this point – if that precision cannot be faithfully maintained over time, it may be better to adopt a more general template that is resilient, and tolerant to deviation.

REFERENCES

Adams, R. C., & Chambers, C. D. (2012). Mapping the timecourse of goal-directed attention to location and color in human vision. *Acta Psychologica, 139,* 512-523.

Alexander, R. G., & Zelinsky, G. J. (2011). Visual similarity effects in categorical search. *Journal of Vision, 11,* 1-15.

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.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57,* 502-518.

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.

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.

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.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28,* 403-450.

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.

Baddeley, A. (2003). Working memory: Looking back and forward. *Nature Reviews: Neuroscience, 4,* 829-839.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321,* 851–854.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91,* 276-292.

Beatty, J., & Kahneman, D. (1966). Pupil diameter and load on memory. *Science, 154,* 1583-1585.

Beck, V. M., Hollingworth, A., & Luck, S. J. (2012). Simultaneous control of attention by multiple working memory representations. *Psychological Science, 23,* 887-898.

Becker, S. I. (2011). Determinants of dwell time in visual search: Similarity or perceptual difficulty? *PLoS One, 6,* 1-5.

Becker, W. (1972). The control of eye movements in the saccadic system. *Bibliotheca Opthalamologica, 82,* 233-243.

Belke, E., Humphreys, G. W., Watson, D. G., Meyer, A. S., & Telling, A. L. (2008). Top-down effects of semantic knowledge in visual search are modulated by cognitive but not perceptual load. *Perception & Psychophysics, 70,* 1444-1458.

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.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177,* 77-80.

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.

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.

Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision, 9,* 1-9.

Bravo, M. J., & Farid, H. (2012). Task demands determine the specificity of the search template. *Attention, Perception & Psychophysics, 74,* 124-131.

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.

Busing, F. M. T. A., Commandeur, J. J. F., Heiser, W. J., Bandilla, W., & Faulbaum, F. (1997). PROXSCAL: A multidimensional scaling program for individual differences scaling with constraints. *Advances in Statistical Software, 6,* 67-73.

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.

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

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.

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

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., Linenweber, M. R., Peterson, S. E., Raichle, M. E., Van Essen, D. C., & Shulman, G. L. (1998). A common network of functional areas for attention and eye movements. *Neuron, 21,* 761-773.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87-185.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11,* 671-684.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review, 12,* 453-459.

De Graef, P, Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52,* 317-329.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193-222.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36,* 1827-1837.

Deubel, H., Wolf, W., & Hauske, G. (1982). Corrective saccades: Effect of shifting the saccade goal. *Vision Research, 22,* 353-364.

Dowd, E. W., & Mitroff, S. R. (*in press*). Attentional guidance by working memory overrides saliency cues in visual search. *Journal of Experimental Psychology: Human Perception and Performance*.

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

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.

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.

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.

Ehringer, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modeling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition, 17,* 945-978.

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.

Ethel, M. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81,* 899-917.

Evans, K. K., Horowitz, T. S., Howe, P., Pedersini, R., Reijnen, E., Pinto, Y., Kuzmova, Y., & Wolfe, J. M. (2011). Visual attention. *Wiley Interdisciplinary Reviews: Cognitive Science, 2,* 503-514.

Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research, 37,* 617-631.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision, 11,* 1-12.

Frings, C., Wentura, D., & Wühr, P. (2012). On the fate of distractor representations. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 570-575.

Giguère, G. (2006). Collecting and analyzing data in multidimensional scaling experiments: A guide for psychologists using SPSS. *Tutorials in Quantitative Methods for Psychology, 2,* 26-37.

Gillund, G. & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1-67.

Godwin, H. H., Hout, M. C., & Menneer, T. (*under review*). Visual similarity is stronger than semantic similarity in guiding visual search for numbers.

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.

Godwin, H. J., Menneer, T., Cave, K. R., Helman, S., Way, R. L., & Donnelly, N. (2010). The impact of relative prevalence on dual-target search for threat items from airport screening. *Acta Psychologica, 134,* 79-84.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251-279.

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review, 11,* 716-722.

Goldinger, S D., & Papesh, M. H. (2012). Pupil dilation reveals the creation and retrieval of memories. *Current Directions in Psychological Science, 21,* 90-95.

Goldstone, R. L. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers, 26,* 381-386.

Goldstone, R. L., & Medin, D. L. (1994). The time course of comparison. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20,* 29-50.

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.

Henderson, J. M. (2012). Eye movements and scene perception. In: S. P. Liversedge, I. D. Gilchrist, & S. Everling (editors), Oxford Handbook on Eye Movements. Oxford, UK: Oxford University Press, pp. 593-606.

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., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory & Cognition, 16,* 417-429.

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.

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review, 15,* 566-573.

Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition, 17,* 1055-1072.

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

Hintzman, D. L. (1984). MINERVA2: A simulation model of human memory. *Behavior Research Methods, 16,* 96-101.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411-428.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528-551.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Pscyhophysics, 57,* 787-795.

Hooge, I. T. C., & Erkelens, C. J. (1998). Adjustment of fixation duration in visual search. *Vision Research, 38,* 1295-1302.

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.

Hollingworth, A., Maxcey-Richard, A. M., Vecera, S. P. (2012). The spatial distribution of attention within and across objects. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 135-151.

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.

Homa, D., Hout, M. C., Milliken, L., & Milliken, A. M. (2011). Bogus concerns about the false prototype enhancement effect. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37,* 368-377.

Homa, D., Proulx, P., & Blair, M. (2008). The modulating influence of category size on the classification of exception patterns. *The Quarterly Journal of Experimental Psychology, 61,* 425-443.

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.

Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception & Psychophysics, 72,* 1267-1282.

Hout, M. C., & Goldinger, S. D. (2011). Multiple-target search increases workload but enhances incidental learning: A computational modeling approach to a memory paradox. *In* Object Perception, Attention & Memory 2011 Conference Report, *Visual Cognition,* 18-21.

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.

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.

Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2012). RSVPupillometry: Incidental memory and psychophysiology in rapid-serial multiple-target search. *Journal of Vision, 12,* 1151-1152.

Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2013). Multidimensional scaling. *Wiley Interdisciplinary Reviews (WIREs): Cognitive Science, 4,* 93-103.

Houtkamp, R., & Roelfsema, P. R. (2009). Matching of visual input to only one item at any one time. *Psychological Research, 73,* 317-326.

Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review, 114,* 599-631.

Huetigg, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixations: Semantic competitor effects and the visual world paradigm. *Cognition, 96,* 23-32.

Huetigg, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing "frog": How object surface colour and stored object colour knowledge influence language-mediated overt attention. *Quarterly Journal of Experimental Psychology, 64,* 122-145.

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.

Ishihara, S. (1993). *Ishihara's tests for colour-blindness.* Tokyo, Japan: Kanehara & Co.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489-1506.

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience, 2,* 194-203.

Jaworska, N., & Chupetlovska-Anastasova, A. (2009). A review of multidimensional scaling (MDS) and its utility in various psychological domains. *Tutorials in Quantitative Methods for Psychology, 5,* 1-10.

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition, 17,* 979-1003.

Kang, M., Hong, S. W., Blake, R., & Woodman, G. F. (2011). Visual working memory contaminates perception. *Psychonomic Bulletin & Review, 18,* 860-869.

Kahneman, D. (1973). Attention and effort. *Prentice Hall: New York.*

Kapoula, Z, & Robinson, D. A. (1986). Saccadic undershoot is not inevitable: Saccades can be accurate. *Vision Research, 26,* 735-743.

Kinchla, R. A., & Wolfe, J. M. (1979). The order of visual processing: “Top-down”, “bottom-up”, or “middle-out”. *Perception & Psychophysics, 25,* 225-231.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4,* 219-227.

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.

Kotowicz, A., Rutishauser, U., & Koch, C. (2010). Time course of target recognition in visual search. *Frontiers in Human Neuroscience, 4,* 1-11.

Kriegeskorte, N., & Marieke, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology, 3,* 1-12.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional Scaling.* Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-011. Beverly Hills and London: Sage Publications.

Kunar, M. A., Flusberg, S., & Wolfe, J. M. (2008b). The role of memory and restricted context in repeated visual search. *Perception & Psychophysics, 70,* 314-328.

Lee, M. D. (2001). Determining the dimensionality of multidimensional scaling representations for cognitive modeling. *Journal of Mathematical Psychology, 45,* 149-166.

Luck, S. J. (2008). Visual short-term memory. In S. J. Luck & A. Hollingworth (Eds.), *Visual Memory* (pp. 43-85). New York: Osford University Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279-281.

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.

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.

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.

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.

McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., & Boot, W. B. (2004). Visual skills in airport-security screening. *Psychological Science, 15,* 302-306.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27,* 817-835.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254-278.

Menneer, 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.

Menneer, 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.

Menneer, 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.

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. L. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review, 14,* 710-716.

Moore, C. M., & Osman, A. M. (1993). Looking for two targets at the same time: One search or two? *Perception & Psychophysics, 53,* 381-390.

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.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45,* 205-231.

Neider, M. B., & Zelinsky, G. J. (2006). Searching for camouflaged targets: Effects of target-background similarity on visual search. *Vision Research, 46,* 2217-2235.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review, 117,* 382-405.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28,* 411-421.

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). *Top-down control of visual attention in object detection.* Paper presented at the Proceedings of the IEEE International Conference on Image Processing, Barcelona, Spain.

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.

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.

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.

Pannasch, S., Schulz, J., & Velichkovsky, B. M. (2011). On the control of visual durations in free viewing of complex images. *Attention, Perception & Psychophysics, 73,* 1120-1132.

Papesh, M. H., & Goldinger, S. D. (2012). Pupil-BLAH-Metry: Cognitive effort in speech planning revealed by pupil dilation. *Attention, Perception & Psychophysics, 74,* 754-765.

Papesh, M. L., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Pscyhophysiology, 83,* 56-64.

Peelen, M. V., Li, F.-F., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature, 460,* 94-97.

Phillips, M. H., & Edelman, J. A. (2008). The dependence of visual scanning performance on saccade, fixation, and perceptual metrics. *Vision Research, 48,* 926-936.

Pietrewicz, A. T., & Kamil, A. C. (1979). Search image formation in the blue jay (*Cyanocitta cristata*). *Science, 204,* 1332-1333.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye movement control. *Cognitive Psychology, 52,* 1-56.

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *Quarterly Journal of Experimental Psychology, 60,* 211-229.

Privatera, C. M., Renninger, L. W., Carney, T., Klein, S., & Aguilar, M. (2010). Pupil dilation during visual target detection. *Journal of Vision, 10,* 1-14.

Rabinowitz, G. B. (1975). An introduction to nonmetric multidimensional scaling. *American Journal of Political Science, 19,* 343-390.

Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research, 42,* 1447-1463.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E–Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 720–732.

Rayner, K., Castelhano, M., & Yiang, J. (2009). Eye movements when looking at unusual/weird scenes: Are there cultural differences? *Journal of Experimental Psychology: Learning, Memory and Cognition, 35,* 254-259.

Reichle, E. D, & Laurent, P. A. (2006). Using reinforcement learning to understand the emergence of “intelligent” eye-movement behavior during reading. *Psychological Review, 113,* 390-408.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105,* 125-157.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using E-Z Reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye-mind link. *Psychological Review, 119*, 155-185.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16,* 1-21.

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. Psychological Science, 12, 48-55.

Reingold, E. M., & Sheridan, H. (2012). Eye movements and visual expertise in chess and medicine. In: S. P. Liversedge, I. D. Gilchrist, & S. Everling (editors), Oxford Handbook on Eye Movements. Oxford, UK: Oxford University Press, pp. 523-550.

Richard, A. M., Luck, S. J., & Hollingworth, A. (2008). Establishing object correspondence across eye movements: Flexible use of spatiotemporal and surface feature information. *Cognition, 109,* 66-88.

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.

Schneider, W., Eschman, A. & Zuccolotto, A. (2002). *E-Prime User's Guide.* Pittsburgh, PA: Psychology Software Tools Inc.

Schwarz, W., & Eiselt, A. K. (2012). Numerical distance effects in visual search. *Attention, Perception & Psychophysics, 74,* 1098‑103. doi:10.3758/s13414-012-0342-8

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317-1323.

Shepard, R. N. (2004). How a cognitive psychologist came to seek universal laws. *Psychonomic Bulletin & Review, 11,* 1-23.

SPSS Inc. (2011). *SPSS Base 20.0 for Windows User's Guide.* SPSS Inc., Chicago, IL.

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 248-261.

Soto, D., Hodsoll, J., Rotshein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences, 12,* 342-348.

Steinhauer, S. R., Siegle, G. J., Condray, R., Pless, M. (2004). Sympathetic and parasympathetic innervations of pupillary dilation during sustained processing. *International Journal of Psychophysiology, 52*, 77-86.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652-645.

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.

Stokes, M., Thompson, R., Nobre, A. C., & Duncan, J. (2009). Shape-specific preparatory activitiy mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences, 106,* 19569-19574.

Stroud, M. J., Menneer, T., Cave, K. R., & Donnelly, N. (2012). Using the dual-target cost to explore the nature of search target representations. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 133-122.

Stroud, M. J., Menneer, 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.

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.

Thiele, A., Henning, P., Kubischik, M, & Hoffman, K. P. (2002). Neural mechanisms of saccadic suppression. *Science, 295,* 2460-2462.

Theeuwes, J., & van der Burg, E. (2011). On the limits of top-down control of visual selection. *Attention, Perception & Psychophysics, 73,* 2092-2103.

Tinbergen, N. (1960). The natural control of insects in pine woods: Vol. I. Factors influencing the intensity of predation by songbirds. *Archives Neelandaises de Zoologie, 13*, 265-343.

Treisman, A. (1991). Search, similarity, and the integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 252-276.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95,* 15–48.

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

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search. *Psychological Review, 113,* 766-786.

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.

Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition, 17,* 159-170.

Vickery, T. J., King, L., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision, 5,* 81-92.

Võ, M. L.-H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schact, A., Hutzler, F. (2008). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology, 45,* 30-40.

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision, 9,* 1-15.

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.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 92-114.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time-course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1436-1451.

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.

Watson, D. G., & Humphreys, G. W. (2000). Visual marking: Evidence for inhibition using a probe-dot paradigm. *Perception & Psychophysics, 62,* 471-481.

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.

Williams, D. E., Reingold, E. M., Moscvitch, M., & Behrmann, M. (1997). Patterns of eye movements during parallel and serial visual search tasks. *Canadian Journal of Experimental Psychology, 51,* 151-164.

Wilschut, A., Theeuwes, J., & Olivers, C. N. L. (2013). The time it takes to turn a memory into a template. *Journal of Vision, 13,* 8-11.

Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology: Human Perception and Performance, 33,* 363-377.

Woodman, G. F., & Luck, S. J. (2010). Why is information displaced from visual working memory during visual search. *Visual Cognition, 18,* 275-295.

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.

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

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9,* 33-39.

Wolfe, J. M. (2005). Watching single cells pay attention. *Science, 308,* 503-504.

Wolfe, J. M. (2007). Guided Search 4.0: Current progress with a model of visual search. In: *Integrated models of cognitive systems,* W. D. Gray (editor). New York, NY, USA: Oxford University Press, pp. 99-119.

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.

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.

Wolfe, J. M., Friedman-Hill, S., Stewart, M., & O'Connell, K. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 34-49.

Wolfe, J. M., & Gancarz, G. (1996). Guided Search 3.0: A model of visual search catches up with Jay Enoch 40 years later. In V. Lakshminrayanan (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.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research, 44,* 1411-1426.

Yang, H., Chen, X., & Zelinsky, G. J. (2009). A new look at novelty effects: Guiding search away from old distractors. *Attention, Perception & Psychophysics, 71,* 554-564.

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research, 49,* 2095-2103.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115,* 787-835.

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science, 20,* 423-428.

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.

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.

FOOTNOTES

1. The Ishihara color-blindness assessment was not used in all experiments for two reasons. First, this test can only be administered to one person at a time, which makes it unfeasible to test participants in the non-eyetracking experiments (wherein many participants took part in the experiment simultaneously). As such, the Ishihara test was only administered during eyetracking studies, and we relied on self-report otherwise. Secondly, when some of the experiments were conducted, we did not yet have access to the Ishihara test, so eyetracking experiments did not universally employ it (only those that were conducted after January 2013 involved this test).

2. We did not entertain the next logical hypothesis – which would be that neither attentional guidance nor decision-making are hindered by imprecise templates – for two reasons. First, because we felt that this was a straw-man hypothesis that no current theories are likely to support. And second, because arguing for this hypothesis would require us to make conclusions based on a set of null effects.

TABLES

Table 1.

Same/different task accuracy from Experiments 1a and 1b, presented as a function of Trial Type and Block.

	Block				
Trial Type	1	2	3	4	Mean
Experiment 1a					
Different	96%	97%	97%	97%	97%
Same, state-pairs	96%	95%	95%	95%	95%
Same, exemplar-pairs	93%	94%	93%	93%	93%
Mean	95%	95%	95%	95%	95%
Experiment 1b					
Different	97%	98%	99%	97%	98%
Same, state-pairs	95%	94%	97%	96%	95%
Same, exemplar-pairs	94%	93%	94%	92%	93%
Mean	95%	95%	96%	95%	95%

Table 2.

Visual search accuracy from Experiment 3a, presented as a function of Template Precision, Set Size, and Block.

Template Precision	Set Size	Block 1	Block 2	Block 3	Block 4
Precise	12	99%	98%	99%	98%
	16	99%	98%	98%	98%
	20	99%	99%	99%	99%
Imprecise	12	100%	100%	99%	99%
	16	98%	95%	97%	99%
	20	99%	96%	93%	99%
Inaccurate	12	93%	99%	95%	95%
	16	97%	98%	96%	94%
	20	96%	96%	94%	96%

Table 3.

Visual search accuracy from Experiment 3b, presented as a function of Template Precision, Match Frequency, Set Size, and Block.

		Match Frequency								
		Low			Medium			High		
Template Precision	Set Size	Block 1	Block 2	Block 3	Block 1	Block 2	Block 3	Block 1	Block 2	Block 3
Precise	12	100%	98%	98%	99%	99%	99%	97%	100%	99%
	16	98%	98%	100%	98%	99%	100%	98%	98%	98%
	20	98%	98%	100%	99%	100%	98%	95%	99%	97%
Imprecise	12	100%	98%	100%	100%	99%	100%	96%	96%	96%
	16	100%	98%	98%	97%	99%	99%	96%	93%	100%
	20	98%	100%	98%	97%	96%	96%	96%	100%	100%
Inaccurate	12	98%	98%	98%	99%	96%	100%	96%	96%	93%
	16	98%	98%	98%	97%	99%	99%	100%	100%	93%
	20	95%	96%	95%	99%	94%	94%	96%	100%	96%

Table 4.

Visual search accuracy from Experiments 3c and 3d, presented as a function of Template Precision, Set Size, and Block.

		Experiment 3c				Experiment 3d			
Template Precision	Set Size	Block 1	Block 2	Block 3	Block 4	Block 1	Block 2	Block 3	Block 4
Precise	12	99%	98%	97%	97%				
	16	99%	99%	97%	98%				
	20	99%	97%	97%	97%				
Similar	12	97%	98%	95%	97%	98%	96%	98%	96%
	16	96%	96%	97%	97%	97%	98%	96%	96%
	20	98%	98%	97%	98%	97%	96%	95%	92%
Moderate	12	97%	98%	97%	98%	97%	97%	97%	97%
	16	95%	98%	96%	97%	97%	98%	95%	96%
	20	95%	96%	97%	99%	96%	96%	95%	95%
Dissimilar	12	97%	100%	98%	99%	98%	97%	97%	95%
	16	97%	98%	96%	97%	97%	97%	97%	95%
	20	96%	99%	97%	96%	97%	96%	95%	96%

Table 5.

Visual search accuracy from Experiments 3e and 3f, presented as a function of Template Precision, Set Size, and Block.

		Experiment 3e				Experiment 3f			
Template Precision	Set Size	Block 1	Block 2	Block 3	Block 4	Block 1	Block 2	Block 3	Block 4
	12	98%	99%	98%	98%				
Precise	16	99%	98%	99%	100%				
	20	97%	96%	98%	98%				
	12	98%	94%	98%	97%	98%	100%	98%	96%
Similar	16	98%	95%	98%	100%	99%	96%	97%	95%
	20	96%	94%	96%	95%	98%	96%	97%	97%
	12	98%	96%	96%	100%	99%	97%	99%	97%
Moderate	16	97%	100%	94%	94%	99%	98%	97%	96%
	20	97%	94%	98%	94%	99%	97%	96%	95%
	12	99%	98%	96%	96%	100%	96%	97%	98%
Dissimilar	16	96%	96%	96%	100%	95%	98%	96%	96%
	20	92%	94%	97%	95%	96%	95%	94%	98%

Table 6.

Visual search accuracy from Experiment 4a, presented as a function of Feature Width, Set Size, and Block.

Feature Width	Set Size	Block 1	Block 2	Block 3	Block 4
Precise	12	99%	98%	98%	97%
	16	99%	99%	98%	97%
	20	99%	99%	97%	96%
Narrow	12	99%	99%	99%	98%
	16	99%	99%	98%	99%
	20	99%	99%	96%	99%
Wider	12	98%	98%	97%	99%
	16	97%	97%	97%	96%
	20	97%	97%	98%	98%

Table 7.

Visual search accuracy from Experiment 4b, presented as a function of Feature Width, Set Size, and Block.

Feature Width	Set Size	Block 1	Block 2	Block 3
Precise	12	99%	100%	99%
	16	99%	98%	99%
	20	99%	99%	100%
Narrow	12	100%	100%	98%
	16	100%	98%	100%
	20	100%	99%	100%
Wider	12	98%	100%	98%
	16	100%	97%	99%
	20	100%	100%	98%
Widest	12	99%	99%	97%
	16	98%	97%	97%
	20	98%	94%	97%

Table 8.

Visual search accuracy from Experiment 4c, presented as a function of Feature Width, Set Size, and Block.

Feature Width	Set Size	Block 1	Block 2	Block 3	Block 4
Precise	12	99%	98%	97%	99%
	16	98%	98%	97%	99%
	20	100%	99%	98%	98%
Similar	12	98%	99%	96%	96%
	16	99%	97%	97%	97%
	20	99%	98%	98%	95%
Moderate	12	99%	99%	98%	96%
	16	98%	98%	98%	97%
	20	98%	98%	98%	96%
Dissimilar	12	98%	97%	97%	98%
	16	98%	98%	97%	96%
	20	99%	98%	96%	97%

Table 9.

Visual search accuracy from Experiment 4d, presented as a function of Feature Width, Set Size, and Block.

Feature Width	Set Size	Block 1	Block 2	Block 3	Block 4
Precise	12	100%	100%	99%	99%
	16	100%	100%	97%	95%
	20	100%	99%	96%	100%
Similar	12	99%	99%	98%	96%
	16	100%	99%	97%	98%
	20	99%	100%	99%	99%
Moderate	12	99%	100%	94%	99%
	16	98%	100%	96%	99%
	20	99%	97%	98%	99%
Dissimilar	12	99%	98%	99%	98%
	16	95%	95%	99%	99%
	20	100%	100%	96%	99%

Table 10.

Target-present visual search accuracy from Experiment 5a, presented as a function of Cue Type, Cue Reminders, Set Size, Target Load, and Search Length.

			1-target				2-targets			
Cue Type	Cue Reminders	Set Size	Search 1	Search 2	Search 3	Search 4	Search 1	Search 2	Search 3	Search 4
Pictures	Present	12	91%	100%	97%	97%	93%	86%	91%	94%
		16	99%	94%	93%	94%	78%	82%	91%	85%
		20	88%	90%	91%	94%	85%	87%	84%	94%
	Absent	12	93%	87%	93%	98%	83%	82%	83%	91%
		16	95%	93%	93%	85%	85%	70%	79%	80%
		20	97%	88%	95%	97%	78%	82%	74%	77%
Words	Present	12	84%	86%	86%	82%	82%	81%	74%	84%
		16	89%	80%	89%	72%	64%	67%	62%	73%
		20	91%	93%	71%	89%	77%	77%	74%	80%
	Absent	12	81%	96%	81%	92%	77%	88%	82%	82%
		16	75%	88%	77%	83%	77%	79%	85%	76%
		20	79%	71%	81%	85%	73%	79%	79%	81%

Table 11.

Target-absent visual search accuracy from Experiment 5a, presented as a function of Cue Type, Cue Reminders, Set Size, Target Load, and Search Length.

			1-target				2-targets			
Cue Type	Cue Reminders	Set Size	Search 1	Search 2	Search 3	Search 4	Search 1	Search 2	Search 3	Search 4
Pictures	Present	12	100%	100%	100%	100%	99%	100%	100%	100%
		16	100%	100%	99%	100%	99%	99%	99%	100%
		20	100%	99%	99%	100%	99%	100%	99%	100%
	Absent	12	100%	100%	100%	98%	99%	98%	97%	98%
		16	100%	99%	99%	100%	98%	99%	100%	97%
		20	99%	99%	100%	98%	100%	99%	98%	100%
Words	Present	12	100%	98%	95%	97%	97%	97%	99%	100%
		16	99%	96%	99%	98%	98%	98%	100%	98%
		20	98%	98%	94%	98%	98%	96%	97%	98%
	Absent	12	97%	98%	99%	93%	99%	94%	98%	95%
		16	98%	99%	100%	94%	96%	99%	97%	94%
		20	98%	97%	98%	98%	97%	99%	97%	97%

Table 12.

Target-present visual search RTs from Experiment 5a, presented as a function of Cue Type, Cue Reminders, Set Size, Target Load, and Search Length.

			1-target				2-targets			
Cue Type	Cue Reminders	Set Size	Search 1	Search 2	Search 3	Search 4	Search 1	Search 2	Search 3	Search 4
Pictures	Present	12	1055	1016	945	931	1391	1389	1523	1590
		16	1034	1010	1127	1068	1366	1799	1381	1610
		20	1101	1103	1152	1203	1664	1763	1679	1694
	Absent	12	1161	1359	1124	1365	1429	1423	1788	1645
		16	1177	1225	1347	1371	1726	1658	2192	2017
		20	1162	1400	1369	1257	1726	2664	2348	2217
Words	Present	12	1738	1823	1817	1573	2359	2204	2414	2313
		16	1985	1847	1902	2164	2497	2407	2760	2342
		20	2091	2215	1913	2093	2701	2656	2361	2842
	Absent	12	1707	1873	1692	1624	2392	2152	2313	2536
		16	2190	2093	2010	2327	3184	2560	3140	3028
		20	2089	2230	1975	2598	3240	2454	3277	2801

Table 13.

Target-absent visual search RTs from Experiment 5a, presented as a function of Cue Type, Cue Reminders, Set Size, Target Load, and Search Length.

			1-target				2-targets			
Cue Type	Cue Reminders	Set Size	Search 1	Search 2	Search 3	Search 4	Search 1	Search 2	Search 3	Search 4
Pictures	Present	12	1473	1449	1619	1602	2232	2268	2338	2435
		16	1716	1779	1922	1862	2610	2609	2783	2878
		20	2160	2085	2191	2345	3001	3077	3121	3456
	Absent	12	1516	1714	1987	1802	2193	2542	2896	2825
		16	1851	2153	2232	2424	2859	3247	3552	3820
		20	2148	2325	2737	2879	3261	3623	3825	4161
Words	Present	12	2592	2507	2571	2756	3357	3122	3424	3596
		16	2869	3066	3133	3148	4103	4017	4292	4137
		20	3369	3456	3406	3788	4604	4534	4685	4634
	Absent	12	2686	2755	2903	3145	3496	3680	3990	4396
		16	3302	3412	3597	3575	4210	4430	4613	4638
		20	3599	3874	3817	3928	4955	5072	5165	5640

Table 14.

Target-present visual search accuracy from Experiment 5b, presented as a function of Cue Type, Cue Reminders, Set Size, Target Load, and Search Length.

			1-target				2-targets			
Cue Type	Cue Reminders	Set Size	Search 1	Search 2	Search 3	Search 4	Search 1	Search 2	Search 3	Search 4
Pictures	Present	12	96%	83%	85%	93%	84%	85%	87%	90%
		16	85%	91%	93%	96%	82%	82%	81%	84%
		20	88%	82%	88%	94%	84%	79%	81%	84%
	Absent	12	96%	80%	89%	95%	80%	83%	81%	82%
		16	86%	89%	87%	86%	82%	81%	85%	89%
		20	88%	87%	87%	88%	76%	74%	72%	77%
Words	Present	12	86%	81%	83%	77%	83%	74%	81%	71%
		16	83%	86%	82%	87%	64%	73%	68%	78%
		20	78%	83%	78%	90%	66%	67%	70%	81%
	Absent	12	86%	80%	76%	82%	74%	70%	63%	79%
		16	71%	79%	70%	89%	59%	64%	80%	69%
		20	78%	70%	75%	86%	70%	68%	75%	67%

Table 15.

Target-absent visual search accuracy from Experiment 5b, presented as a function of Cue Type, Cue Reminders, Set Size, Target Load, and Search Length.

			1-target				2-targets			
Cue Type	Cue Reminders	Set Size	Search 1	Search 2	Search 3	Search 4	Search 1	Search 2	Search 3	Search 4
Pictures	Present	12	100%	100%	100%	100%	99%	100%	100%	100%
		16	100%	100%	100%	99%	99%	100%	100%	100%
		20	99%	100%	100%	100%	99%	100%	98%	100%
	Absent	12	98%	100%	99%	99%	99%	99%	99%	100%
		16	99%	99%	99%	100%	99%	99%	97%	100%
		20	100%	98%	99%	99%	99%	99%	99%	97%
Words	Present	12	98%	98%	99%	99%	98%	99%	99%	96%
		16	98%	97%	97%	100%	94%	99%	99%	99%
		20	98%	98%	99%	98%	97%	97%	98%	96%
	Absent	12	98%	98%	97%	97%	97%	97%	97%	97%
		16	98%	98%	98%	99%	95%	97%	98%	99%
		20	97%	97%	97%	97%	97%	97%	95%	98%

Table 16.

Target-present visual search RTs from Experiment 5b, presented as a function of Cue Type, Cue Reminders, Set Size, Target Load, and Search Length.

			1-target				2-targets			
Cue Type	Cue Reminders	Set Size	Search 1	Search 2	Search 3	Search 4	Search 1	Search 2	Search 3	Search 4
Pictures	Present	12	966	936	947	937	1220	1464	1326	1393
		16	997	1094	977	1053	1260	1690	1432	1312
		20	1015	1018	1258	1147	1527	1531	1544	1382
	Absent	12	974	1022	1039	1114	1445	1415	1483	1425
		16	968	1213	1072	1068	1268	1676	1638	1852
		20	1166	1234	1392	1231	1486	1868	1811	1725
Words	Present	12	1876	1623	1904	1793	2528	2297	2345	2150
		16	1996	2236	2103	2169	2345	3082	3285	2579
		20	2254	2172	2537	2503	3031	3183	3231	3150
	Absent	12	1816	2001	1651	1790	2110	2503	2510	2496
		16	1872	2012	2400	2330	2558	2839	3349	2798
		20	1980	2271	2408	2205	2640	3207	3389	2792

Table 17.

Target-absent visual search RTs from Experiment 5b, presented as a function of Cue Type, Cue Reminders, Set Size, Target Load, and Search Length.

			1-target				2-targets			
Cue Type	Cue Reminders	Set Size	Search 1	Search 2	Search 3	Search 4	Search 1	Search 2	Search 3	Search 4
Pictures	Present	12	1405	1410	1444	1516	2014	2071	2211	2197
		16	1619	1643	1708	1743	2413	2320	2481	2616
		20	1852	1834	1988	1986	2769	2780	2874	3398
	Absent	12	1512	1584	1701	1825	2129	2289	2527	2525
		16	1677	1774	2005	1995	2496	2702	2942	3141
		20	1964	2070	2214	2198	2862	3065	3144	3285
Words	Present	12	2834	2848	2995	3202	3731	3645	3869	4039
		16	3614	3495	3688	3745	4501	4473	4620	4894
		20	3934	3913	4204	4086	5222	5131	5425	5798
	Absent	12	2757	2861	3064	2891	3717	3763	4090	3934
		16	3291	3343	3442	3305	4486	4298	4608	4615
		20	3787	3820	4053	4027	4989	5053	5402	5580

FIGURES

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

Figure 2. Mean reaction time (on accurate trials) from Experiments 1a and 1b. Error bars represent 1 standard error (SE) of the mean.

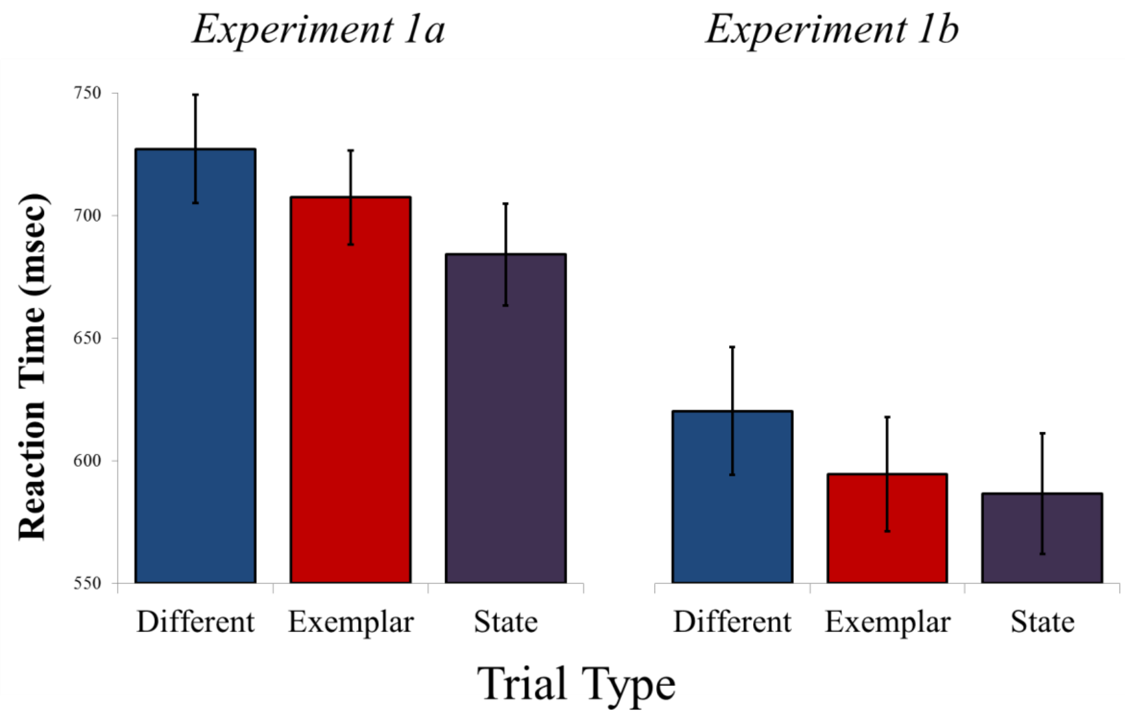


Figure 3. Trial progression from Experiment 1b. Dotted lines on the final display indicate alternative target locations; outlines were not actually shown.

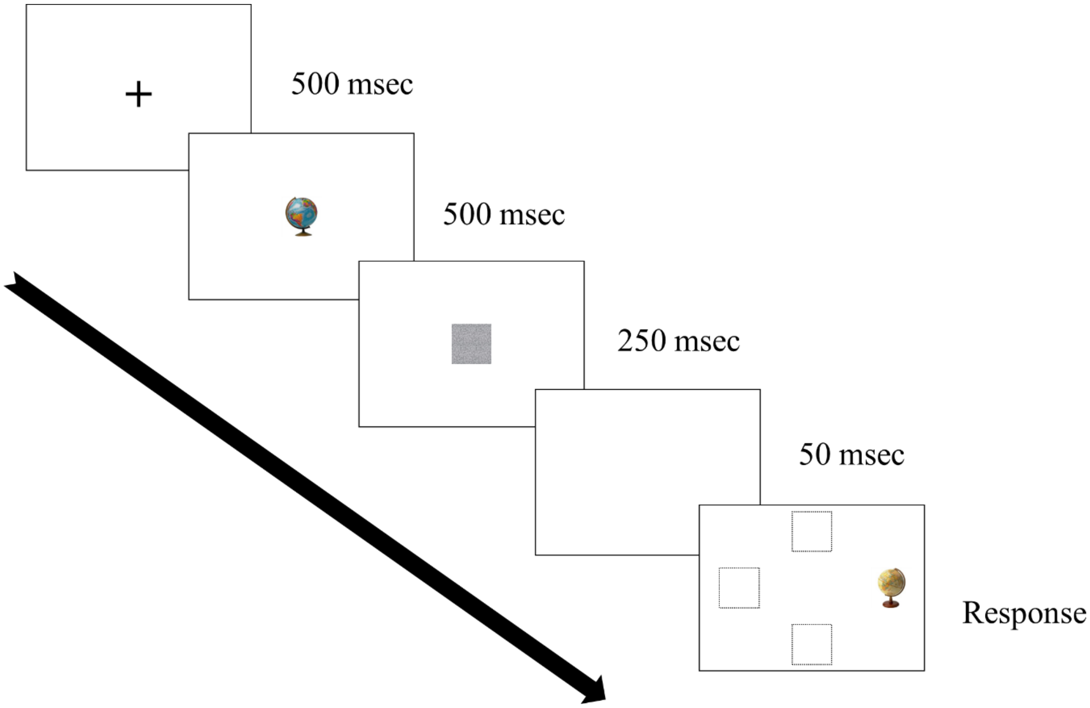


Figure 4. Scree plots for the teddy bear and butterfly categories, from Experiment 2. Stress values are plotted as a function of the dimensionality in which the MDS data were scaled.

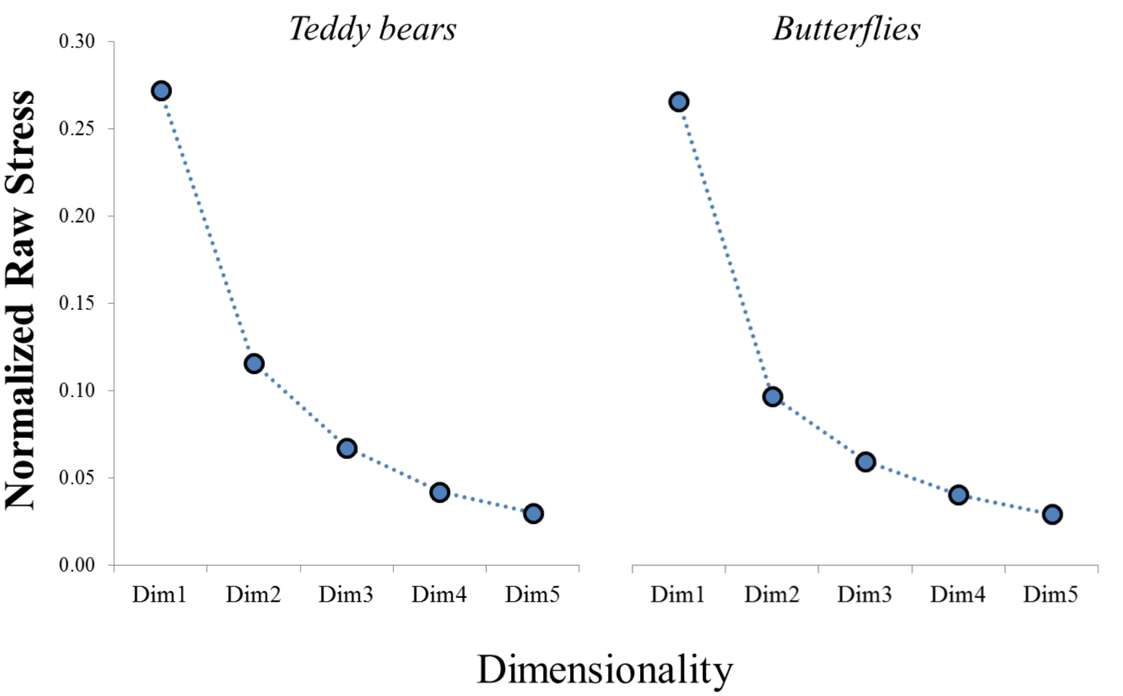


Figure 5. MDS solutions for the teddy bear stimuli, from Experiment 2.

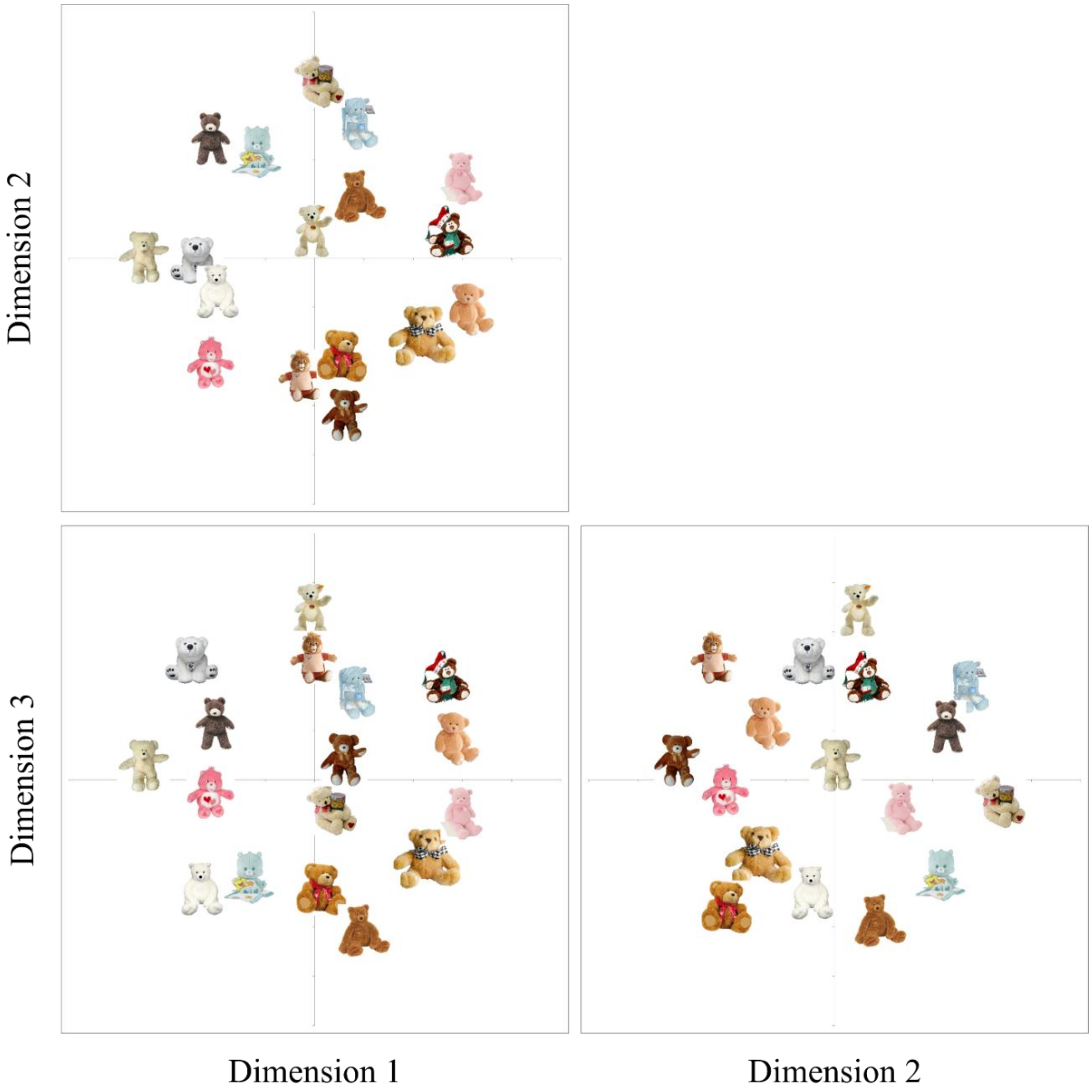


Figure 6. MDS solutions for the butterfly stimuli, from Experiment 2.

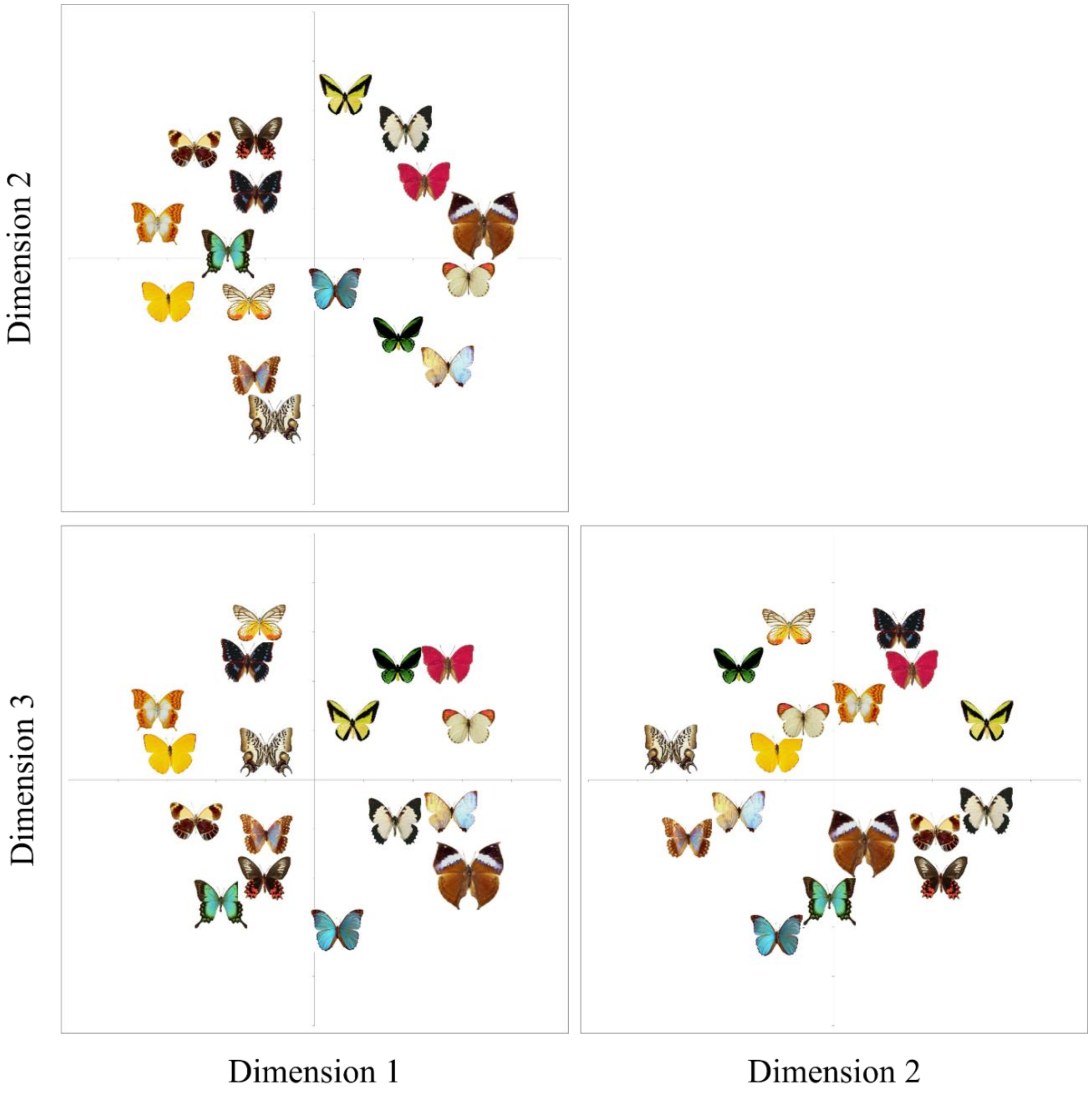
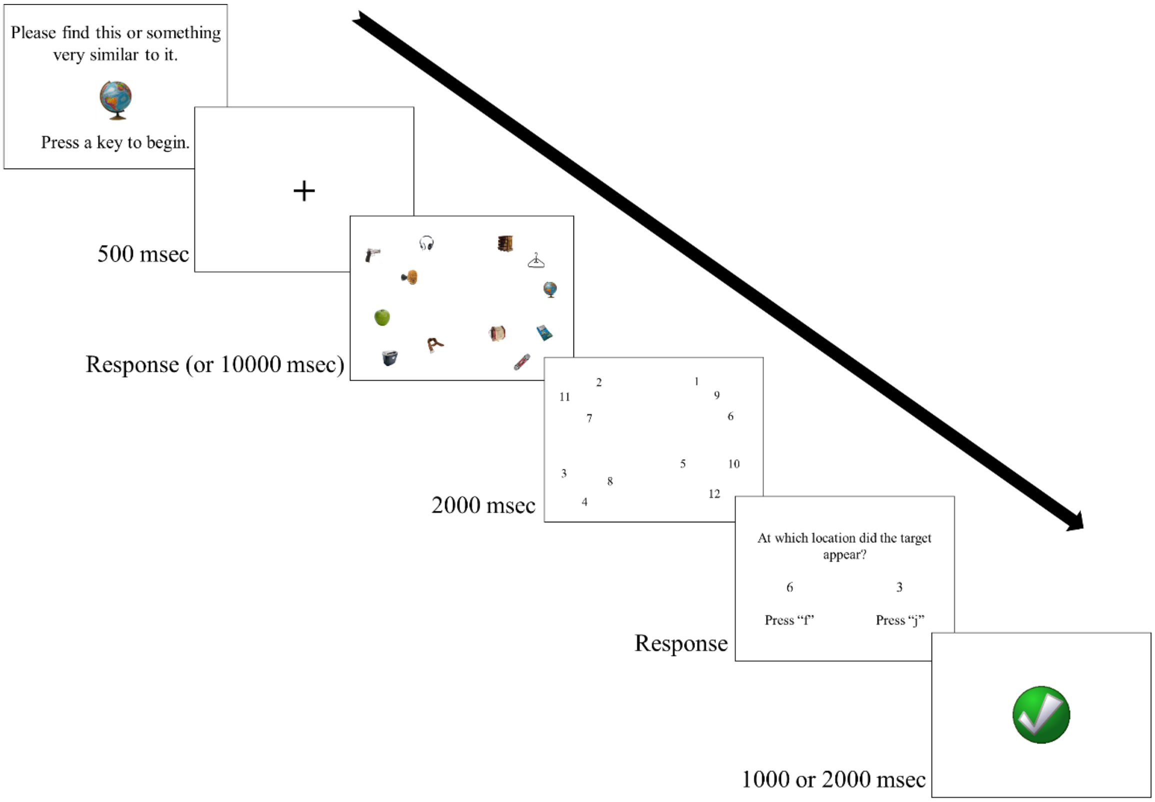


Figure 7. Visual search trial progression, from Experiment 3a.



Please find this or something very similar to it.

Press a key to begin.

At which location did the target appear?

Figure 8. Sample visual search display, from Experiment 3a, 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.

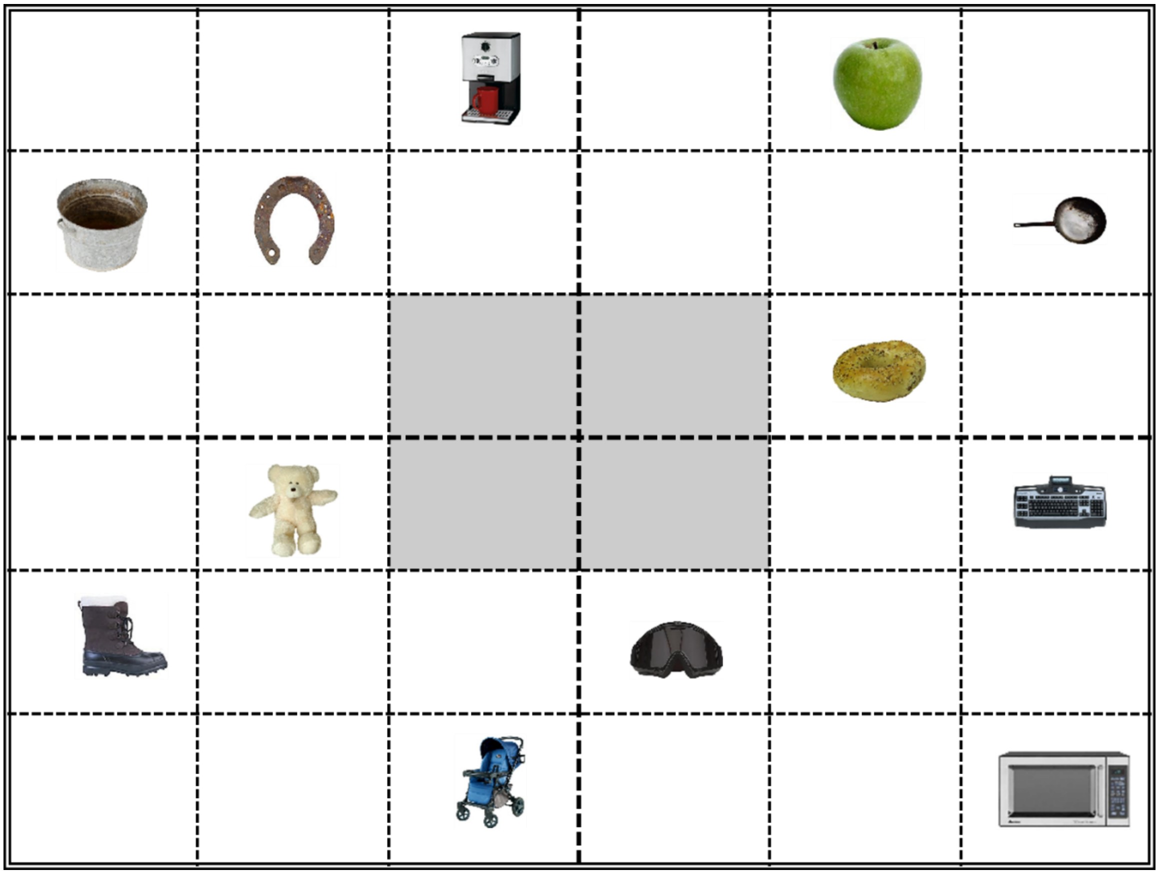


Figure 9. Mean visual search reaction time (on accurate trials) from Experiment 3a, presented as a function of Template Precision and Block. Error bars represent 1 standard error (SE) of the mean.

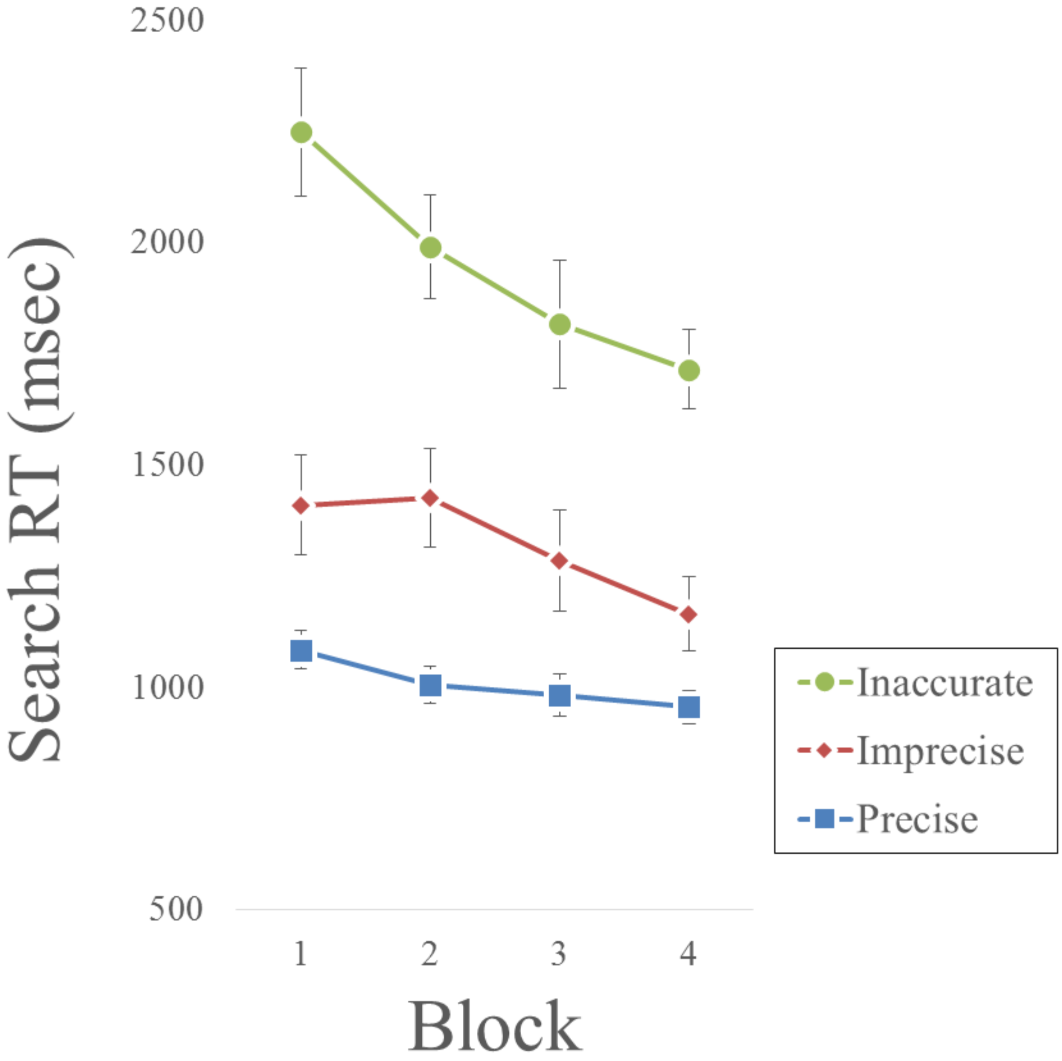


Figure 10. Mean visual search reaction time, scan-path ratio, and decision time (on accurate trials) from Experiment 3b, presented as a function of Template Precision and Block. Error bars represent 1 standard error (SE) of the mean.

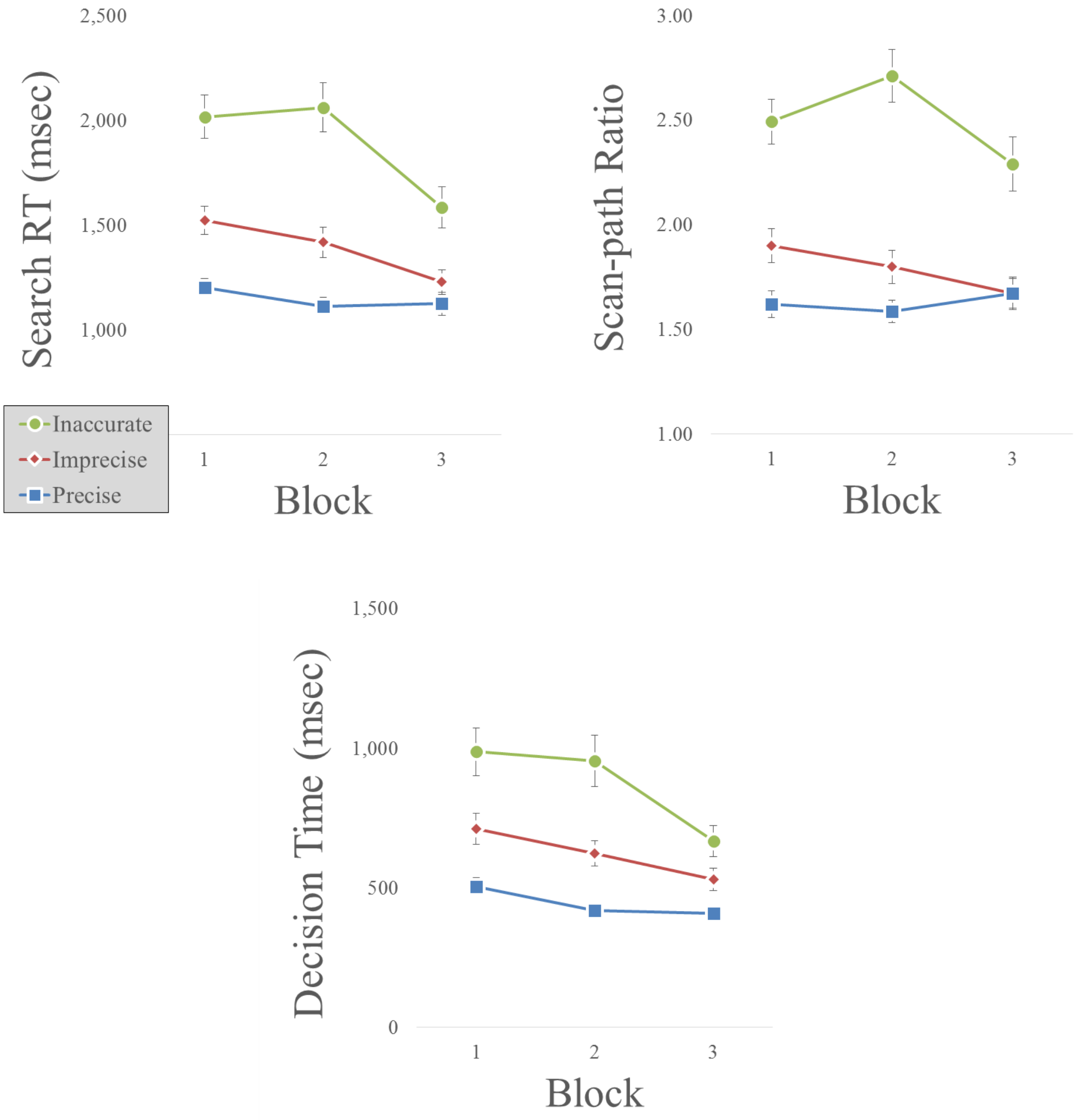


Figure 11. Mean visual search reaction time from Experiments 3c and 3d, presented as a function of Template Precision. Error bars represent 1 standard error (SE) of the mean.

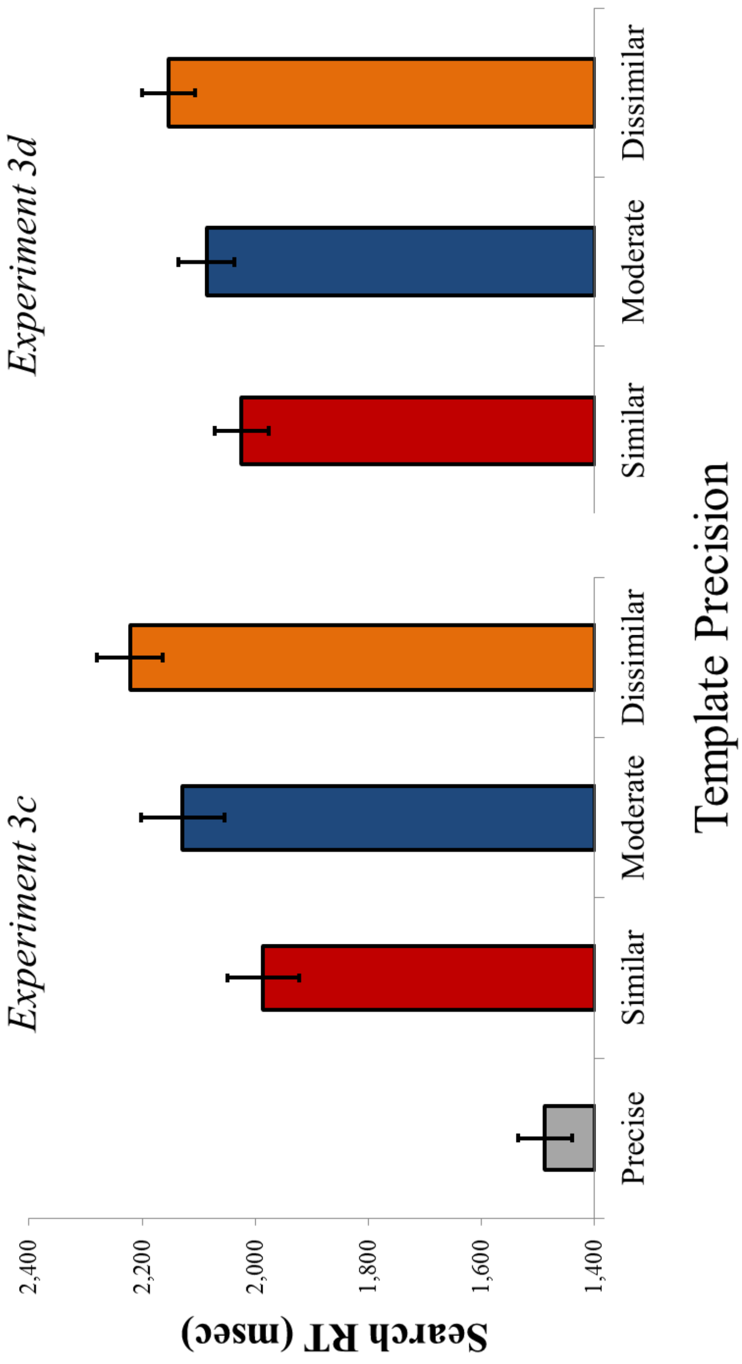


Figure 12. Mean visual search reaction time, scan-path ratio, and decision time (on accurate trials) from Experiments 3e and 3f, presented as a function of Template Precision. Error bars represent 1 standard error (SE) of the mean.

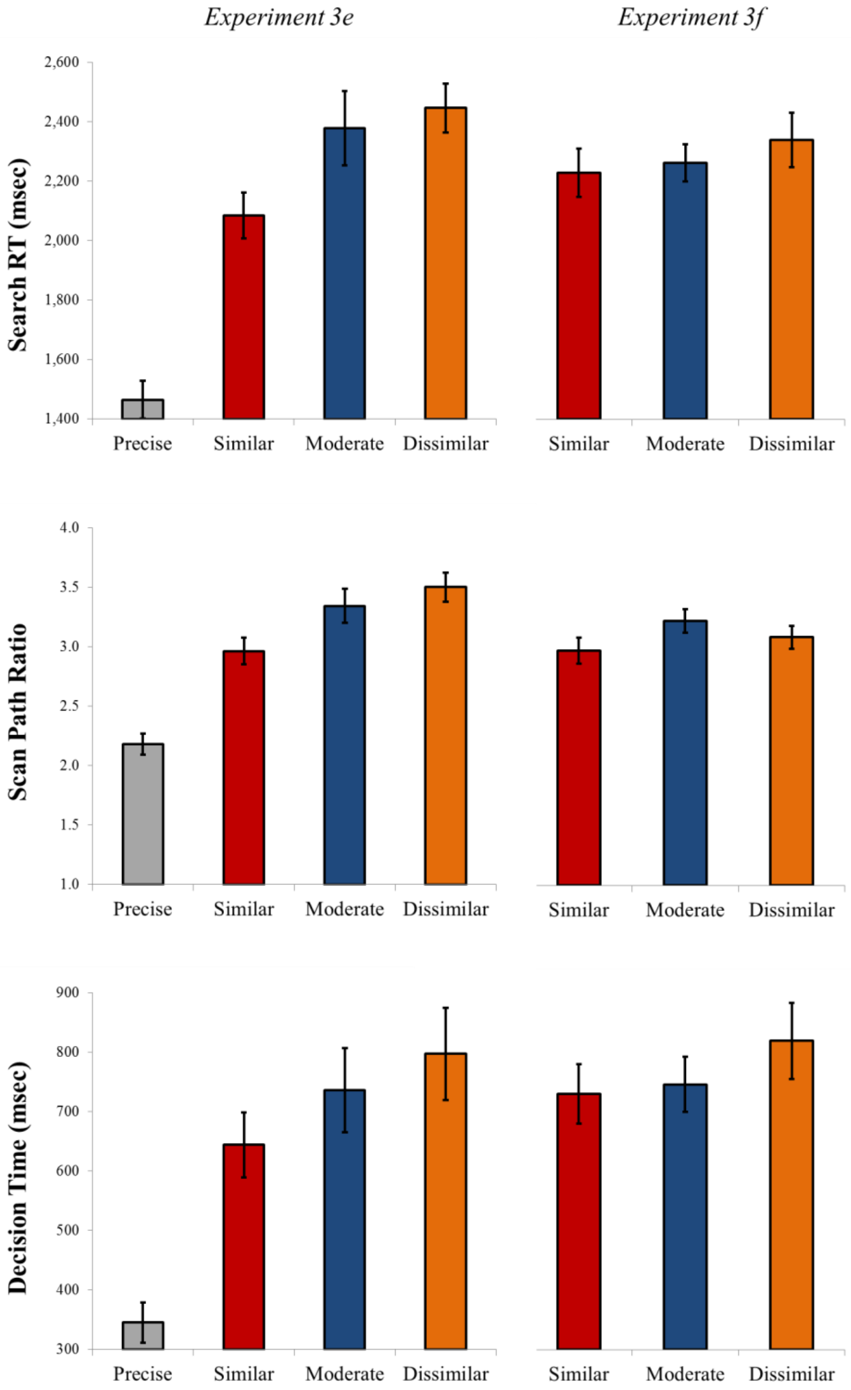


Figure 13. Sample single- and two-target visual search cues, from Experiment 4. Participants saw one or a pair of images and were instructed to find one of them. In the *precise* Feature Space condition, only a single item was shown as the cue. For the *narrow* condition a state-pair was used, and for the *wider* condition an exemplar-pair was selected. In the *widest* condition, two images were selected from different semantic categories.

Figure 14. Mean visual search reaction time (on accurate trials) from Experiment 4a, presented as a function of Feature Width. Error bars represent 1 standard error (SE) of the mean.

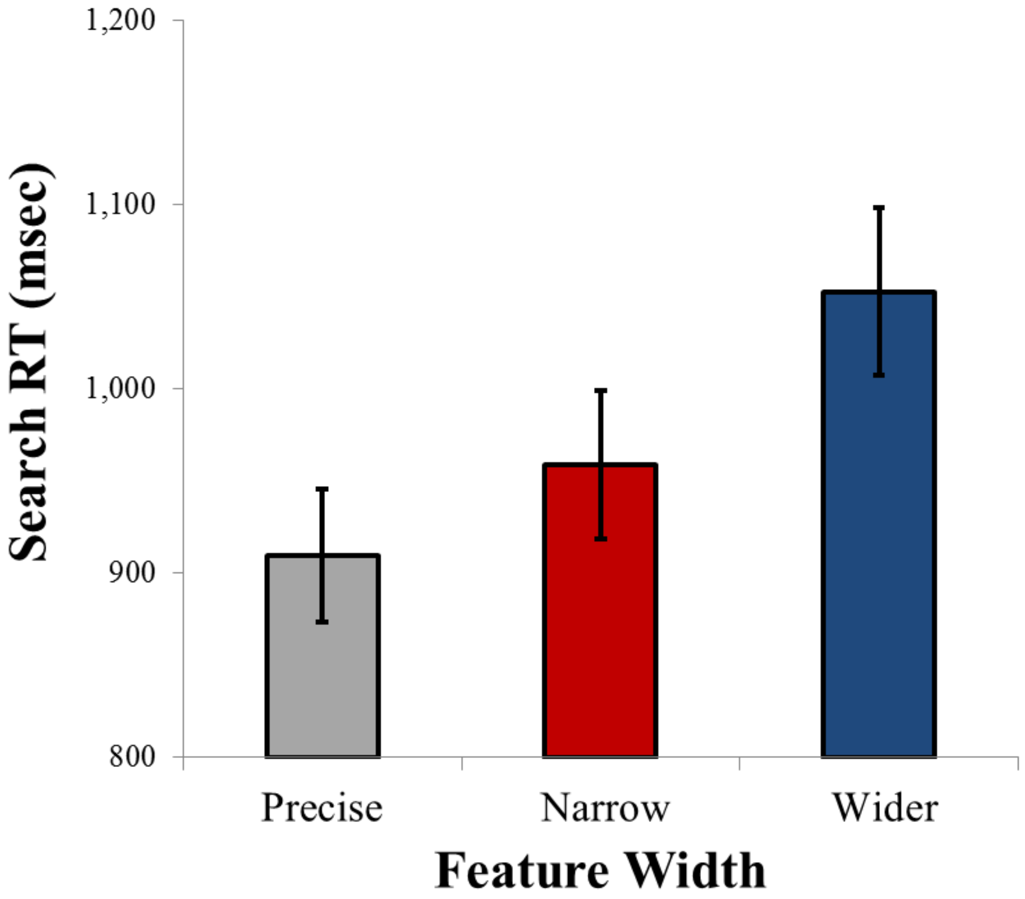


Figure 15. Mean visual search reaction time, scan-path ratio, and decision time (on accurate trials) from Experiment 4b, presented as a function of Feature Width. Error bars represent 1 standard error (SE) of the mean.

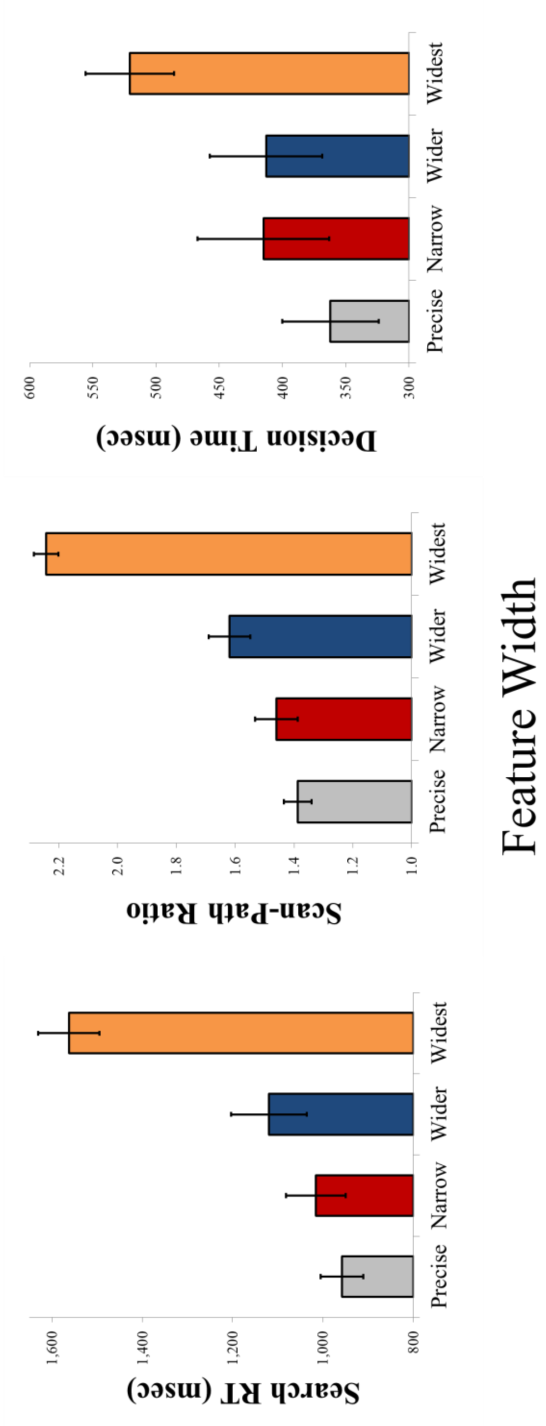


Figure 16. Mean visual search reaction time (on accurate trials) from Experiment 4c, presented as a function of Feature Width. Error bars represent 1 standard error (SE) of the mean.

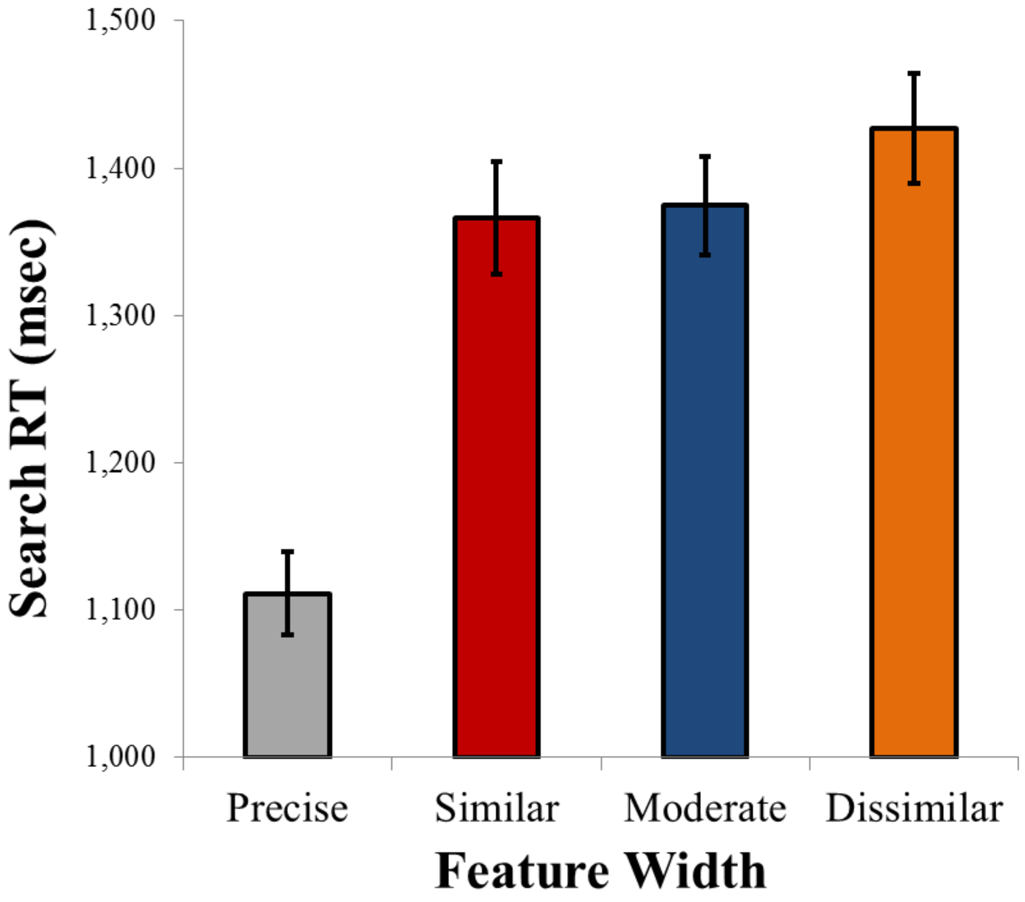


Figure 17. Mean visual search reaction time, scan-path ratio, and decision time (on accurate trials) from Experiment 4d, presented as a function of Feature Width. Error bars represent 1 standard error (SE) of the mean.

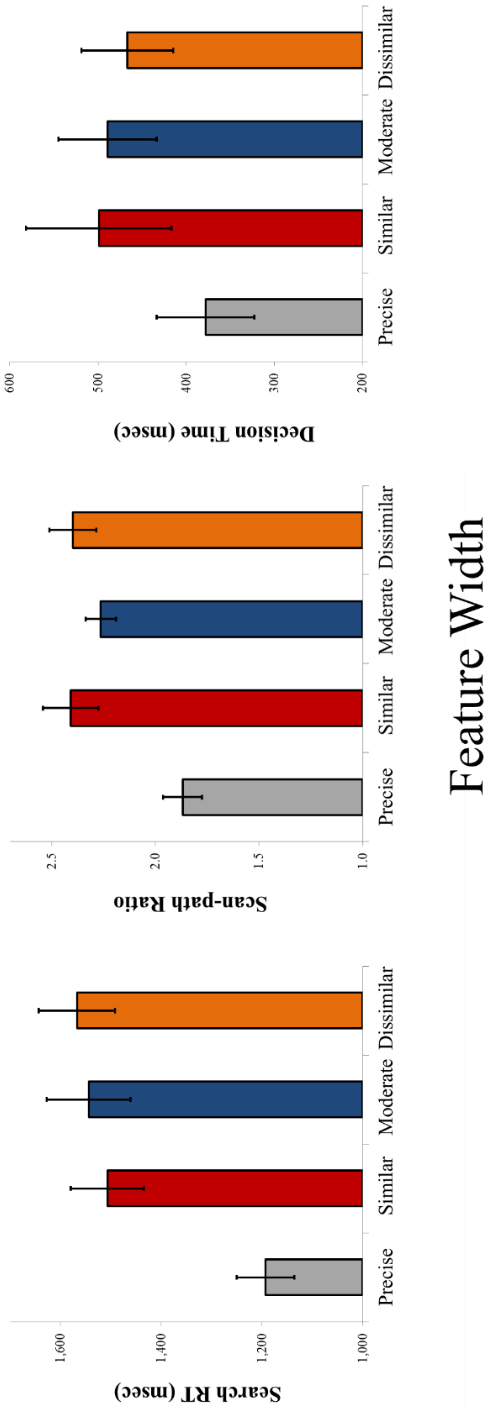


Figure 18. Visual search trial progression, from Experiment 5.

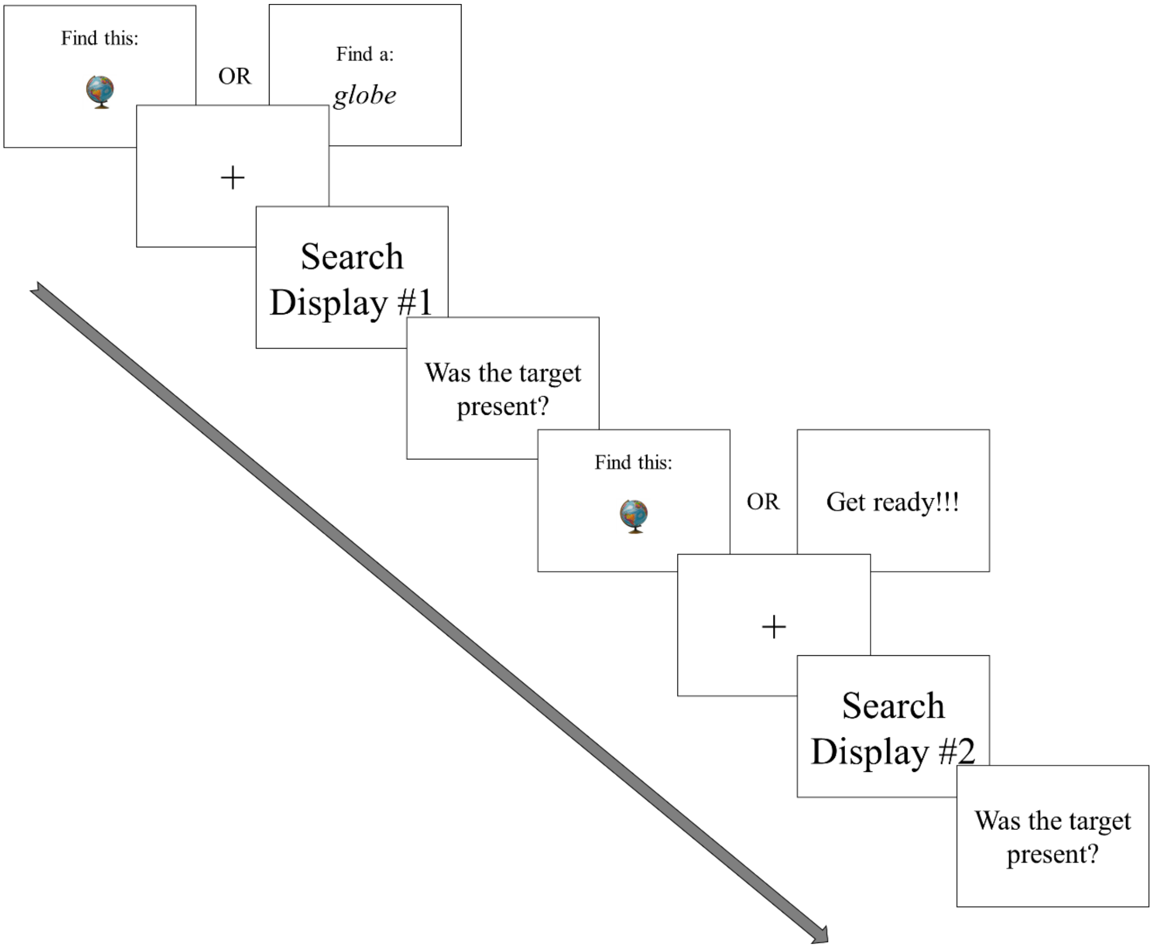


Figure 19. Target-present visual search accuracy, from Experiment 5a, presented as a function of Cue Type, Target Load, and Cue Reminders. Error bars represent 1 standard error (SE) of the mean.

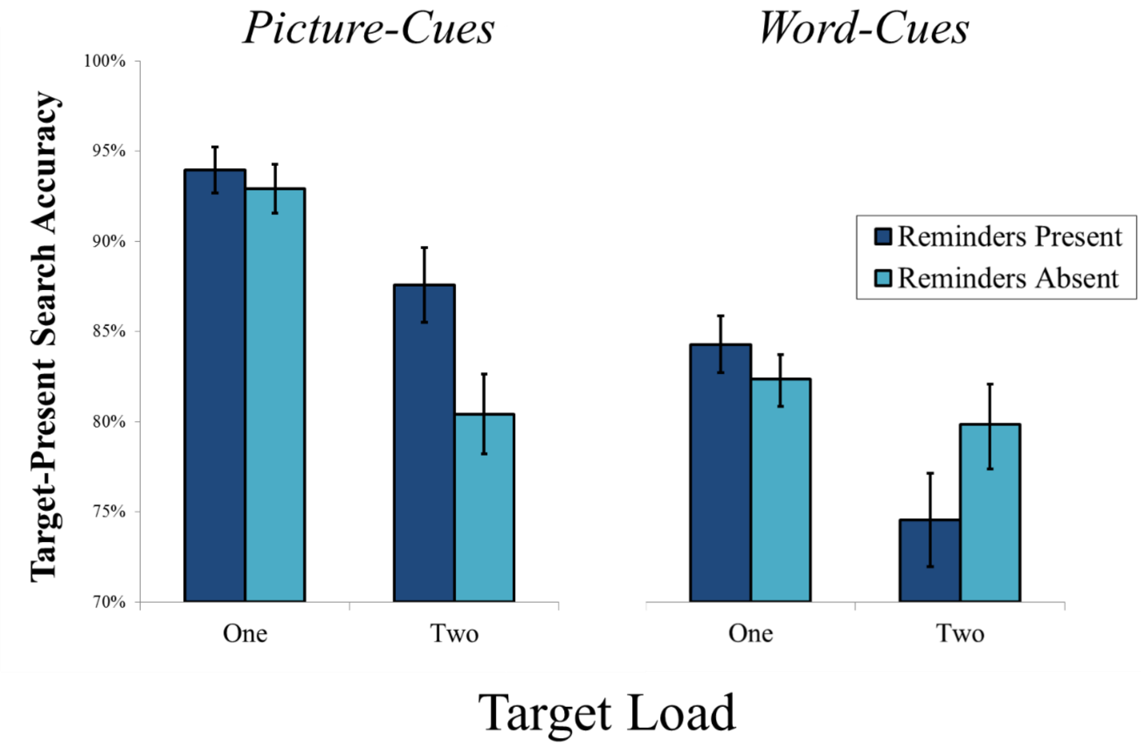


Figure 20. Target-absent visual search accuracy, from Experiment 5a, presented as a function of Cue Type, Target Load, and Cue Reminders. Error bars represent 1 standard error (SE) of the mean.

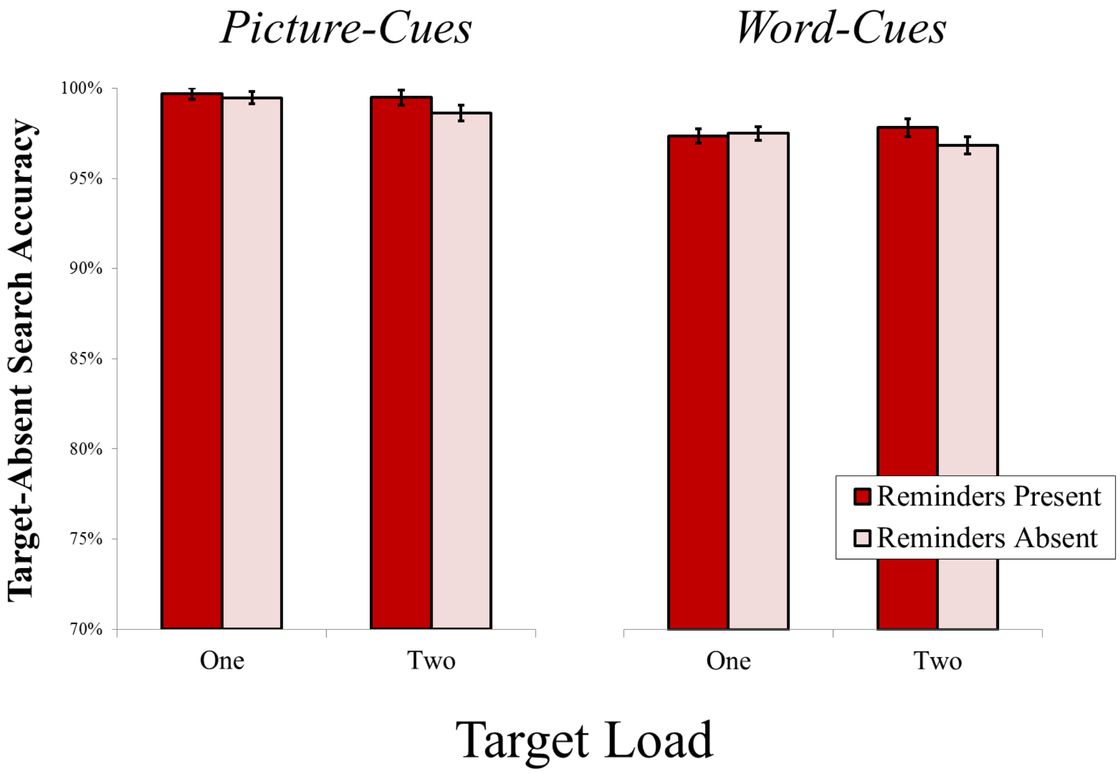


Figure 21. Target-present visual search RTs (on correct trials), from Experiment 5a, presented as a function of Cue Type, Cue Reminders, Target Load, and Search Length. Error bars represent 1 standard error (SE) of the mean.

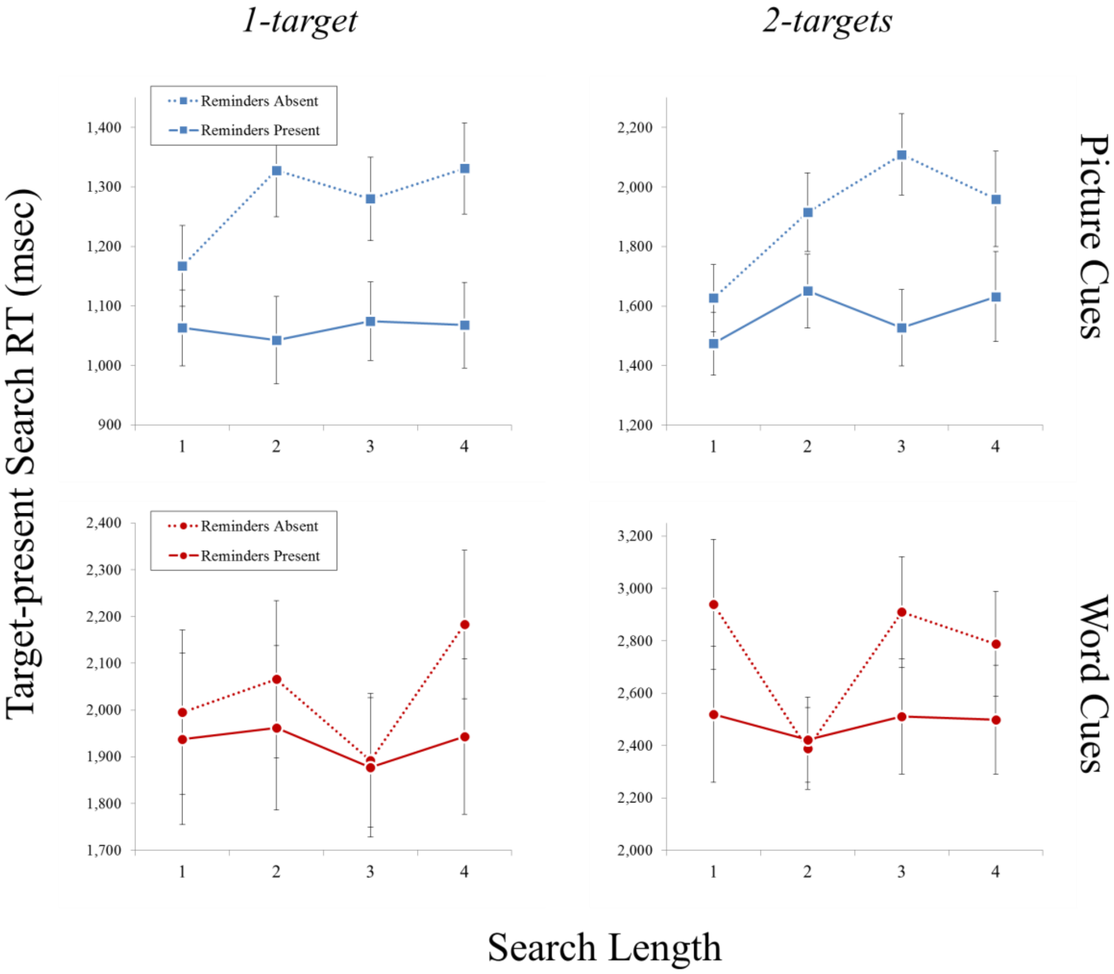


Figure 22. Target-absent visual search RTs (on correct trials), from Experiment 5a, presented as a function of Cue Type, Cue Reminders, Target Load, and Search Length. Error bars represent 1 standard error (SE) of the mean.

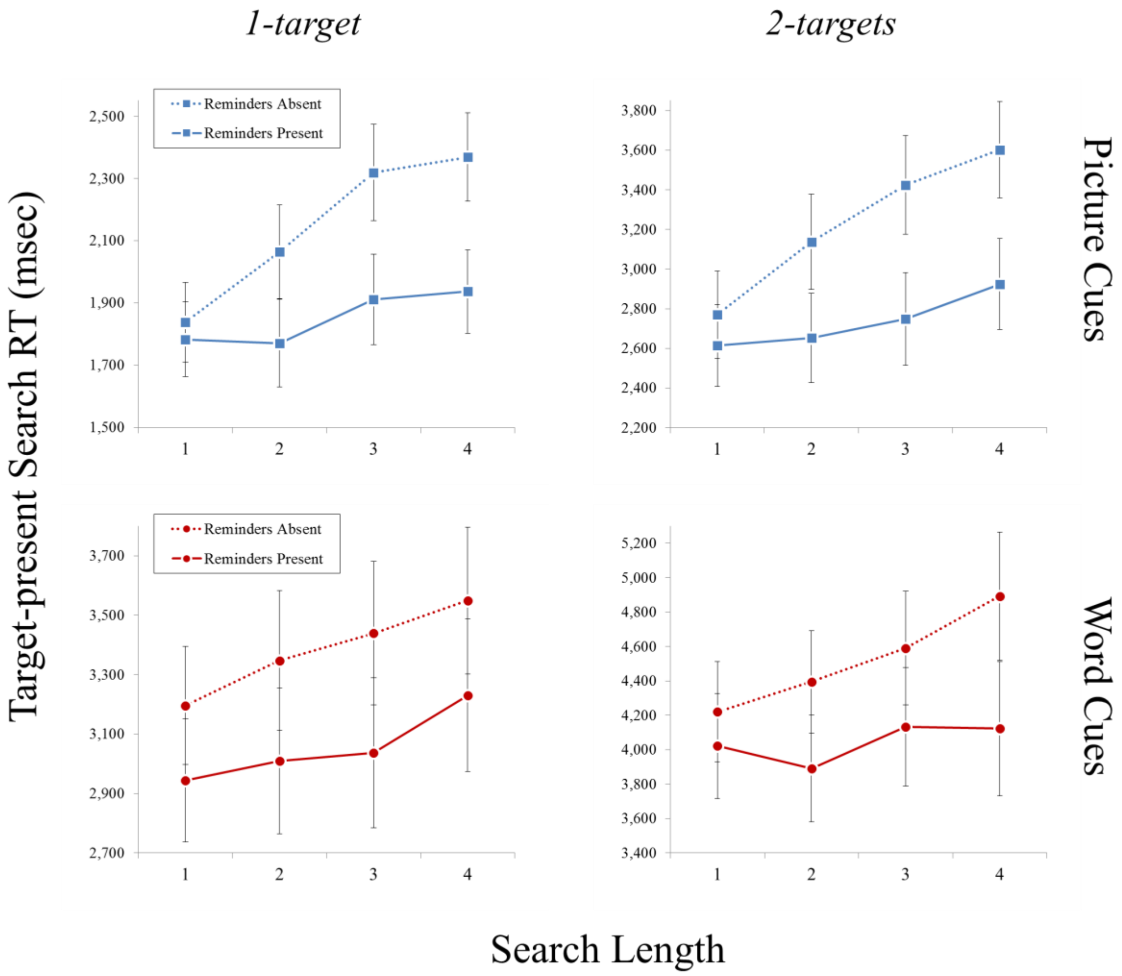


Figure 23. Target-present visual search accuracy, from Experiment 5b, presented as a function of Cue Type, Target Load, and Cue Reminders. Error bars represent 1 standard error (SE) of the mean.

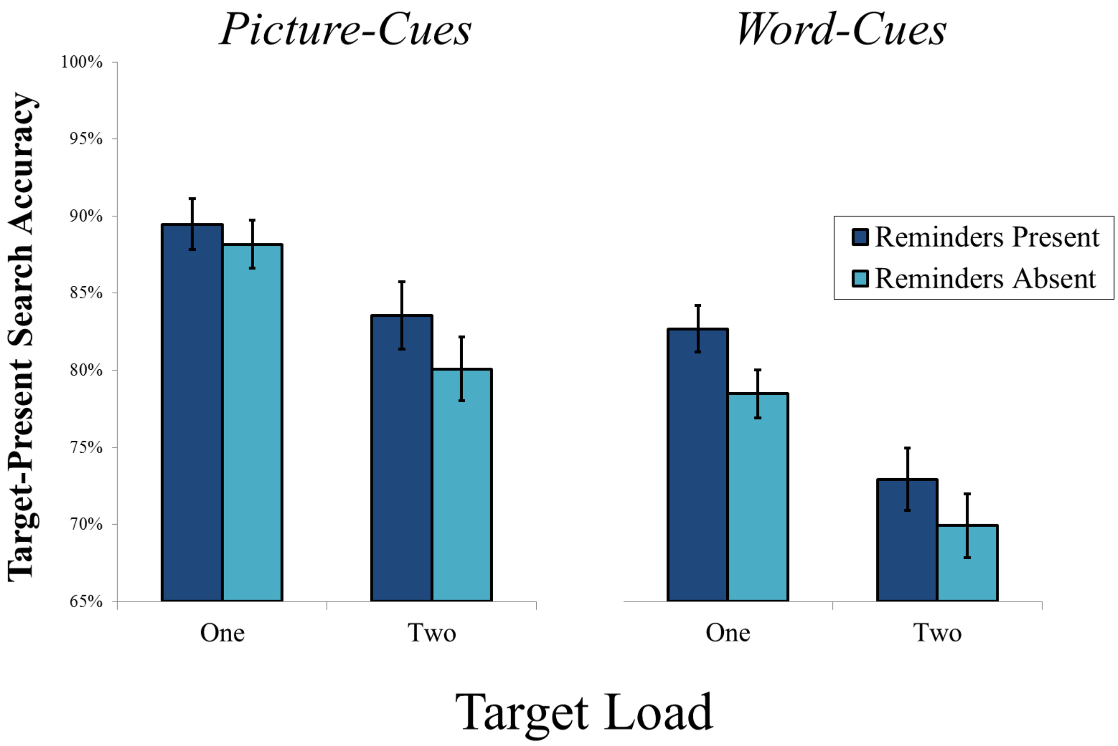


Figure 24. Target-absent visual search accuracy, from Experiment 5b, presented as a function of Cue Type, Target Load, and Cue Reminders. Error bars represent 1 standard error (SE) of the mean.

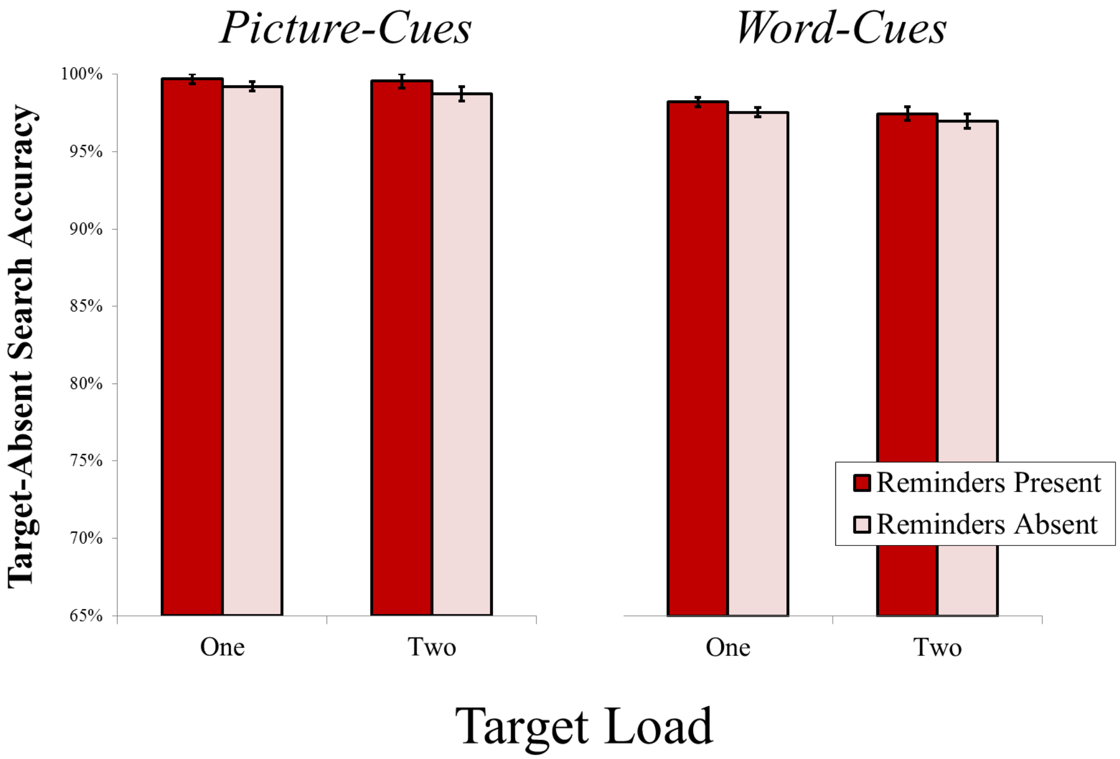


Figure 25. Target-present visual search RTs (on correct trials), from Experiment 5b, presented as a function of Cue Type, Cue Reminders, Target Load, and Search Length. Error bars represent 1 standard error (SE) of the mean.

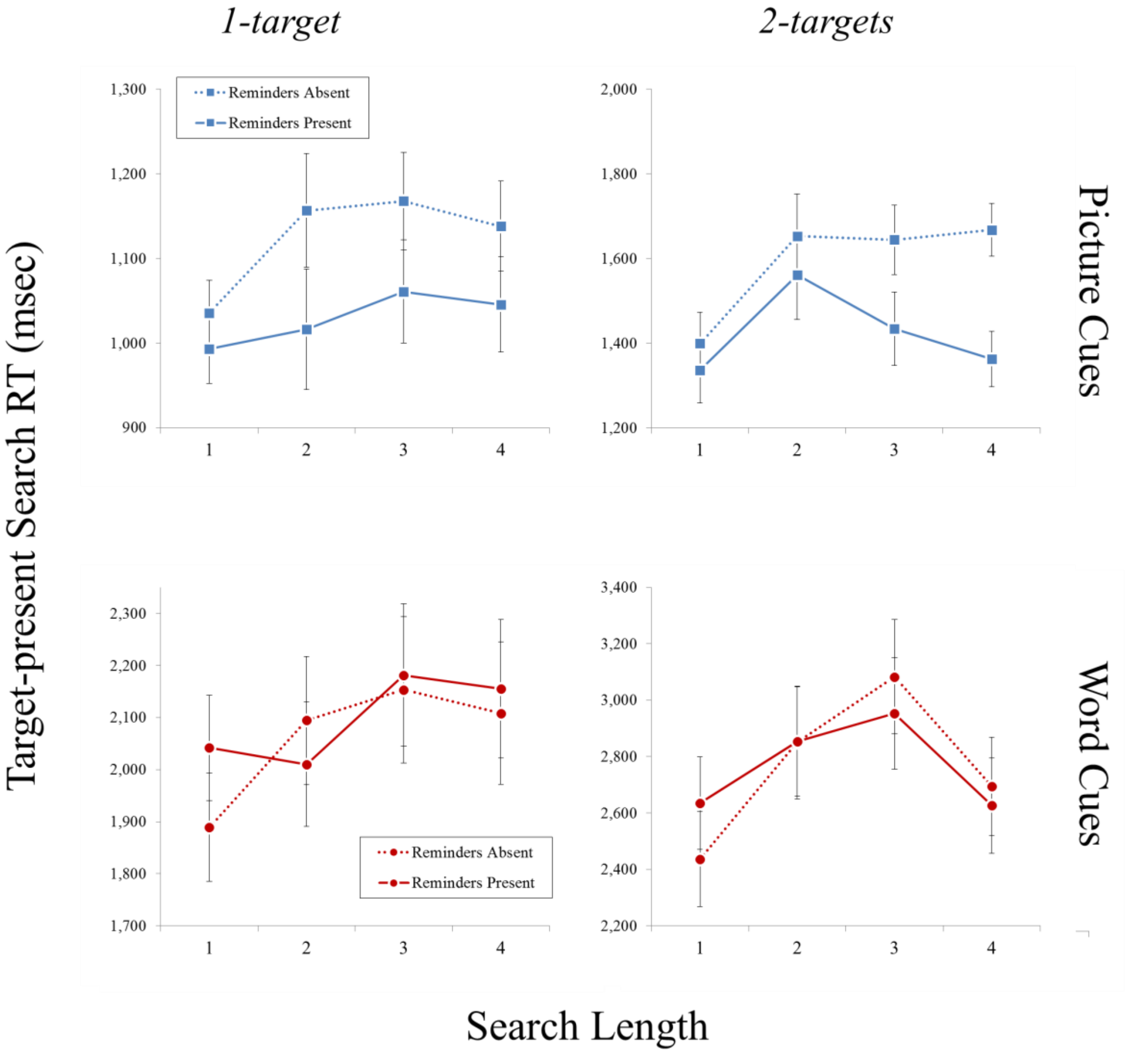


Figure 26. Target-absent visual search RTs (on correct trials), from Experiment 5b, presented as a function of Cue Type, Cue Reminders, Target Load, and Search Length. Error bars represent 1 standard error (SE) of the mean.

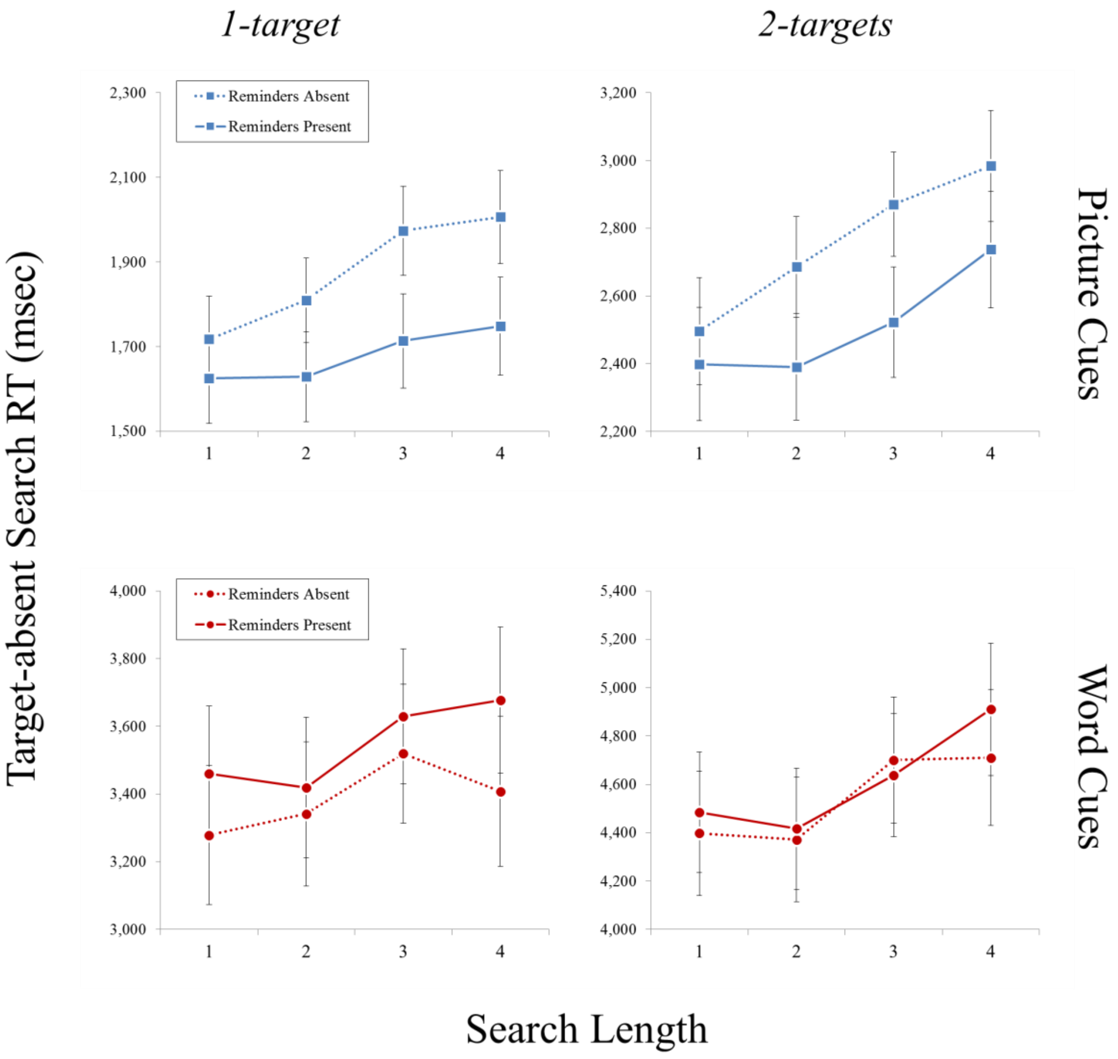


Figure 27. Pupil dilation, plotted as a function of the number of targets searched for (1 target, bottom two lines; 3 targets, top two lines), and trial type (target-present, solid lines; target-absent, dotted lines). On the X-axis time is represented from the start of the trial to the end; each quartile represents 6 items being presented in the stream. From Hout, Papesh, & Goldinger, 2012.

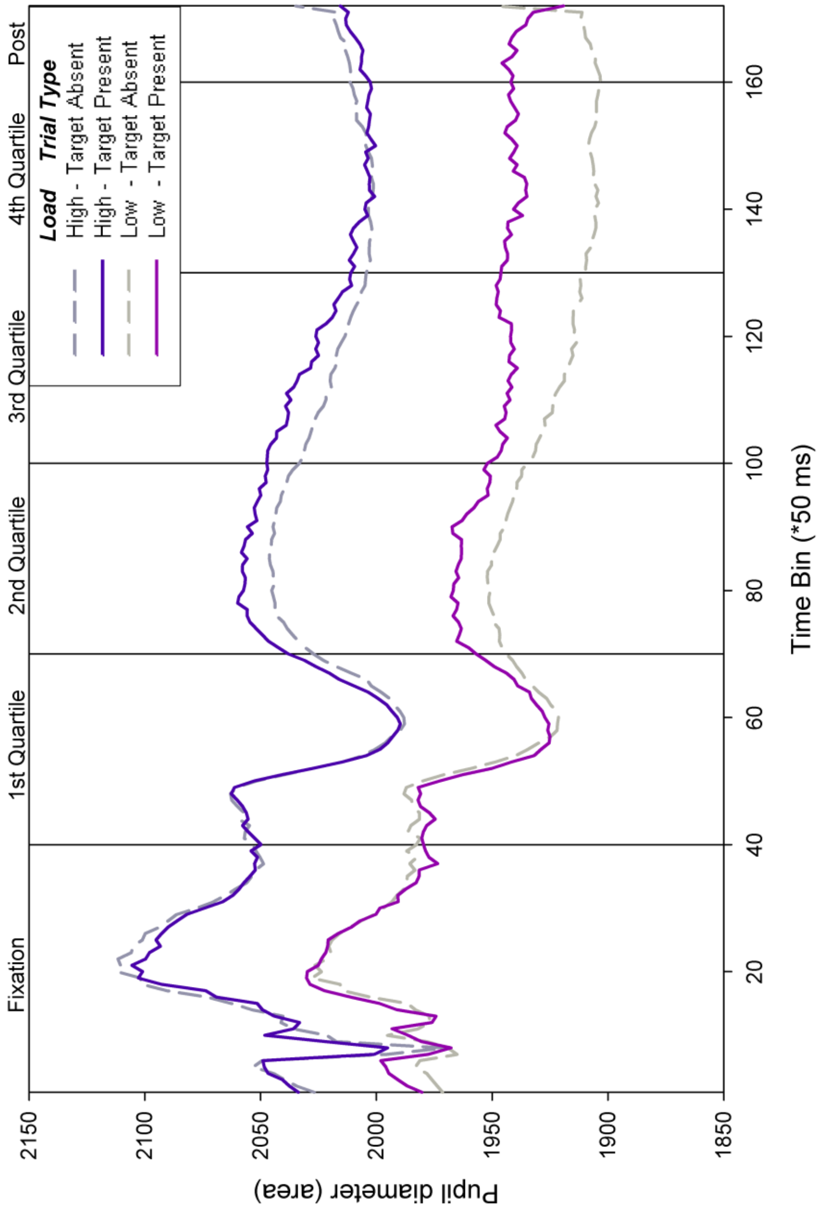


Figure 28. Pupil dilation, plotted as a function of cue type (picture cues in purple; word cues in green). These are target-present trials wherein the target was one of the first six items in the stream. From Hout, Papesh, & Goldinger, 2012.

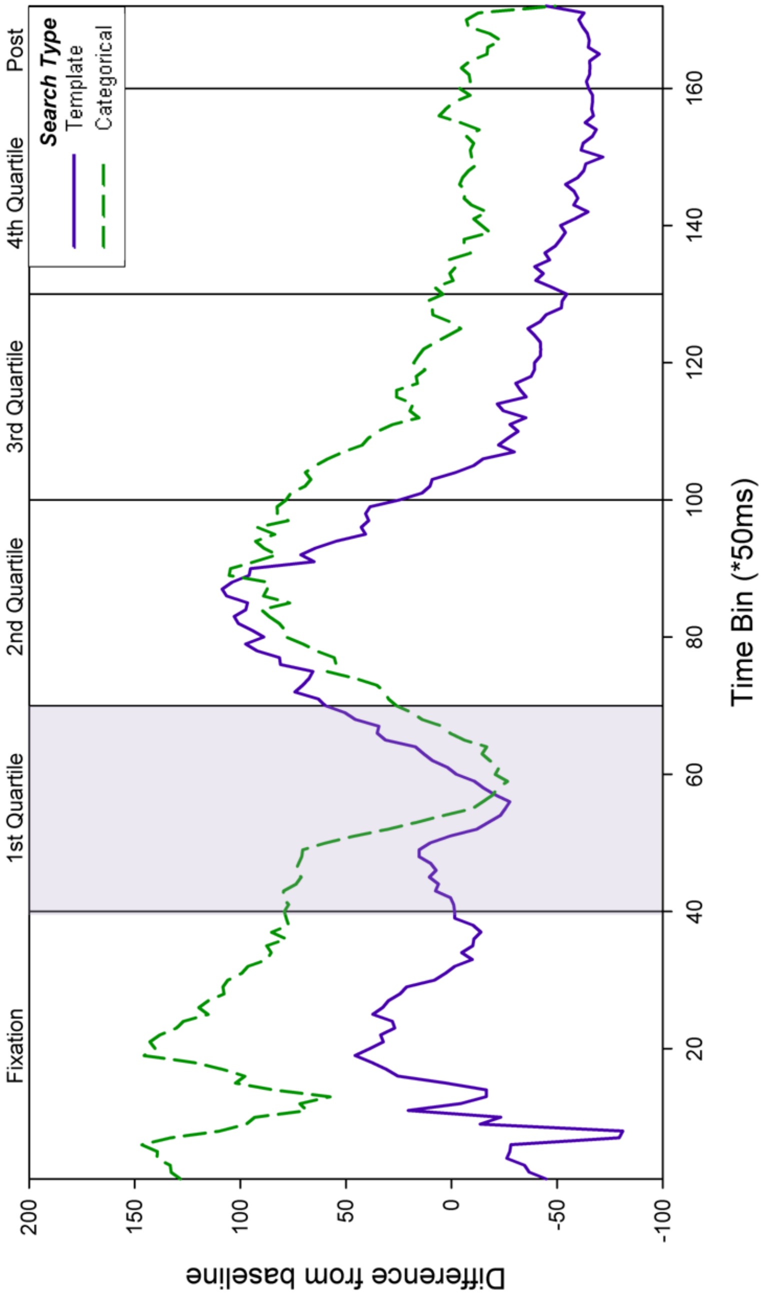
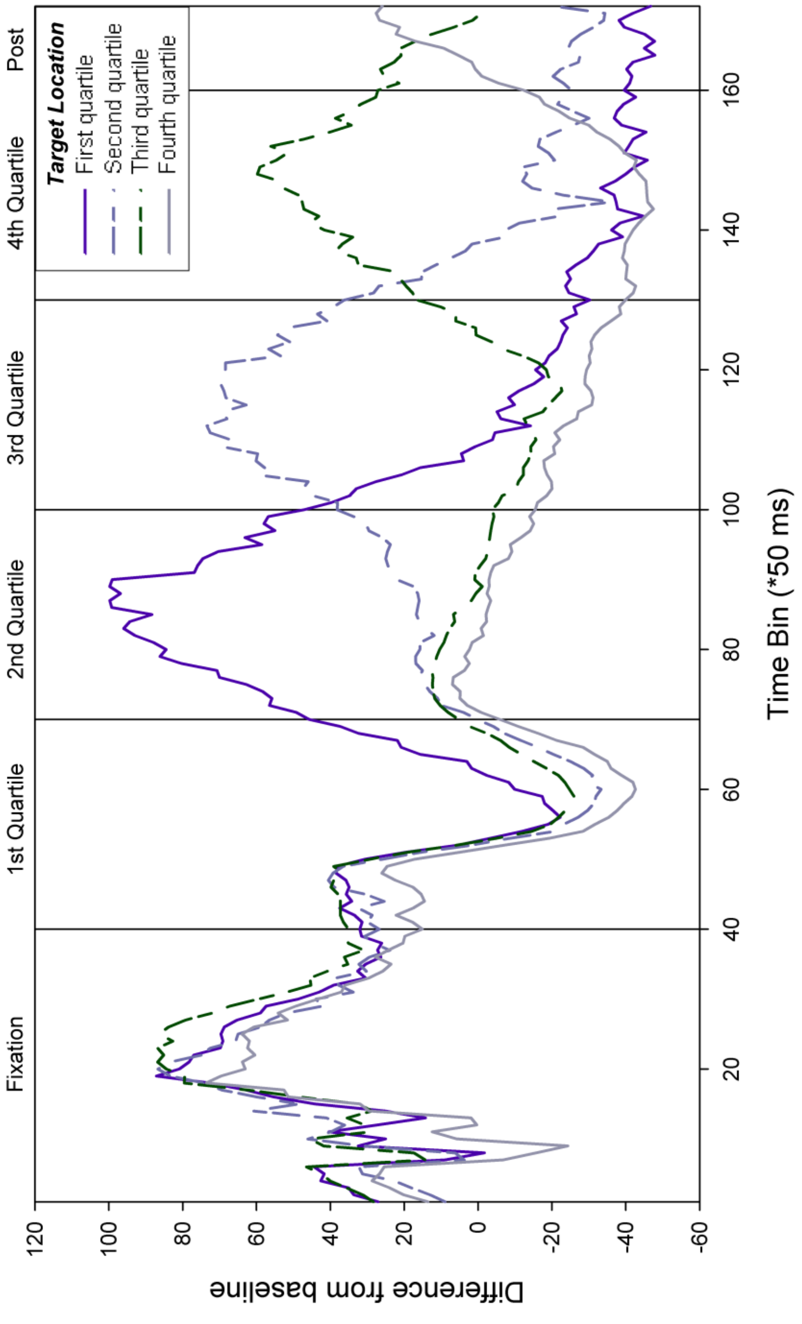


Figure 29. Pupil dilation, plotted as a function of when the target appeared (first, second, third or fourth quartile), collapsed across all other factors (target-absent trials excluded). From Hout, Papesh, & Goldinger, 2012.



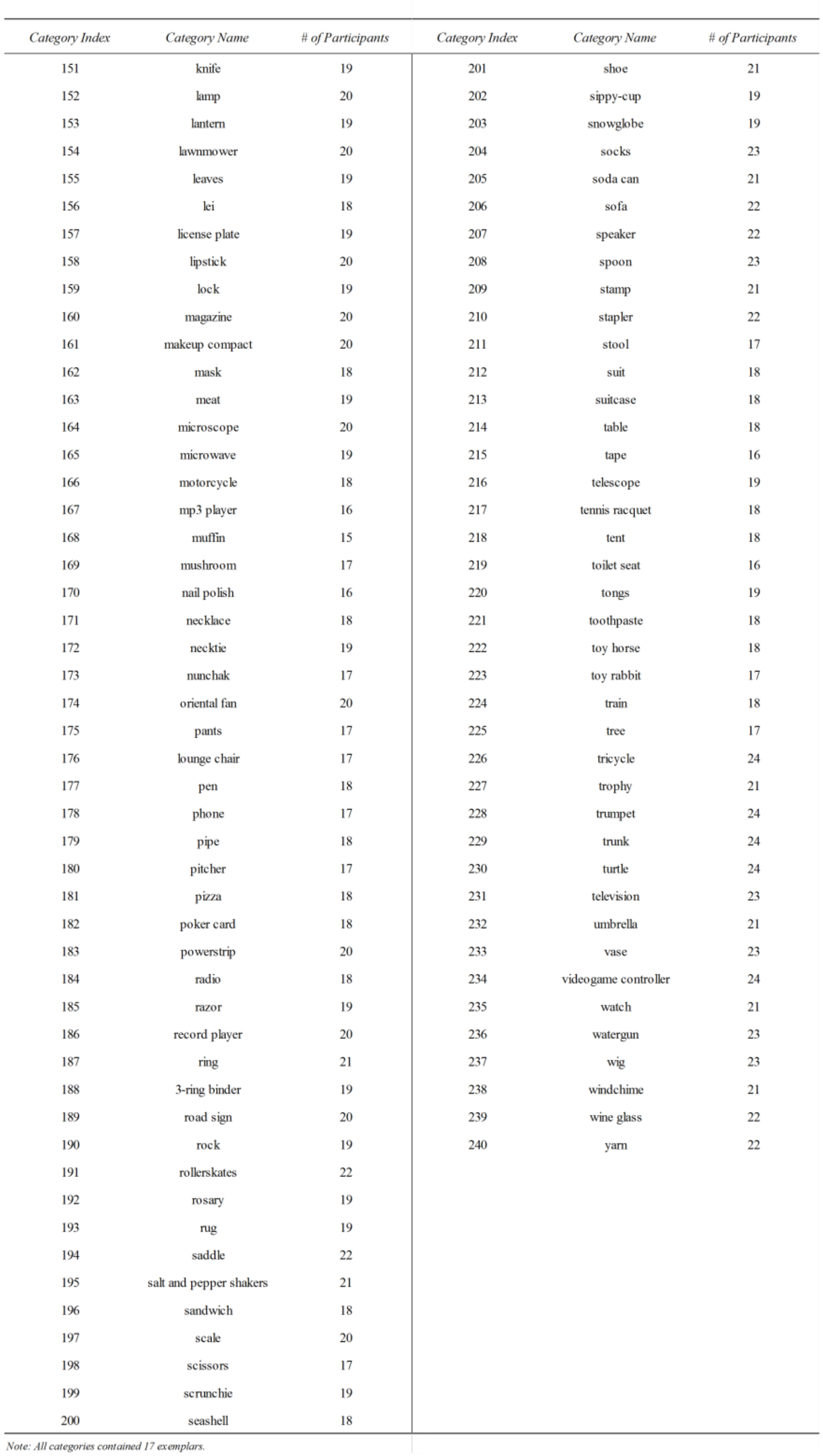
APPENDIX A

Table A1. *Stimulus categories 1-150, from Experiment 2. Displayed are the number of participants who contributed data to each set.*

Category Index	*Category Name*	*# of Participants*	*Category Index*	*Category Name*	*# of Participants*	*Category Index*	*Category Name*	*# of Participants*
1	playpen	21	51	balloon	19	101	cookie	20
2	birdhouse	21	52	barbie doll	19	102	cooking pan	18
3	bowling pin	23	53	baseball card	20	103	cooking pot	18
4	carabiner	21	54	basket	17	104	crib	17
5	chocolate	19	55	bathing suit	17	105	cup and saucer	18
6	circuit board	21	56	beanbag chair	20	106	cushion	19
7	coffee maker	21	57	teddy bear	19	107	decorative screen	20
8	cooler	19	58	bed	17	108	desk	18
9	door knocker	22	59	beer mug	17	109	dog	20
10	feather	24	60	bell	19	110	doll	18
11	fruit parfait	21	61	bench	18	111	dollhouse	20
12	board game	22	62	bicycle	17	112	domino	18
13	eyeglasses	23	63	dollar bill	17	113	donut	18
14	plant	22	64	binoculars	19	114	doorknob	20
15	hair dryer	20	65	bird	18	115	dresser	19
16	highchair	21	66	bongo	17	116	dumbbell	17
17	highlighter	20	67	bonsai tree	18	117	earrings	19
18	horseshoe	21	68	boot	18	118	easter egg	20
19	ice skates	21	69	bottle	19	119	exercise equipment	19
20	laptop	21	70	bowl	16	120	fan	20
21	leatherman	21	71	bowtie	18	121	fish hook	21
22	mailbox	20	72	bread	18	122	flag	19
23	protractor	20	73	broom	18	123	flashlight	20
24	mattress	20	74	bucket	17	124	frame	19
25	microphone	24	75	butterfly	17	125	frisbee	18
26	pasta	23	76	button	20	126	handheld game	19
27	printer	21	77	cake	21	127	garbage can	21
28	remote control	22	78	calculator	20	128	gift	19
29	robot	19	79	camcorder	20	129	glove	21
30	router	21	80	camera	21	130	goggles	21
31	scooter	20	81	candle holder	19	131	golfball	21
32	shower curtain	19	82	car	21	132	grill	20
33	shredder	20	83	cat	21	133	guitar	20
34	sink	20	84	ceiling fan	19	134	hair brush	20
35	skateboard	20	85	cellphone	20	135	hammer	21
36	sleeping bag	19	86	chair	22	136	handbag	18
37	typewriter	19	87	cheese	20	137	handgun	17
38	vacuum	20	88	cheese grater	21	138	hanger	19
39	wall lamp	19	89	cherub statue	21	139	hat	19
40	wooden box	19	90	chessboard	20	140	headband	19
41	abacus	21	91	christmas stocking	19	141	headphones	19
42	airplane	19	92	christmas ornament	18	142	helmet	19
43	apple	19	93	cigarette	20	143	hourglass	18
44	army guy	18	94	clock	20	144	jacket	20
45	axe	18	95	coat rack	17	145	jack-o-lantern	18
46	russian doll	23	96	coffee mug	18	146	juice	18
47	carriage	21	97	coin	18	147	kayak	18
48	backpack	21	98	collar	15	148	key	19
49	bagel	21	99	compass	18	149	keyboard	19
50	ball	19	100	computer key	18	150	keychain	20

Note: Categories 1-40 contained 16 exemplars each. All others contained 17 exemplars.

Table A2. *Stimulus categories 151-240, from Experiment 2. Displayed are the number of participants who contributed data to each set.*



Category Index	*Category Name*	*# of Participants*	*Category Index*	*Category Name*	*# of Participants*
151	knife	19	201	shoe	21
152	lamp	20	202	sippy-cup	19
153	lantern	19	203	snowglobe	19
154	lawnmower	20	204	socks	23
155	leaves	19	205	soda can	21
156	lei	18	206	sofa	22
157	license plate	19	207	speaker	22
158	lipstick	20	208	spoon	23
159	lock	19	209	stamp	21
160	magazine	20	210	stapler	22
161	makeup compact	20	211	stool	17
162	mask	18	212	suit	18
163	meat	19	213	suitcase	18
164	microscope	20	214	table	18
165	microwave	19	215	tape	16
166	motorcycle	18	216	telescope	19
167	mp3 player	16	217	tennis racquet	18
168	muffin	15	218	tent	18
169	mushroom	17	219	toilet seat	16
170	nail polish	16	220	tongs	19
171	necklace	18	221	toothpaste	18
172	necktie	19	222	toy horse	18
173	nunchak	17	223	toy rabbit	17
174	oriental fan	20	224	train	18
175	pants	17	225	tree	17
176	lounge chair	17	226	tricycle	24
177	pen	18	227	trophy	21
178	phone	17	228	trumpet	24
179	pipe	18	229	trunk	24
180	pitcher	17	230	turtle	24
181	pizza	18	231	television	23
182	poker card	18	232	umbrella	21
183	powerstrip	20	233	vase	23
184	radio	18	234	videogame controller	24
185	razor	19	235	watch	21
186	record player	20	236	watergun	23
187	ring	21	237	wig	23
188	3-ring binder	19	238	windchime	21
189	road sign	20	239	wine glass	22
190	rock	19	240	yarn	22
191	rollerskates	22			
192	rosary	19			
193	rug	19			
194	saddle	22			
195	salt and pepper shakers	21			
196	sandwich	18			
197	scale	20			
198	scissors	17			
199	scrunchie	19			
200	seashell	18			

Note: All categories contained 17 exemplars.

Figure A1. MDS coordinates for each of the 17 teddy bear exemplars, from Experiment 2.

Item	Picture	Dim1	Dim2	Dim3
1		-0.40	0.47	0.19
2		0.07	0.68	-0.16
3		0.56	-0.26	0.17
4		0.60	0.26	-0.16
5		-0.43	-0.42	-0.08
6		-0.27	0.43	-0.43
7		-0.46	-0.13	-0.46
8		-0.01	-0.48	0.49
9		0.17	0.55	0.33
10		0.00	0.10	0.65
11		-0.48	-0.11	0.46
12		0.13	-0.64	0.06
13		0.55	0.11	0.38
14		0.45	-0.32	-0.36
15		0.04	-0.42	-0.52
16		0.18	0.19	-0.60
17		-0.68	0.00	0.06

Figure A2. MDS coordinates for each of the 17 butterfly exemplars, from Experiment 2.

Item	Picture	Dim1	Dim2	Dim3
1		-0.18	-0.63	0.09
2		-0.36	-0.05	-0.52
3		0.10	-0.17	-0.63
4		-0.23	0.44	-0.41
5		-0.25	-0.19	0.56
6		-0.47	0.43	-0.23
7		-0.59	-0.21	0.10
8		-0.62	0.10	0.27
9		-0.26	0.27	0.51
10		0.63	-0.12	0.19
11		0.46	0.28	0.45
12		0.35	0.50	-0.20
13		0.52	-0.43	-0.18
14		0.60	0.12	-0.37
15		0.36	-0.37	0.44
16		0.15	0.61	0.21
17		-0.21	-0.59	-0.27

Figure A3. Sample *similar*, *moderate*, and *dissimilar* item pairings for the teddy bear exemplars, from Experiment 2. *Rank* indicates each pair's status, relative to other pairs, with 1 being the pair that are closest to one another, and 136 being the pair that are farthest apart. *Distance* is the three-dimensional Euclidean distance measured in arbitrary MDS units. *Item Pair* indicates the numerical identifier of each item in the pair.

Rank	Identity	Item Pair	Picture 1	Picture 2	Distance
1	Similar	3 – 13			0.42
2	Similar	14 – 15			0.45
3	Similar	11 – 17			0.46
67	Moderate	4 – 15			0.95
68	Moderate	3 – 16			0.97
69	Moderate	8 – 14			0.97
134	Dissimilar	2 – 8			1.33
135	Dissimilar	7 – 13			1.34
136	Dissimilar	2 – 12			1.34

Figure A4. Sample *similar, moderate,* and *dissimilar* item pairings for the butterfly exemplars, from Experiment 2. *Rank* indicates each pair's status, relative to other pairs, with 1 being the pair that are closest to one another, and 136 being the pair that are farthest apart. *Distance* is the three-dimensional Euclidean distance measured in arbitrary MDS units. *Item Pair* indicates the numerical identifier of each item in the pair.

Rank	Identity	Item Pair	Picture 1	Picture 2	Distance
1	Similar	4 – 6			0.30
2	Similar	7 – 8			0.35
3	Similar	1 – 17			0.36
67	Moderate	5 – 16			0.95
68	Moderate	5 – 10			0.96
69	Moderate	1 – 10			0.96
134	Dissimilar	8 – 13			1.34
135	Dissimilar	16 – 17			1.34
136	Dissimilar	8 – 14			1.38

Figure A5. Mean scan-path ratio (on accurate trials), from Experiment 3b, presented as a function of Set Size and Block. Error bars represent 1 standard error (SE) of the mean.

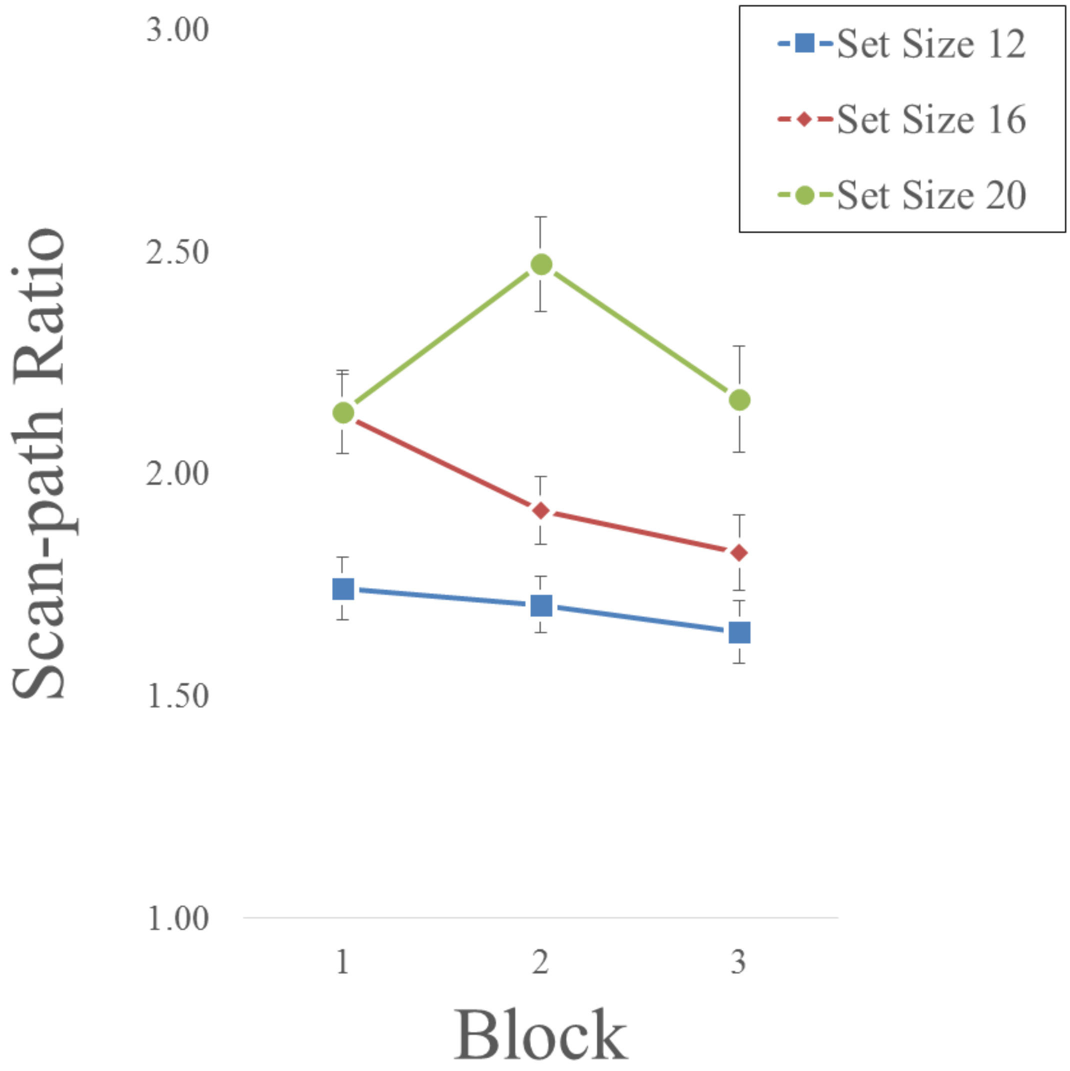


Figure A6. Mean scan-path ratio (on accurate trials), from Experiment 3b, presented as a function of Precision, Match Frequency (MF), and Block. Error bars represent 1 standard error (SE) of the mean.

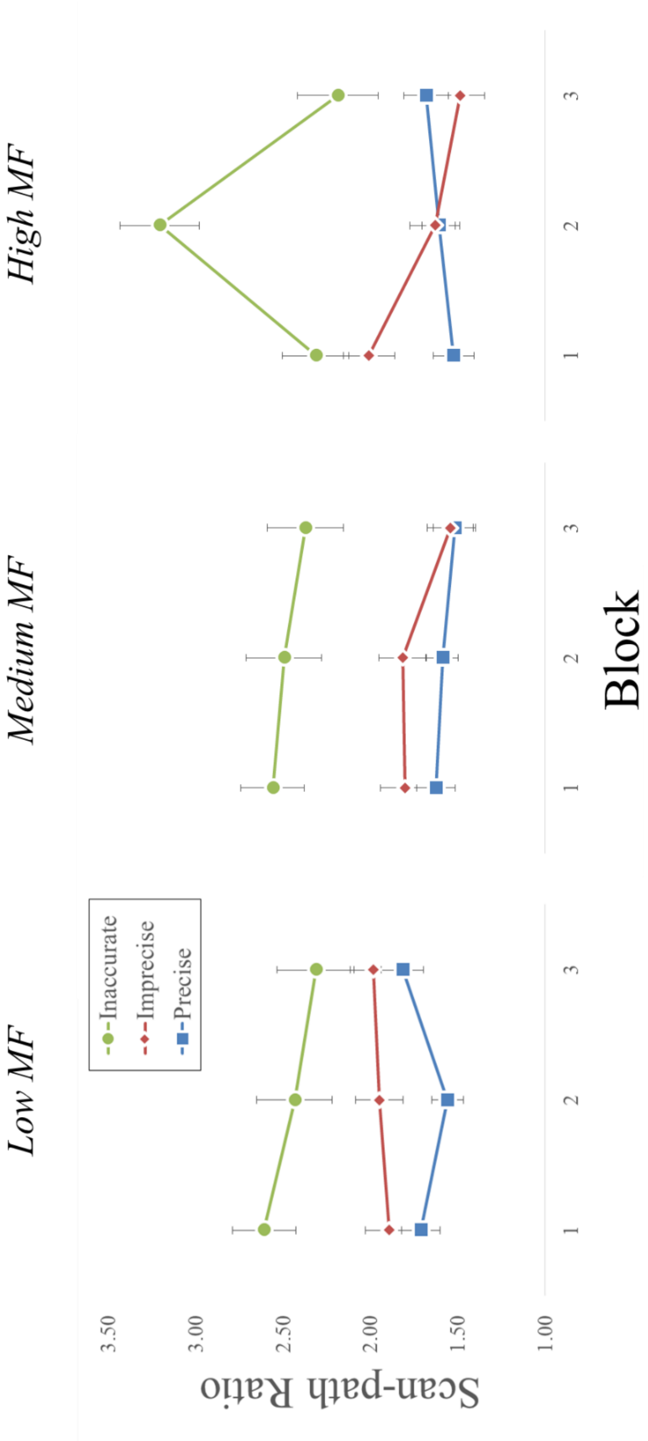


Figure A7. Mean decision time (on accurate trials) from Experiment 3f, presented as a function of Block and Set Size. Error bars represent 1 standard error (SE) of the mean.

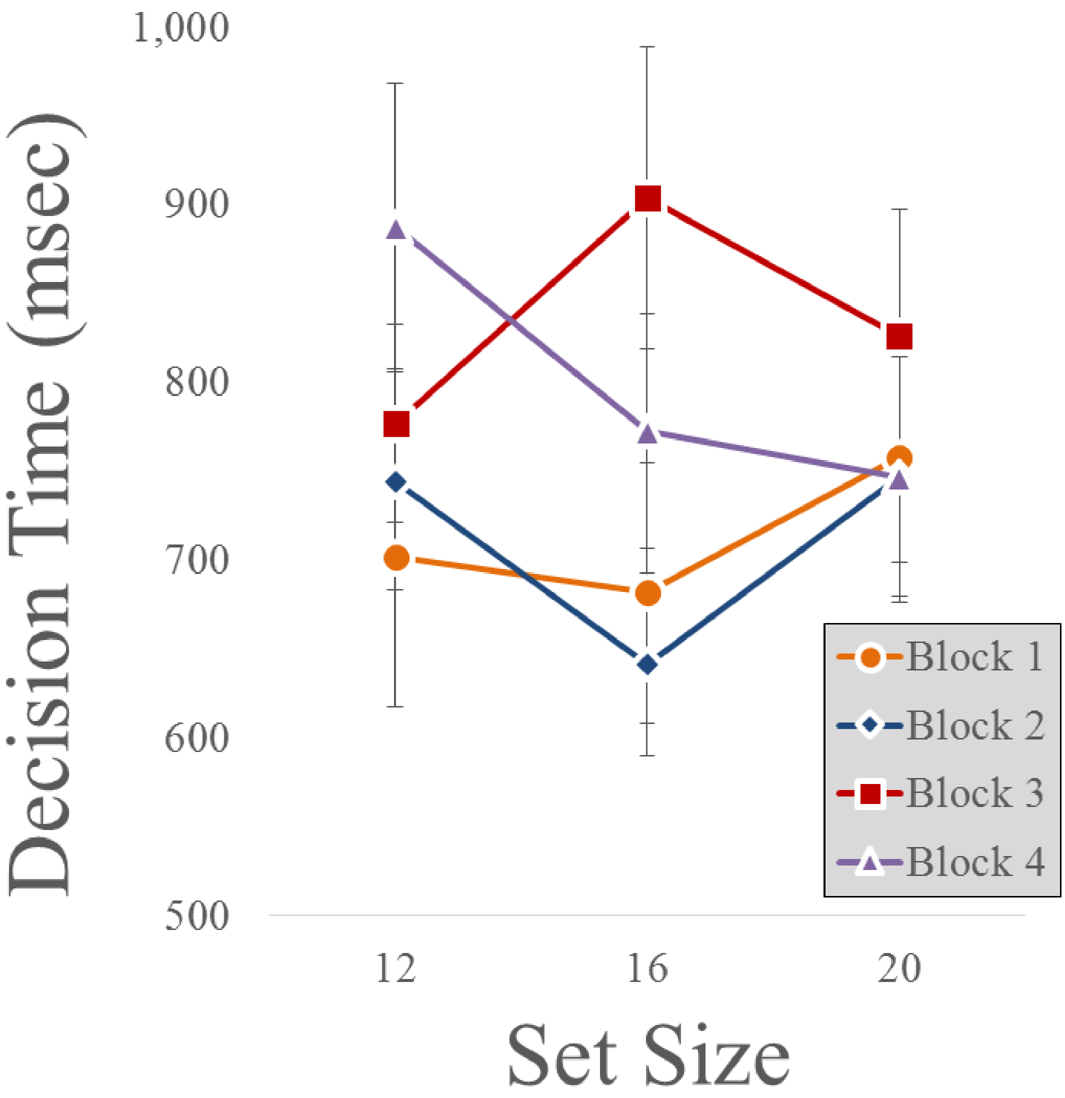


Figure A8. Mean search RT (on accurate trials) from Experiment 4b, presented as a function of Feature Width and Set Size. Error bars represent 1 standard error (SE) of the mean.

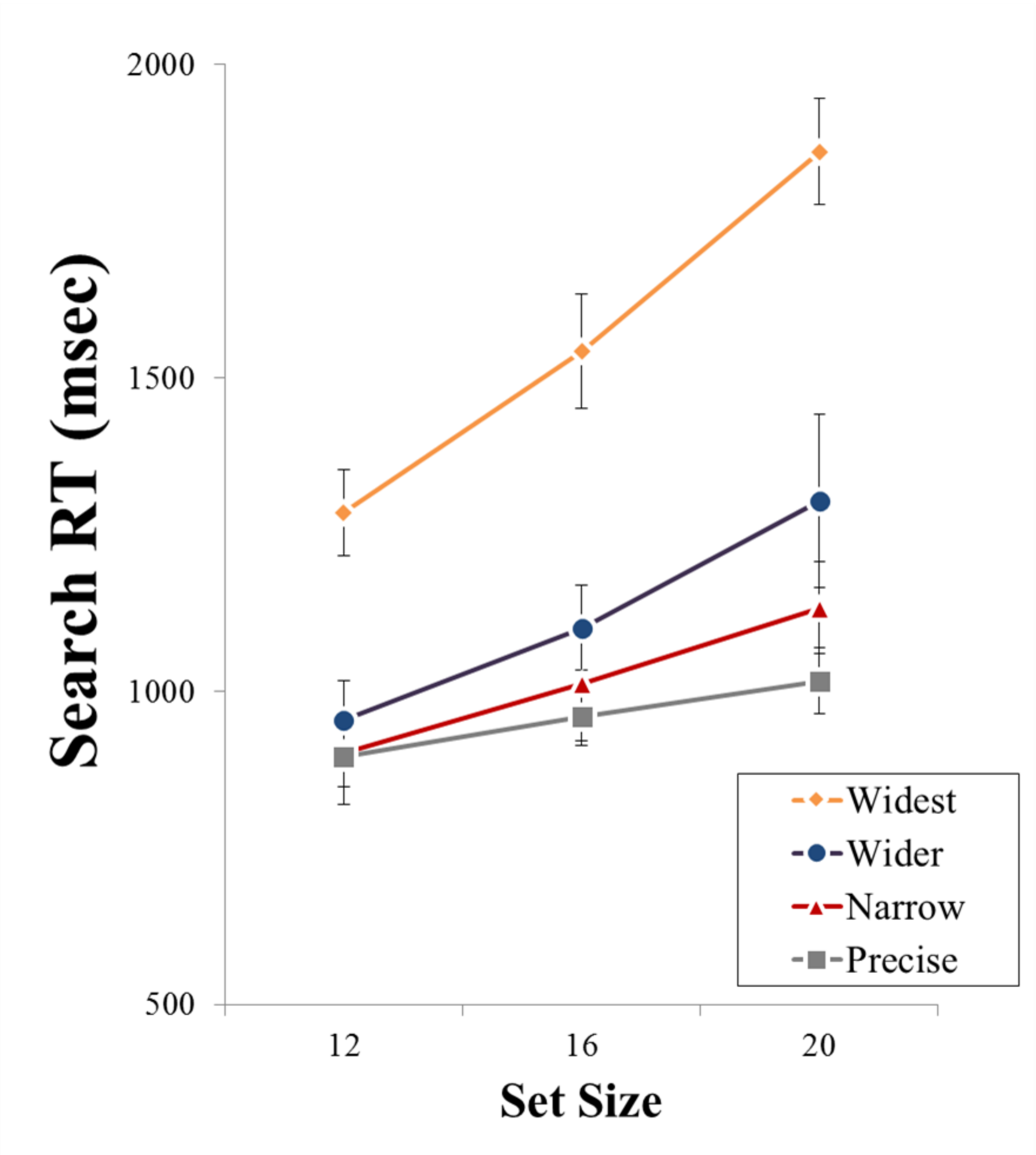


Figure A9. Mean decision time (on accurate trials) from Experiment 4b, presented as a function of Feature Width, Set Size, and Block. Error bars represent 1 standard error (SE) of the mean.

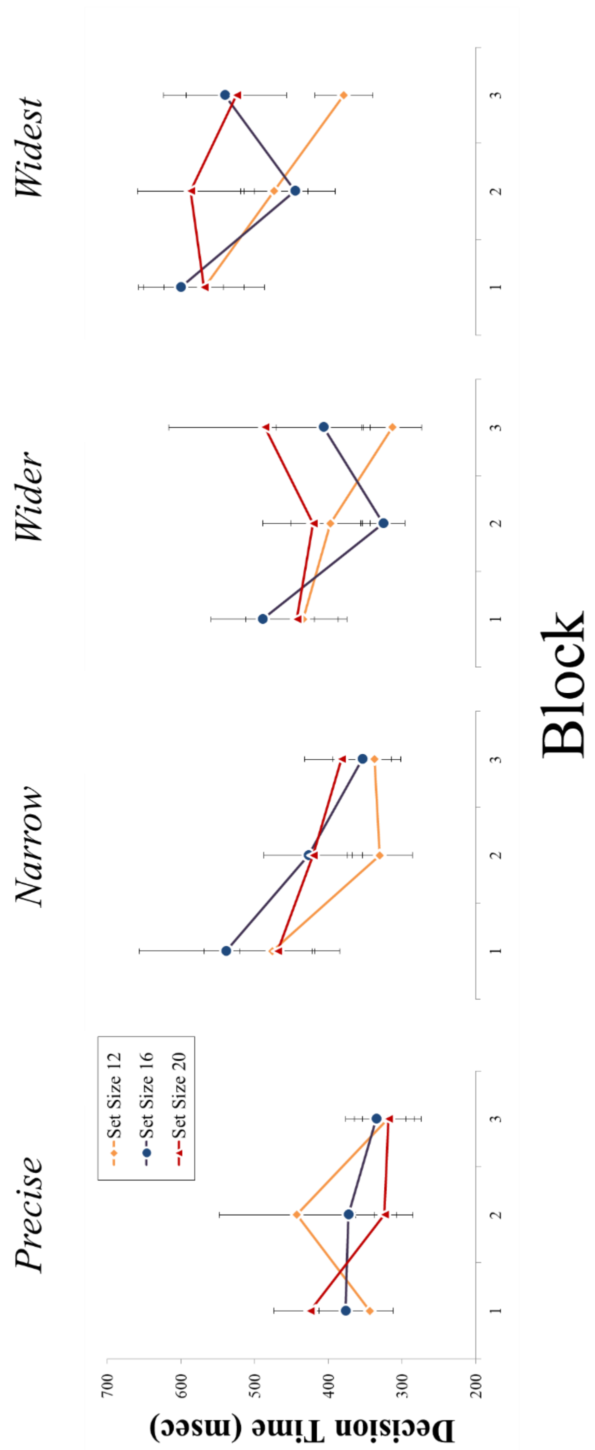


Figure A10. Mean search RT (on accurate trials) from Experiment 4c, presented as a function of Feature Width and Set Size. Error bars represent 1 standard error (SE) of the mean.

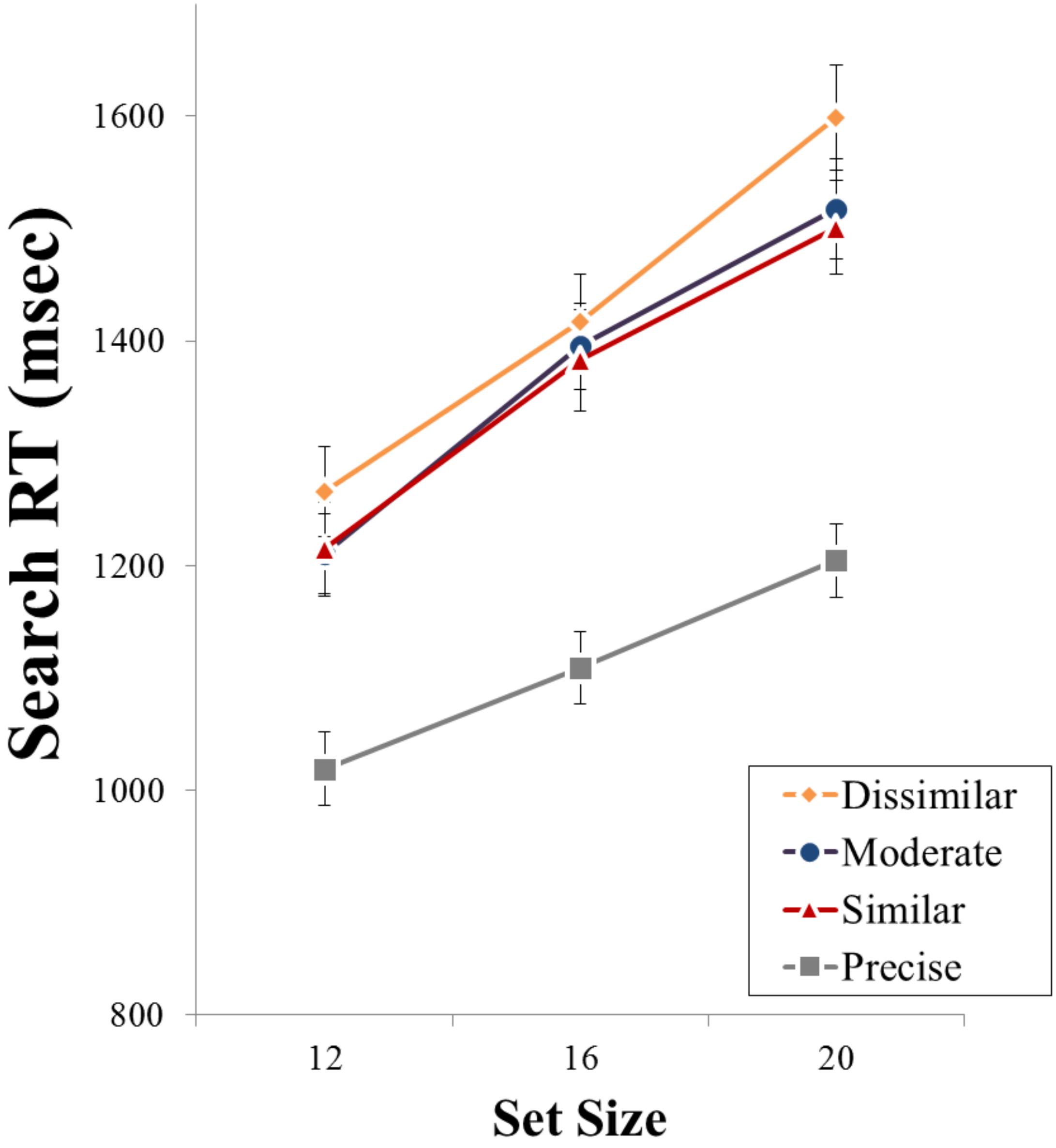


Figure A11. Mean search RT (on accurate trials) from Experiment 4d, presented as a function of Feature Width and Block. Error bars represent 1 standard error (SE) of the mean.

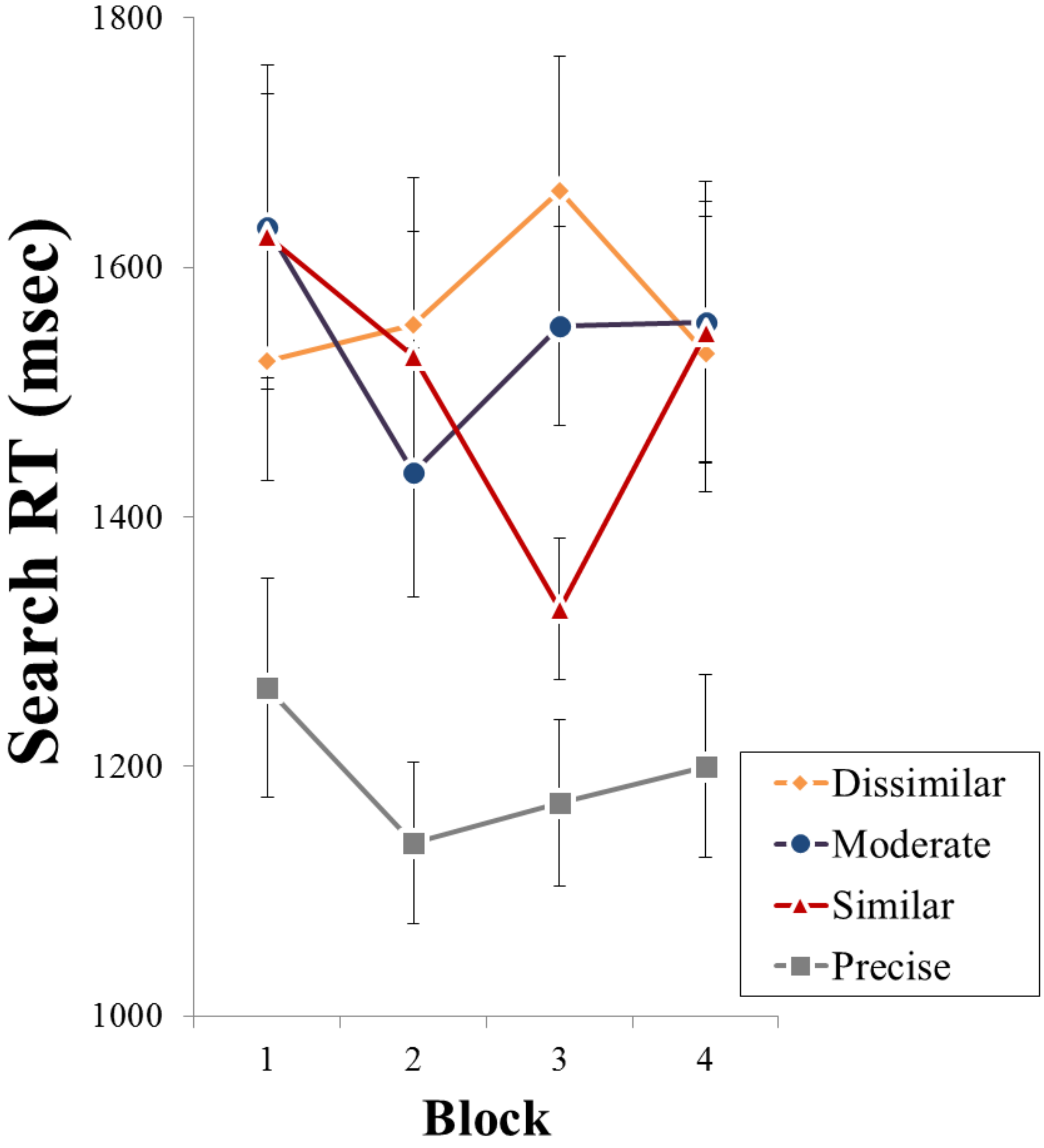


Figure A12. Target-absent visual search accuracy, from Experiment 5a, presented as a function of Set Size and Search Length. Error bars represent 1 standard error (SE) of the mean.

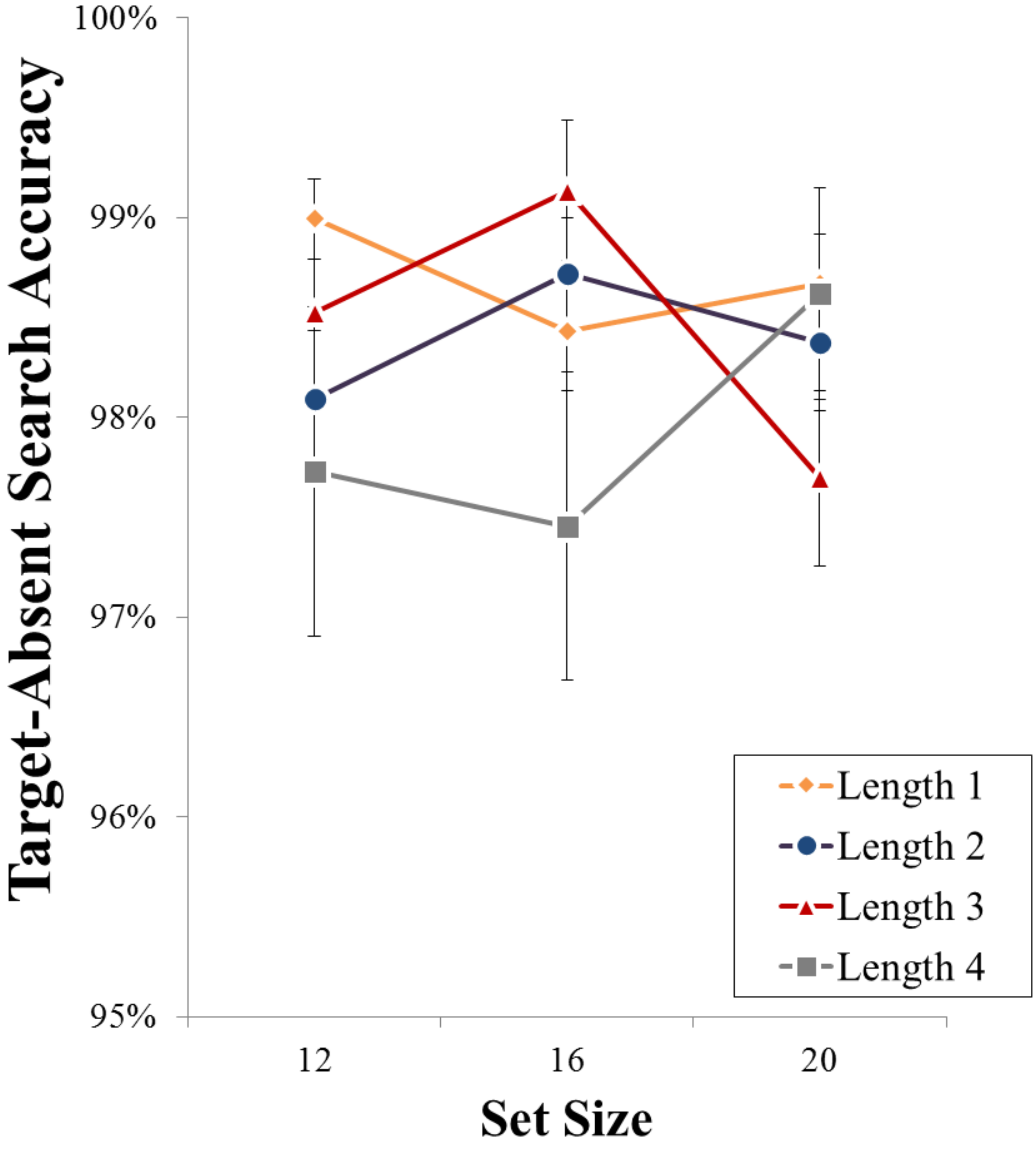


Figure A13. Target-present visual search accuracy, from Experiment 5b, presented as a function of Set Size and Search Length. Error bars represent 1 standard error (SE) of the mean.

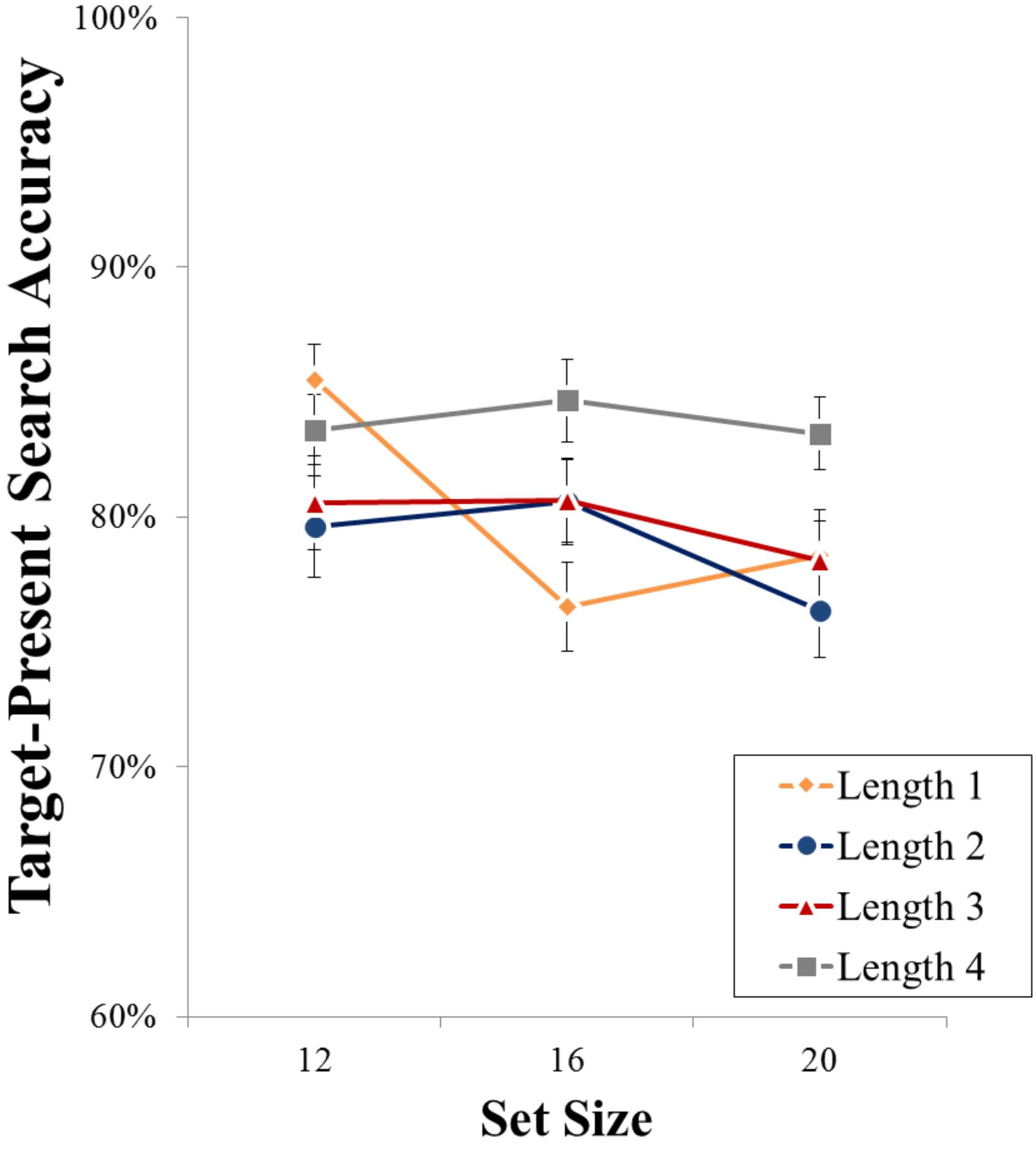
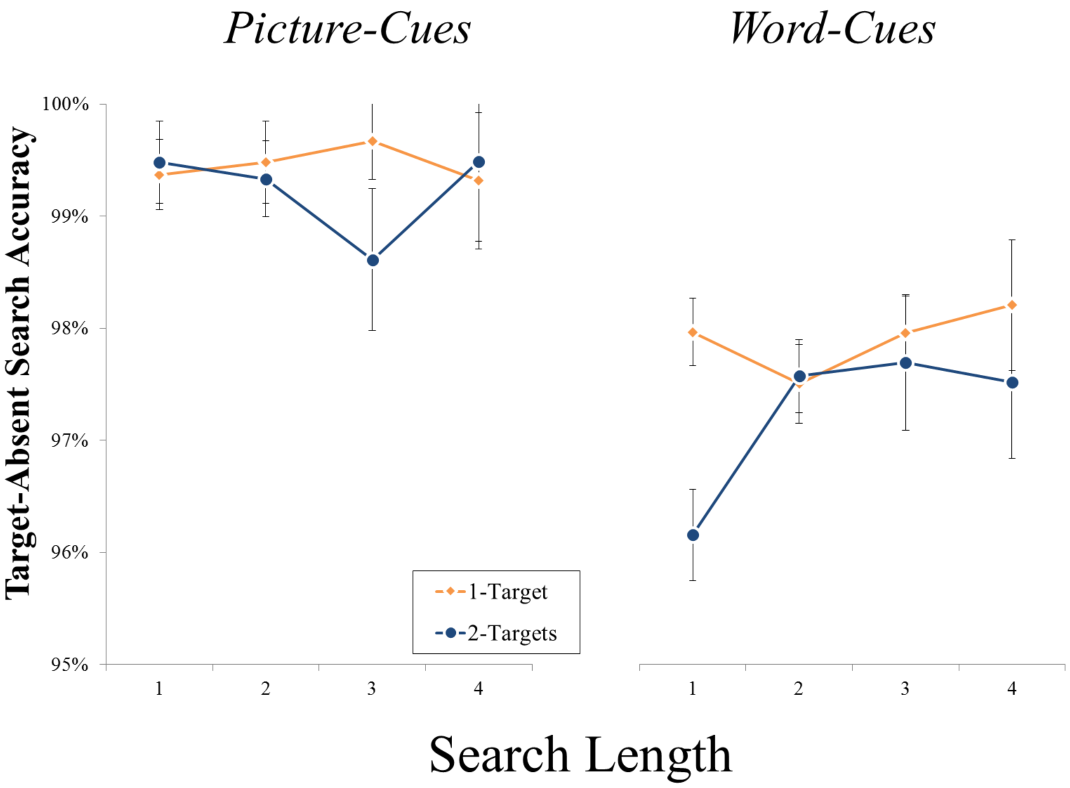


Figure A14. Target-absent visual search accuracy, from Experiment 5b, presented as a function of Cue Type, Target Load and Search Length. Error bars represent 1 standard error (SE) of the mean.



APPENDIX B

Figure B1. Arizona State University Institutional Review Board Human Subjects Research Approval.

ASU Knowledge Enterprise Development

Office of Research Integrity and Assurance

To:	Stephen Goldinger Psychology
From:	Mark Roosa, Chair Soc Beh IRB
Date:	09/05/2012
Committee Action:	**Exemption Granted**
IRB Action Date:	09/05/2012
IRB Protocol #:	1208008196
Study Title:	Target Templates in Visual Search

The above-referenced protocol is considered exempt after review by the Institutional Review Board pursuant to Federal regulations, 45 CFR Part 46.101(b)(2) .

This part of the federal regulations requires that the information be recorded by investigators in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects. It is necessary that the information obtained not be such that if disclosed outside the research, it could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the subjects' financial standing, employability, or reputation.

You should retain a copy of this letter for your records.