The role of working memory in visual search guidance for multiple targets

A Dissertation Presented

by

Hyejin Yang

to

The Graduate School

in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in

Experimental Psychology

Stony Brook University

August 2010

Stony Brook University
The Graduate School

Hyejin Yang

We, the dissertation committee for the above candidate for the
doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Gregory Zelinsky, Ph.D. – Dissertation Advisor
Associate Professor, Department of Psychology

Nancy Franklin, Ph.D. - Chairperson of Defense
Associate Professor, Department of Psychology

Hoi-Chung Leung, Ph.D.
Associate Professor, Department of Psychology

Lester Loschky, Ph.D.
Assistant Professor, Department of Psychology, Kansas State University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

The role of working memory in visual search guidance for multiple targets

by

Hyejin Yang

Doctor of Philosophy

in

Experimental Psychology

Stony Brook University

2010

It is an open question whether the search for multiple targets is less efficient than the search for a single target. Here we explored visual search guidance for multiple targets by tracking subjects' eye movements, with the broad goal being to better understand the close relationship between visual search and working memory. One series of experiments showed that the search for multiple targets is less efficient and less guided than the search for a single target. Using a retro-cue paradigm we were able to distinguish between two potential causes of this load effect; one related to capacity limitations on visual working memory during encoding, and the other due to a mismatch between the features of multiple targets in working memory and the single target in the search display. We found that the proportion of initial fixations on the target, a conservative measure of search guidance, was influenced by feature mismatch but not by a memory encoding limitation; guidance was affected by the targets indicated by the retro cue (feature mismatch) and not the number of targets shown at preview (memory encoding). We therefore conclude that multiple target search is less guided than single target search due to too many features in visual working memory (from multiple targets) weakening the guidance to the actual target appearing in the search display. A second series of experiments explored how multiple targets are represented, in terms of feature shared by two targets (common features) or features unique to two targets (distinctive features), and how this representation changes with the conceptual relationship between the search targets. We found that two dissimilar targets are represented by distinctive features, whereas two similar targets are represented by common features, although in both cases search guidance improves with the use of distinctive features.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank, with all my heart, Prof. Gregory J. Zelinsky, my advisor, for his patience, encouragement, and guidance over the years. He showed me what a researcher should be with his passion, diligence, and, most of all, intelligence. Thanks to Dr. Christopher Dickinson for showing me how sincere efforts are paid off. Dr. Xin Chen, thanks for helping me with programming and for saving me from bugs. Thanks to Dr. Mark Neider for showing me how productive a parent can be. Many many thanks to Joe Schmidt for all the chats and discussion, for helping me with EB, and for dealing with all the small but time consuming things in the lab. Thanks to Robert Alexander for bringing sunshine and noises in the lab. Thanks to Westri Stalder for guarding our lab during nights after Xin. Not to mention that all of them inspired me by their humors, brilliance, and endeavor, which I really appreciate.

Thanks to Dr. Hwamee Oh for putting so much time and effort for my first fMRI project. Thanks to Vera for her friendship and supports.

I really deeply appreciate all the love and supports from my parents and parents in law. Especially my mom and dad, without them, this wouldn’t be possible. My dearests, Irene and Amber, you made me a mom and strong. You will never know what it meant to me. My husband, my companion, Jongsik, I am so blessed to be with you. I love you all.

1. Overview

We search for a book from a book shelf. We search for quarters to pay meters. Like these examples, visual search is a task where you try to find a given target item. It may be performed efficiently regardless of the number of items in a search display when the target can be distinguished from distractors easily. However, it may be performed inefficiently when the target looks similar to distractors. In this case, attention needs to be allocated to the location of each item in a serial manner (Treisman & Gelade, 1980). Allocation of attention is not random. It has been shown that attention is guided to the items that share visual features with a target (Wolfe, 1994). For example, if you search for a red square, a red circle might attract your attention because they have the color red in common. This topic of guidance in visual search was studied widely, and it is this potential for guidance that creates a relationship between visual search and working memory (WM).

During visual search, target representations are maintained online so as to efficiently guide search to the target and to recognize the target once it is found. This is especially true when there is a delay between the target preview and the search display. Both of these functions require the involvement of memory in search. Target representations are likely to be held and maintained in WM during search, as WM is responsible for storing goal-relevant information. Not surprisingly, most visual search theories assume the involvement of WM during visual search. These theories suggest that target representation in WM guides the allocation of attention to items that have target features (Wolfe, 1994; Desimone & Duncan, 1995).

In line with those theories, many visual search studies have shown that attention is biased to search items having target features. Importantly, this bias has been quantified using eye movement measures of search guidance. For example, distractors sharing target features are fixated more than distractors that do not (Williams, 1967; Luria & Strauss, 1975; Scialfa & Joffe, 1998; Williams & Reingold, 2001). Distractors having target features are also preferentially fixated by initial saccades (Findlay, 1997). Other studies have shown that search guidance improves with the level of target specification (Chen & Zelinsky, 2006; Malcolm & Henderson, 2009; Schmidt & Zelinsky, 2009); when subjects have more information about what a target will look like, they are more efficient in shifting their gaze to the target. Collectively, these findings suggest a role for WM in search; target features are held in WM and compared to the search display, with search being preferentially directed to those items best matching these target features.

Several studies have directly focused on the relationship between WM and visual search. Woodman, Luck, and Schall (2007) examined if target templates are maintained in WM during visual search, as assumed by visual search theories. They compared search efficiency when subjects were engaged in dual visual search and WM tasks to search efficiency when subjects were only engaged in a visual search task. Targets in the visual search tasks were either held constant throughout the whole experiment or varied on every trial. They found that search efficiency in the dual task condition was degraded compared to search efficiency in the single task condition only when the target changed on every trial. They concluded that targets are held in WM, and need to be encoded anew when the target identities change from trial to trial.

Other studies have focused on whether visual search and WM share the same mechanism (Oh & Kim, 2004; Woodman & Luck, 2004; Woodman, Vogel, & Luck, 2001). Participants in these studies performed either a dual WM and visual search task, or just a visual search task. Two different types of WM tasks were used; a spatial WM task and an object WM task. For the spatial WM task, several dots were presented and subjects had to remember the locations of the dots while performing visual search. For the object WM task, subjects were presented with colored squares or boxes with gaps and had to remember the colors of the squares or shapes of the boxes while performing visual search. They found that visual search efficiency decreased when participants performed visual search and spatial WM tasks together compared to when participants performed only visual search. However, search efficiency did not decrease when participants performed the dual visual search and object WM task. These results suggest that visual search and visual working memory share a spatial attentional mechanism, but perhaps use different features.

A related study asked whether visual search uses the same executive functions used by WM (Han & Kim, 2004). Again, subjects performed either a dual WM and visual search task, or visual search alone. The WM tasks either required an executive function (backward counting or alphabet reordering) or simple storage (remembering digits or letters). Visual search efficiency in the dual task condition was degraded compared to the search alone condition only when the WM task required an executive function. This finding suggests that executive functions are also shared by WM and visual search tasks.

Other studies provided evidence that the representation in WM affects the allocation of attention during visual search (e.g., Soto, Heinke, Humphreys, & Blanco, 2005; Woodman & Luck, 2007). The task in these studies required holding an object in WM while performing visual search. A target or a distractor in the search display could match the object in WM. When targets matched the WM representation, visual search efficiency increased (Soto, Heinke, Humphreys, & Blanco; Woodman & Luck). When distractors matched the WM representation, different studies reported different findings. Some studies found that search efficiency decreased when distractors were inside geometric shapes that matched the WM representation (Soto, Heinke, Humphreys, & Blanco). Other studies found that search efficiency remained the same regardless of the presence of distractors matching the WM representation (Downing & Dodds, 2004; Houtkamp & Roelfsema, 2006). Still other studies found that search efficiency actually increased when only the distractors, not the targets, always matched the WM representation (Woodman & Luck, 2007).

Most recently, Hollingworth and Luck (2009) directly investigated how WM contents affected the control of eye movements during visual search. In their study, subjects were holding a color in WM for a later discrimination task. As subjects made a saccade to a target, the search display was rotated so that the saccade landed on a location between a target and another object. When this object had the same color as the color for the later discrimination task, the proportion of the next saccades to the target decreased compared to when the object had a different color. The authors concluded from this finding that eye movements during visual search are controlled by visual WM. In sum, there is reasonable consensus that WM representations affect search guidance by orienting attention to items matching the contents of WM, with the strength of this effect depending on the nature of the search item.

This involvement of WM during visual search gets more obvious when we search for several things at the same time, which is the focus of this study. When we search simultaneously for multiple targets, for example a note and a pen, we need to hold multiple target representations in WM. Holding multiple things for a later test is a typical way to manipulate WM load in the WM literature. In this study, we addressed search guidance for multiple targets using this standard WM manipulation.

Chapter 2 examines the interaction between search guidance and WM load: is search guidance different between single target search and multiple target search? Experiment 1 shows that subjects holding multiple targets in WM were able to guide their search less efficiently compared to when they were holding only a single target in WM. Following this initial demonstration, Experiment 2 asked why search guidance decreased with multiple targets. Specifically, was this decreased guidance for multiple targets due to subjects picking only one of the targets to guide their search (a working memory limitation), thereby producing no guidance for an incorrectly chosen target, or was this reduced guidance due to subjects holding both targets in WM, but having an imperfect feature match between the single object appearing in the search display and the features from both target objects in WM. We found that subjects were holding both target objects in WM, and that decreased guidance was due to too many target features guiding search; given that only one target appeared in the search display, the mismatching target in WM was essentially guiding search *away* from the actual target.

Chapter 3 asks *how* people use WM representations to guide their search. Specifically, we wanted to see if multiple targets were represented by features common to both targets or by features distinctive to each of the targets. In experiment 1, we tested this question using two targets that were from different categories and found that people used distinctive features to guide multiple target search. In experiment 2, we asked the same question using two targets from the same category and found that people used common features for two targets when the targets were very similar to each other.

Chapter 4 summarizes these findings, discusses their implications, and points to potential directions for future work.

2. Too many features spoil the guidance: Multiple target search is inefficient due to feature mismatch

2.1. Introduction

We search for a pen and a piece of paper. We search for a fork and a knife. As these examples illustrate, visual search in our everyday lives very often involves searching for more than one thing at the same time. This topic of multiple target search has been addressed in the visual search literature for several decades. The main question is if the search for multiple targets is less efficient than search for a single target, and there have been mixed results. In some studies, search for multiple targets was as efficient as single target search (e. g., D'Zmura, 1991; Quinlan & Humphreys, 1987). Other studies have found that the efficiency of multiple target search could be improved after practice (Neisser, Novick, & Lazar, 1963; Schneider & Shiffrin, 1977; but see Menneer, Cave, & Donnelly, 2009 for a different finding), or after a decrease of the number of possible targets from all 26 English letters to five letters (Metlay, Sokoloff, & Kaplan, 1970). In still other studies, however, multiple target search was found to be less efficient than the search for a single target; subjects needed more time to detect multiple targets than to detect a single target (Kaplan & Carvellas, 1965; Menneer, Barrett, Phillips, Donnelly, & Cave, 2007; Moore & Osman, 1993).

Interestingly, one study showed that search for multiple targets could be efficient or inefficient depending on the feature dimensions that defined the targets (Treisman, 1988). Subjects did not need more time to search for three different colored bars than to search for a single colored bar. However, they did need more time to search for a blue bar, a horizontal bar, and a large bar, in the presence of small green vertical distractor bars, than to search for a single bar. A recent study also showed that the cost for multiple target search depends on the featural similarity between the targets (Menneer, Cave, & Donnelly, 2009). When subjects searched for x-ray images of a knife and a gun, both of which looked blue, the accuracy for multiple target search was similar to that for single target knife search, which showed lower accuracy than a single target gun search. When subjects searched for x-ray images of a metal threat weapon (blue) and an improvised explosive device (multiple colors), the accuracy for multiple target search was lower than that for single target search. Based on these findings, they suggested that multiple target search could be efficient when the targets share common features between them. So it seems that the search for multiple things at the same time requires more time than search for one thing unless you are searching for a red pen and a blue pen or for a red pen and a red apple at the same time.[1]

The primary goal of this experiment is to determine if there is a cost associated with the search for multiple targets (relative to single target search performance) when these targets are visually complex real-world objects. Most of the previous work on this question used simple stimuli, such as letters or geometric shapes. The search for multiple letters might be as efficient as the search for a single letter (e.g., Neisser et al., 1963; Schneider & Shiffrin, 1977) because of perceptual grouping caused by the small number of features shared by letters. It is therefore unclear whether the previous findings, which have all been obtained using simple stimuli, will generalize to multiple target search in more naturalistic contexts. Also, given that familiarization with stimuli can increase

[1] It is unclear whether a common feature search, searching for just a pen or a red thing, should not be considered a multiple target search, as this single common feature (or set of features) might constitute a new single search target. We will return to this question in Chapter 3.

search efficiency (e.g., Wang, Cavanagh, & Green, 1994), the fact that previous studies used stimuli repeatedly might be another reason why multiple target search was found to be as efficient as single target search. To address these concerns and to mimic multiple target search in our everyday lives more closely, we investigated whether there was a cost in searching for multiple real-world objects compared to searching for a single real-world object, while avoiding potential effects of stimulus familiarity on search efficiency by never repeating stimuli.

If multiple target search is less efficient than single target search, *why* is it less efficient? Answering this question is the second major goal of this experiment. We will consider two factors that might plausibly affect the efficiency of multiple target search. First, a limit on WM capacity might in turn limit the number of targets that can be held in memory for the purpose of visual search. The visual search task requires finding a designated target item among distractors, and the efficient performance of this task requires holding the target representation in WM so that it can be actively compared to patterns in the search display. It is this process of ongoing comparison between a WM target representation and a search display that is believed to result in the efficient guidance of search to a target (Desimone & Duncan, 1995; Duncan & Humphreys, 1989; Wolfe, 1994). Although there have been several studies exploring the relationship between WM and visual search, the focus of these studies has been on determining how the contents of WM affects search, such as the guidance of attention to search items matching the WM representation (e.g., Soto, Heinke, Humphreys, & Blanco, 2005; Woodman & Luck, 2007) and the maintenance of targets in WM only when their identities change on every trial (Woodman, Luck, &Schall, 2007). Despite the close relationship between WM and visual search, no study of multiple target search has spotlighted the role of WM.

Multiple target search has obvious implications for WM, as there are multiple target templates that must be held in WM in order to guide search. The distinction between a single representation and multiple representations is referred to as a load manipulation in the WM community, and there are ongoing debates on the capacity of visual WM to accommodate such loads. Some studies concluded that visual WM could hold up to four objects regardless of their complexity by showing that change detection performance for both single-feature and multiple-feature objects degraded dramatically when a test array had more than four objects (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). Others have argued that visual WM capacity depends on stimulus complexity; we can hold fewer visually complex objects compared to simple objects (Alvarez & Cavanagh, 2004; Eng, Chen, & Jiang, 2005). These studies used stimuli that varied in visual complexity (colored squares, letters, Chinese characters, and shaded cubes) and found that change detection performance decreased as object complexity increased. This suggestion was challenged by a later study that found evidence for a fixed WM capacity regardless of object complexity, arguing that the previous evidence for a reduced capacity with complex objects resulted from comparison error that was due to high sample-test array similarity rather than a capacity limit (Awh, Barton, & Vogel, 2007).

Rather than assuming that visual WM capacity is determined by the number of objects *or* the number of features, other studies have suggested that visual WM is constrained both by the number of objects *and* number of features so its capacity could be both fixed and variable (Wheeler & Treisman, 2002; Xu & Chun, 2006). Wheeler and Treisman (2002) suggested that visual WM has separate stores for features and objects with bounded features. They also suggested that there is a limit on the numbers of features that can be stored in each feature dimension, but that there is no limit on the numbers of features that can be stored across different feature dimensions. A recent fMRI study would seem to support this position by showing that different brain areas hold different numbers of objects in visual WM (Xu & Chun, 2006). Inferior intra-parietal sulcus (IPS) seems to represent up to four objects regardless of complexity, while superior IPS and lateral occipital complex (LOC) seem to represent variable numbers of objects depending on complexity. Representations in inferior IPS do not have detailed feature information about the objects, while representations in superior IPS and LOC have detailed feature information, which is why these areas show differences regarding visual WM capacity (Xu, 2009).

Related to the above studies, other recent efforts have tried to characterize WM in terms of the resolution of its representations rather than capacity limits on the number of objects or features. For example, Zhang & Luck (2008) showed that six items could be represented as accurately as three items, but that the probability for six items to be represented in visual WM was about half of that for three items. This suggests that when the number of items to be encoded exceeds visual WM capacity limits, we selectively represent a subset of items with high precision. Bays and Husain (2008) posited a different view. While adopt the concept of 'slots' when explaining visual WM capacity limits (e.g., Luck & Vogel, 1997; Alvarez & Cavanagh, 2004), they explained visual WM capacity in terms of more graded limited resources. They found that the precision of representations gradually decreased as the number of items increased from one to six. Given this finding, they concluded that multiple objects compete for limited resources, but that there is no a priori limit on the number of objects that can be represented in visual WM; as the number of objects increase, their resolution simply decreases.

The suggestion that visual WM is capacity limited, either in terms of the number of objects in WM or in terms of the precision of the representations, provides one explanation for why multiple target search might be less efficient than single target search. For example, suppose that a subject is shown a vase with flowers and a lamp as target previews in a multiple target search task. Because either of these objects might appear in the search display, the subject will need to encode detailed feature information from both, thereby involving both the superior IPS and LOC regions. And because the vase with flowers and the lamp consist of a potentially large number of features, with perhaps many features (e.g., red, green, blue, etc.) existing on each feature dimension (e.g., color), one might expect the capacity limit on visual WM to limit the encoding these objects. If the number of features of multiple targets normally exceeds visual WM capacity, then not all of these target features will be encoded, resulting in poorer search guidance and reduced efficiency for multiple target search. Similarly, if graded resources are the source of limitations, then the representation of the lamp and vase with flowers will both be less precise compared to a representation of each individually. Regardless of whether the limitation is coming from slots or graded resources, the ultimate result is that information about complex objects is lost at encoding when the task requires that multiple objects be represented in visual WM, and that these impoverished representations are responsible for the degraded guidance observed in multiple target visual search. We will refer to this as the *capacity-limitation hypothesis*.

An alternative reason for why multiple target search might be less efficient than single target search has to do with of the comparison operation underlying search guidance. Visual search theories have suggested that a target definition or template is stored in WM and that search is guided by this WM representation of the target (Desimone & Duncan, 1995; Duncan & Humphreys, 1989; Wolfe, 1994; Zelinsky, 2008). For example, the guided search model proposed that targets are analyzed in terms of basic features such as color and orientation, with the specific features on each of these dimensions (e.g., red or vertical) being stored in visual WM. These WM representations are then assumed to guide attention to the locations of the objects that have the target features in the search scene; the more features the search target has in common with the WM target representation, the greater this item's priority is in the selection of objects to inspect with focal attention (Wolfe, 1994; Zelinsky, 2008). This relationship between search guidance and the target features held in visual WM has implications for the current discussion of multiple target search. To the extent that the features from multiple targets are represented in visual WM, and to the extent that the features in visual WM are used to guide search (Wolfe, 2004; Zelinsky, 2008), then some discrepancy or mismatch is inevitable between the features from the multiple targets held in WM and the features of the single target appearing in the search display. We will refer to this feature mismatch, and the consequent reduction in search guidance that it will inevitably bring, as the *feature-mismatch hypothesis*. For example, if the features of a vase with flowers and the features of a lamp are all encoded into visual WM as a result of having these objects designated as targets, but it is only the lamp that appears in the search display, then poorer guidance is expected (relative to a single target search for only the lamp) due to the features of the vase and flowers not matching the features of the lamp in the search display. Decreased search guidance for multiple target search is therefore a necessary consequence of having two (or more) targets in WM, but only one of these targets in the search display.

It is also important to note that this effect of mismatching features is independent of any effect of visual WM capacity limits on guidance in a multiple target search task. If you are searching for a red vertical bar and a green horizontal bar, which are well within visual WM capacity limits, and if a red vertical bar appears as a target in search display, there is still a mismatch between target features in WM and target features in the search display, and this mismatch would be expected to worsen guidance relative to a single target search for a red vertical bar. The efficiency of multiple target search can therefore be affected by either capacity limitations on visual WM, mismatches between target features in visual WM and the search display, or both.

The remainder of this chapter describes two experiments that explore more fully the relationship between visual WM load and search guidance. In Experiment 1 we set out to answer two questions. First we ask whether the search for multiple targets is less efficient and less guided than the search for a single target using real-world objects. Second, to the extent that multiple target search is less guided than single target search, we seek to determine whether this load effect is due to limitations on WM capacity or feature mismatch. To answer this second question we adopted a retro cue paradigm. These were cues presented after the target previews but before the search display. By combining these cues with a manipulation of the number of target previews presented to subjects, we can independently assess the effects of WM limits and feature mismatch on search efficiency and guidance. Guidance costs arising from WM capacity limitations would be revealed by varying the number of target previews appearing before the retro cue, whereas costs attributable to feature mismatch would be revealed by using the cue to manipulate the number of the previewed targets that subjects could expect to appear in the search display. In Experiment 2, we adopt a dual-target paradigm (having multiple targets appear in the actual search display) to independently assess whether load effects on search guidance are due to mismatching features, or subjects using only a subset of the previewed targets to guide their search.

2.2. Experiment 1

2.2.1. Introduction

Discussion of WM load effects in the context of search are complicated by the fact that there are at least two components, capacity limits on WM and feature mismatch, that might affect search performance. Previous studies directly comparing single target to multiple target search failed to distinguish between these components, thereby leaving ambiguous the cause of any observed load effect. In this experiment we attempt to resolve this ambiguity.

We will introduce a retro cue to tease apart these WM and search guidance components to see how each affects efficiency and guidance for multiple search targets. A retro cue will be used to manipulate the number of target preview representations that are held in WM, thereby directly manipulating WM load and indirectly manipulating the number of targets subjects will search for at the same time (i.e., the number of target templates that might be used to guide search). The retro cue is presented after the target preview(s) to inform subjects as to which of the previewed targets (one of them, some of them, or all of them) they should search for. For example, after seeing four target objects during preview, the retro cue inserted before the onset of the search display will indicate to subjects whether they could search for only one of the previewed objects, two of them, or all of them. Using this retro cue paradigm, we can untangle the contributions of WM capacity limits and feature mismatch on search guidance, enabling us to see how each of these components might contribute to any observed costs associated with searching for multiple real-world objects.

According to the capacity-limitation hypothesis, search guidance should degrade with the number of target previews (WM load) regardless of how many of these previews are designated by the retro cue to be relevant to search. This is because the subject would be forced to encode all of the multiple target previews into visual WM, as at this point in the trial the cue would not yet have indicated which of these targets might appear in the actual search display. If a limitation on WM capacity prevents the complete encoding of all of these targets, these incomplete or partially-encoded target representations would be expected to lessen search guidance and efficiency and produce a load effect.[2] Finding a main effect of the number of target previews would therefore provide evidence for encoding limitation as being the cause of load effects, and support for the capacity-limitation hypothesis.[3]

---

[2] Note that we don't distinguish here between cases in which all of the target previews are partially encoded, or some subsets of the target previews are fully encoded. For the purpose of this study we consider these both to be expressions of an encoding limitation on visual WM.

[3] Note that this definition of a load effect limits the role of WM on search specifically to a capacity limitation expressed at encoding. We do not consider any forgetting from WM that might occur between preview offset and search display onset, or any effect of a maintenance process in offsetting this forgetting

The feature-mismatch hypothesis makes a very different prediction. If search guidance is degraded purely as a result of feature mismatch, then the number of target previews (one, two, etc.) designated at the start of each trial should not affect search guidance, as the assumption is that visual WM capacity is sufficient to encode all of the target previews. Rather, this hypothesis predicts an effect of the retro cue, which determines the number of targets that will actually be compared to the search display and used to guide search. When the retro cue is non-specific, then multiple targets will be compared to the search display, resulting in a relatively high degree of feature mismatch and poor search guidance. As the retro cue becomes more specific, fewer extraneous targets will be compared to the search display, resulting in less feature mismatch and better search guidance. Finding a main effect of retro cue would therefore support the feature-mismatch explanation for a load effect. Of course it might also be the case that WM constraints and feature mismatch might both affect search, or potentially interact. These patterns, too, would be revealed using our retro cue paradigm.

2.2.2. Method

Participants

Forty-five students from Stony Brook University participated in the experiment for course credit. All had normal or corrected to normal visual acuity and were naive with regard to the experimental hypotheses.

Stimuli and Apparatus

Targets were images of teddy bears, butterflies, cars, and fish. Teddy bear images were obtained from Cockrill (2001) and modified for this experiment. All

---

(e.g., Woodman & Vogel, 2005), when evaluating the relationship between WM load and search guidance in this study.

butterfly images and most of the car and fish images were selected from the Hemera Photo Objects Collection (Gatineau, Quebec, Canada). Other car and fish images were obtained from random web sites and modified for this experiment. The distractors were images of real-world objects randomly selected from the Hemera Photo Objects Collection. None of the distractors belonged to the four target categories. We resized all of the objects such that the smallest bounding box enclosing each consisted of exactly 5000 pixels. Although this manipulation allowed the heights and widths of the objects to vary, the widest object was about 4° and the tallest object was about 3.6°. All objects were presented in color, and no object, target or distractor, was presented more than once. This was done in order to have a relatively pure representation in WM without the potential for contamination from long-term memory (LTM).

The experiment was designed and controlled using the Experiment Builder software package (SR Research Ltd, version 1.5.1.). Eye movements were recorded using an EyeLink II eye tracking system (SR Research Ltd). This video-based eye tracker sampled eye position at 500 Hz and had a spatial resolution of about 0.2°. We used a chinrest to prevent head movement and to keep a fixed viewing distance between the monitor and subjects. Manual responses were collected using a Game Pad controller interfaced through the computer's USB port. Stimuli were displayed using a ViewSonic 19" flat-screen CRT monitor at a refresh rate of 100 Hz. The viewing angle of the display was 26° horizontally and 20° vertically.

Procedure

The experiment started with an initial calibration procedure, needed to map screen coordinates to eye position. This was followed by a short block of practice trials,

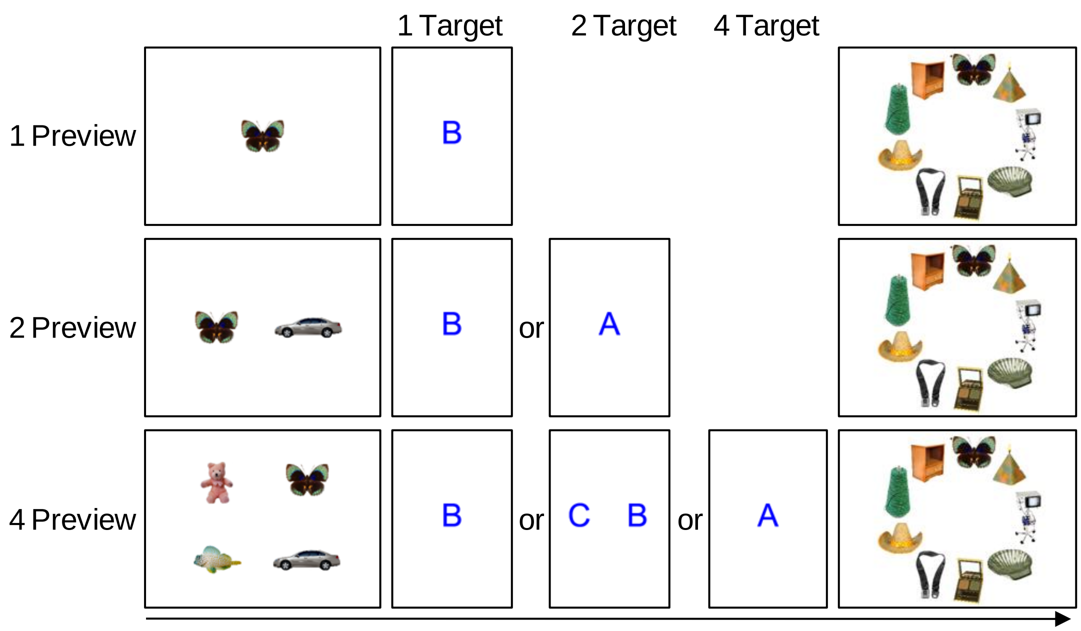

Figure 1. Examples of trials in each condition. Note that stimuli are not drawn to scale.

and another calibration procedure immediately preceding the main experiment. A trial began when the subject looked at a dot in the center of the screen and pressed a button on a game pad with their right thumb. Three different groups of subjects participated in three preview conditions (Figure 1). In the 1 Preview (1P) condition, one image from the sets of teddy bears, butterflies, cars, and fish was presented in the center of the screen for 1.5 seconds. In the 2 Preview (2P) condition, two images out of the four categories (for example, one car and one butterfly) were presented for 3 seconds, one on the left side of the display and the other on the right. In the 4 Preview (4P) condition, a teddy bear, a butterfly, a car, and a fish, were presented for 6 seconds, with one object appearing in each quadrant of the display. Target previews in the 2P and 4P conditions were randomly placed at each location, with the constraint that each preview category appeared equally often at each location. Given that we are interested in capacity limitations on WM, and not encoding limitations per se, we increased the preview duration with the number of target previews so as to make sure that there was sufficient opportunity for subjects to encode all of the previews, to the extent that their WM capacity enabled such encoding.

After a 500 msec delay, again with a fixation dot appearing at the center of the screen, a retro cue was displayed for 2 seconds indicating to subjects the specific targets from the preview that they should look for in the search display. This was to manipulate the number of actual target(s) to be used to guide search. Specifically, the retro cue consisted of three levels: In the 1 Target search condition (1T), the subjects searched for only a single target; in the 2 Target search condition (2T), the subjects had to search for two targets; and in the 4 Target search condition (4T), the subjects searched for four targets at the same time. Using this cue, we can therefore dissociate the number of targets actually used to guide search (1T, 2T, 4T) from the number of target previews presented initially to subjects (1P, 2P, 4P).

In the 1P condition, there was only one target preview, and subjects searched for this one previewed target in the search display. All the trials in the 1P condition, therefore, were 1 Target search trials, and a retro cue in this 1P-1T condition was a letter standing for the target preview; a 'T' for teddy bears, a 'B' for butterflies, a 'C' for cars, and an 'F' for fish. For example, a 'T' was presented if subjects saw a teddy bear preview.

In the 2P condition, there were two target previews, and subjects searched for one of them (1T) or both of them (2T) depending on the retro cue. The retro cue in the 2P-1T conditions was either a T, B, C, or F, meaning that one of the two target previews designated by the cue would appear in the search display. For example, if a fish and a butterfly were followed by a 'B' retro cue, this indicated to subjects that only the butterfly would appear in the search display. The retro cue in the 2P-2T conditions was an ‘A’, meaning that any one of the two target previews could appear in the search display. Subjects would therefore have to search for both targets, as the search display would be equally likely to contain either of the previewed objects.

In the 4P condition, there were four target previews, and subjects searched for one of them (1T), two of them (2T), or all of them (4T) depending on the retro cue. Cues in the 4P-1T conditions, again, were either a T, B, C, or F, meaning that only the one target preview designated by the cue would appear in the search display. The retro cues in the 4P-2T conditions were two letters standing for two of the four target preview types, meaning that one of the two target previews designated by the retro cue would appear in the search display. For example, if the cues ‘F C’ were presented following the four target previews, this indicated to subjects that they should look for both a fish and a car as either might appear in the search display. The retro cue in the 4P-4T conditions was an ‘A’, again indicating that the search display might contain any one of the four target previews. This condition therefore represented the most difficult of our multiple target search tasks.

Following the retro cue, there was another 500 msec delay, again with a central fixation dot, followed by the presentation of the search display. The search display depicted nine objects, one target and eight distractors. These objects were arranged on an imagery circle with a radius of 8.9°. Each object was equally distant from its nearest neighbor (about 6°, center to center). Inserted near the target was a small (0.14°) ‘+’ or an ‘×’ character, placed at some arbitrary location within 2 pixels of the target’s edge. The subject’s task was to find the target and to report whether this character was a ‘+’ or an ‘×’. This discrimination task had been used successfully in previous work (Chen & Zelinsky, 2006; Schmidt & Zelinsky, 2009), and its adoption enabled us to exclusively use target present trials, which are of primary interest to search guidance. Subjects indicated a ‘+’ by pressing the left trigger of the Game Pad with their left index finger, and an ‘×’ by pressing the right trigger of the Game Pad with their right index finger. After their response, there was a feedback display indicating whether the judgment was correct or incorrect. There were 96 1T trials, which were evenly distributed over the four target categories in the 1P condition. In the 2P condition there were also 96 1T trials, but now an additional 24 2T trials. In the 4P condition there were 96 1T trials, 24 2T trials, and 24 4T trials.

2.2.3. Results and Discussion

Error rates were less than 3% in all conditions, and error trials were excluded from all subsequent analyses. These low error rates indicate that subjects were able to perform the +/× discrimination task successfully.

We introduced retro cues to independently assess the contributions of the capacity-limited WM component and the feature mismatch component on search guidance. Before the presentation of the retro cue subjects would have to hold all the target previews in WM. For example, in the 4P condition, all four target previews must be encoded and held in WM until the retro cue. However, depending on the retro cue, subjects in the 4P conditions could restrict their search to one, two, or four targets. Before the retro cue the 4P conditions therefore held constant the WM load at four objects, meaning that these conditions did not differ with respect to their WM demands, but after the cue the number of templates that would be used for search guidance varied from one to four. Any difference among the 4P 1T, 2T, and 4T conditions could therefore be attributed to imprecise guidance resulting from mismatch between features in WM and features in the search display. Contrast this scenario with a comparison between the 1P-1T, 2P-1T, and 4P-1T conditions. Here, only the WM load was manipulated, not the number of guiding templates, which in each case was only a single target. Any difference among these conditions could therefore be attributed to a limitation on WM capacity.

The mean times taken by subjects to make their button press responses (RTs) are shown in Table 1. Note, however, that these measures reflect the time needed for subjects to not only find the target, but also to locate the + or × adjacent to this object and to discriminate between these characters. Because they lump together guidance with various delays specific to the discrimination judgment, overall RTs in our task cannot be directly compared to the RT measures of search efficiency more commonly reported in target present/absent search tasks. To properly analyze our data we divided the overall RTs into separate search guidance and discrimination components. Search guidance, the primary focus of this study, is indicated in this measure by the time between the onset of the search display and the subject's first fixation on the target, what we refer to as RT-to-target (or RTT). These data are shown in Figure 2a. The second discrimination component reflects the time between the first fixation on the target and the button press

Table 1
Mean manual RTs in Experiment 1. Numbers in parentheses indicate standard error.

	1 Target	2 Target	4 Target
1 Preview	1529.04 (70.41)		
2 Preview	1600.20 (70.38)	1771.76 (94.39)	
4 Preview	1677.18 (80.36)	2041.33 (93.84)	2034.21 (86.78)

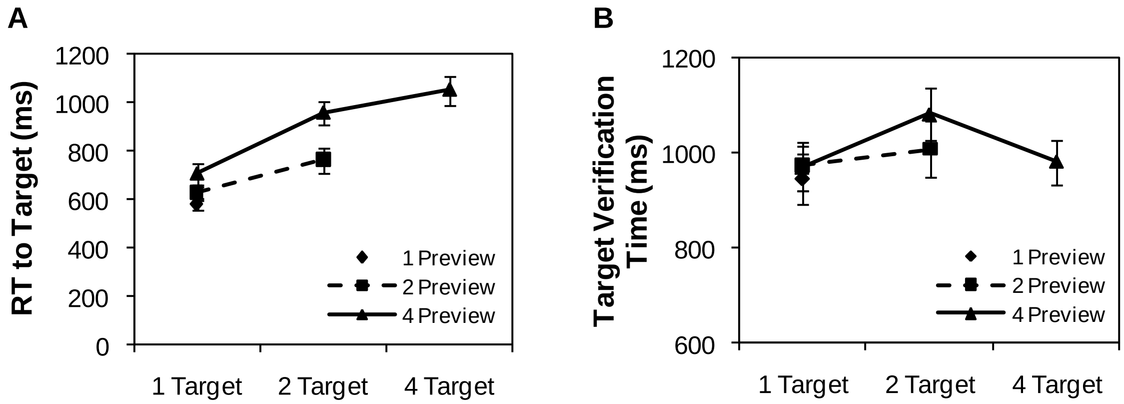

Figure 2. RT to target (a) and discrimination time (b) in experiment 1. Error bars indicate standard error.

response. These data are shown in Figure 2b for completeness, but are largely uninterpretable and not central to the questions of this study.

To determine whether subjects needed more time to search for multiple targets, we compared RTT in the 1P-1T, 2P-2T, and 4P-4T conditions (Figure 2a). This analysis revealed a significant difference between conditions, $F(2,42)=24.1$, $p<.01$, which we then clarified using LSD post-hoc tests. Consistent with a standard demonstration of a multiple target effect, we found that RTTs were longer in the 2P-2T condition compared to the 1P-1T condition ($p<.05$). We also found a further increase in RTT in the 4P-4T condition compared to the 2P-2T condition ($p<.001$). This indicates that the load effect is not limited to a difference between one and multiple targets; as the number of previewed targets increase, subjects require more time to find the target.

To better understand the nature of this load effect, we conducted separate ANOVAs to test the capacity-limitation and feature-mismatch hypotheses. The capacity-limitation hypothesis predicts that an increase in the number of previews should result in worse guidance, and therefore longer RTTs. Consistent with this prediction, the main effect of preview number was significant in the 1T conditions, $F(2,42)=3.6$, $p<.05$. Post-hoc tests revealed faster RTTs in 1P-1T compared to 4P-1T ($p<.05$). However, RTT in 2P-1T was only marginally faster than 4P-1T ($p<.1$), and did not differ from 1P-1T ($p>.3$). The main effect of preview number was significant in the 2T conditions, $F(1,28)=7.6$, $p<.05$, meaning faster RTT in 2P-2T than in 4P-2T.

In contrast to this mixed evidence for an effect number of previews on search, we found a robust effect of the number of cue-specified targets. As predicted by the feature-mismatch hypothesis, the main effect of number of targets was highly significant in 2P condition, $F(1,14)=9.5$, $p<.01$, showing faster RTT in 2P-1T than in 2P-2T. The main effect of number of targets was also highly significant in the 4P condition, $F(2,28)=34.1$, $p<.001$. Through post-hoc testing we found faster RTT in 4P-1T than in 4P-2T ($p<.001$), which was faster than 4P-4T ($p<.05$). Together, this preliminary analysis indicates some support for effects of both a WM capacity limitation and feature mismatch on search guidance, with the contribution of mismatching features being the clearer effect.

As expected, analysis of the discrimination times failed to reveal meaningful differences between conditions (Figure 2b). There were no significant main effects of number of previews in 1T, $F(2,42)=.09$, $p>.92$, and in 2T conditions, $F(1,28)=.85$, $p>.36$, nor was there a significant main effect of number of cue-specified targets in the 2P conditions, $F(1,14)=2.5$, $p>.14$. However, there was a significant main effect of number of cue-specified targets in the 4P conditions, $F(2,28)=9.6$, $p<.01$. Post-hoc analysis showed that this effect was due to longer discrimination times in the 4P-2T condition than in the 4P-1T ($p<.01$) and in the 4P-4T conditions ($p<.01$). We speculate that this difference was due to subjects needing extra time to confirm that the target was one of the two cued target previews. There was no difference between the 4P-1T and 4P-4T conditions (p>.67). Except for the 4P-2T condition, we did not find any effect of the number of previews and the number of targets on discrimination time. This means that the effects of multiple targets on search are limited largely to effects on search guidance, at least for the discrimination task used in the present study.

Although RTT is an accepted measure of search guidance (e.g., Castelhano & Henderson, 2007; Yang & Zelinsky, 2009), it is by no means the best measure. First, it includes the time needed to reject all of the distractors fixated before the target. Given that each of these distractor rejections require verifying that the fixated object is indeed a distractor, the accumulation of these individual decisions adds to the measure a potentially large amount of variability that is unrelated to the actual guidance process. Second, RTT has no clear baseline against which a chance level of guidance can be compared. This requires the limitation of conclusions to only relative guidance differences between conditions. To address these concerns we analyzed our data using a more conservative measure of search guidance, one that defines guidance in terms of the percentage of trials in which the target was the first object fixated during search.

Figure 3 shows the percentage of immediate target fixations for the preview and target conditions. If search was unguided and randomly directed to objects, the target should be fixated initially on only 11% of the trials. Analyses revealed immediate target fixation rates that were clearly above chance in all conditions, ts(14)>=2.8, ps<.05. This suggests that search was guided even when subjects searched for four objects at the same time.

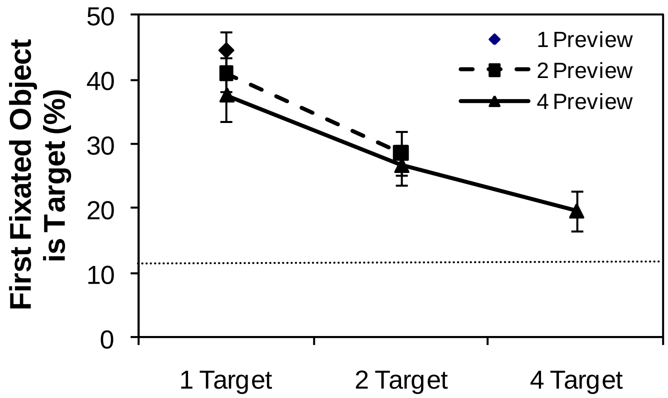

Figure 3. Percentage of trials where the target was the first fixated object. Error bars indicate standard error. The dotted line indicates chance level of guidance.

However, it is also clear from Figure 3 that search guidance varied between conditions. As in the case of RTT, we compared immediate target fixation rates in the 1P-1T, 2P-2T, and 4P-4T conditions to determine that guidance decreased with the addition of multiple search targets. We found a significant difference between conditions, $F(2,42)=15.8$, $p<.001$. Post-hoc tests again confirmed that this difference was significant between both the 1P-1T and 2P-2T conditions ($p<.01$) as well as between the 2P-2T and 4P-4T conditions ($p=.05$). Increasing the WM and search load from one to two to four targets not only increases the time needed to fixate the target, it also reduces the degree of immediate search guidance.

So as to better understand the nature of this load effect, we again subjected our data to separate ANOVAs to test the capacity-limitation and feature-mismatch hypotheses. The capacity-limitation hypothesis predicts that an increase in the number of previews should result in worse guidance, and therefore fewer immediate fixations of the target. Contrary to this prediction, and to our analysis of the RTT data, we found no main effects of preview on initial target fixations in either the 1T, $F(2,42)=1.2$, $p>.3$, or 2T conditions, $F(1,28)=.18$, $p>.7$. In contrast, there was a highly significant main effect of target in both the 2P condition, $F(1,14)=17.0$, $p<.01$, and the 4P condition, $F(2,28)=13.7$, $p<.001$, a pattern consistent with the feature-mismatch hypothesis. Subsequent analyses confirmed that guidance decreased between the 4P-1T and 4P-2T conditions ($p<.001$) and also between the 4P-2T and 4P-4T conditions ($p<.07$). This analysis helps to clarify the slightly unstable patterns found in the RTT data; it is the feature mismatch created by holding multiple target representations held in visual WM that reduces search guidance, not an encoding constraint imposed by a limited WM capacity.

Perhaps the absence of an effect of previews in the above analysis is due in part to subjects differentially coding these objects, thereby causing some to be more heavily weighted than others. For example, if subjects spend a longer time viewing one object, or have a bias for the most recently viewed object, then this single object might be used to guide search in those cases when the retro cue fails to narrow the selection of guiding target representations. To explore this possibility we analyzed the oculomotor scanning behavior of subjects as they viewed the two and four target preview displays.

Figure 4 breaks down the target preview viewing behavior in three ways: cases in which the object that would appear as the target in the search display was the first object fixated during preview viewing (to look for a preview primacy effect), cases in which the search target was the last object fixated during preview viewing (to look for a preview recency effect), and cases in which the search target was the object fixated longest during preview viewing (to look for a duration-related encoding bias).

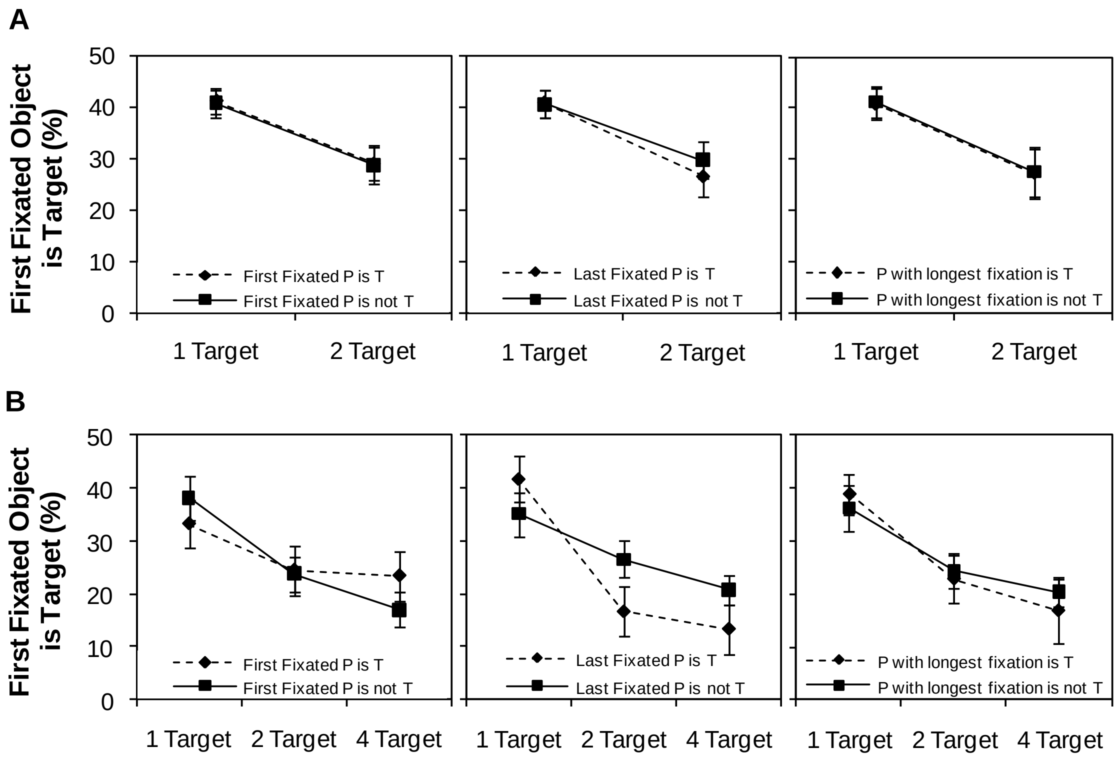

Figure 4. Preview analyses for 2P (a) and 4P (b) conditions. Error bars indicate standard error.

Separate analyses were conducted for the 2P and 4P conditions. In the 2P condition, there were no guidance benefit associated with the first object fixated during preview, $F(1,14)=1.0$, $p>.3$, the last previewed object fixated, $F(1,14)=1.0$, $p>.3$, nor the previewed object fixated the longest, $F(1,14)=.005$, $p>.9$. For all three measures there was only a general decrease in guidance with an increase in the number of cue-specified targets, $F(1,14)=15.9$, $p<.01$ for the first previewed object, $F(1,14)=18.9$, $p<.01$ for the last previewed object, and $F(1,14)=19.1$, $p<.01$ for the previewed object fixated longest. Similarly, in the 4P conditions there were no guidance benefits associated with the first fixated preview, $F(1,14)=.07$, $p>.7$, with the last fixated preview, $F(1,14)=1.6$, $p>.2$, or with the longest fixated preview, $F(1,14)=.03$, $p>.8$. Planned comparison showed that there was a guidance difference only for the last fixated preview in the 1T condition,

$t(14)=2.5$, $p<.05$. There were no other significant differences. Again, there was a general decrease in guidance with an increase in the number of cue-specified targets in all three measures, $F(2,28)=9.1$, $p<.01$ for the first previewed object, $F(2,28)=19.8$, $p<.001$ for the last previewed object, and $F(2,28)=11.4$, $p<.001$ for the previewed object fixated longest. These analyses therefore provide no reason to believe that differential encoding of targets during preview viewing resulted in the absence of preview-specific guidance in this experiment.

In summary, our goals in this experiment were to determine if multiple target search is less efficient and less guided than single target search, and if so, is this load effect better explained as an encoding limitation on visual WM or in terms of mismatching features between the search target and the guiding target representations. Using both RT-to-target and immediate target fixation measures of guidance, we found that search guidance indeed does decrease with the addition of multiple targets. Our findings extend previous work on this topic by demonstrating that this load effect is reflected in actual search guidance (rather than non-guidance related decisional factors), appears in the context of real-world objects, and is not specific to the search for one versus two targets; guidance decreases out to the maximum of four targets tested in this experiment. However, as discussed this load effect is ambiguous as to its cause. By adopting a retro cue paradigm, we were able to partially disambiguate the cause of this load effect. Although the RTT measure suggested some evidence for the contribution of both an encoding limitation and feature mismatch, analysis of immediate target fixations, the more conservative guidance measure, clearly suggested that mismatching features were responsible for the effect of target load on search guidance. The fact that this evidence appeared most clearly in the immediate fixation measure may suggest that mismatching features play their greatest role very early in search, with WM factors exerting more of an influence on guidance as the search decision starts to converge on the target.

2.3. Experiment 2

2.3.1. Introduction

In experiment 1 we showed that multiple target search is less guided than single target search. We also showed that this load effect was not due to capacity limitations on visual WM, arguing instead that mismatching features were the cause of this reduced guidance. Because there was only one target in the search display yet multiple target representations in visual WM, the features from these target representations that did not match the actual search target would not only weaken the target guidance signal, but also potentially match distractor objects, thereby guiding search *away* from the target. However, there is an alternative explanation for the reduced guidance that we found with multiple targets. It may be that these multiple target representations were indeed available to guide search (i.e., no WM capacity limitations), but that subjects were simply choosing not to use all of these representations, instead perhaps picking only one to guide their search. When subjects happen by chance to pick the right target representation (the one actually matching the search target), there will be good guidance. When they happen to guess the wrong target, there will be no guidance. Guidance should therefore decrease with the addition of each new target representation to visual WM, as the probability of correctly guessing the search target becomes smaller.

In experiment 2 we directly tested whether people search for multiple targets simultaneously or whether they pick one of these possible target representations to guide their search. Two target previews were shown on each trial, one a teddy bear and the other a car; search displays were manipulated using three conditions. In the 'both T' condition, both of the two previewed targets appeared in search display. In the 'same T' condition, one of two previewed targets appeared twice in search display; that is, either two instances of the previewed teddy bear or the previewed car. In the 'one T' condition, only one of the previewed targets appeared in search display.

If both target previews are held in WM and used to guide search, guidance should be same regardless whether the search display depicts both targets or two instances of the same target. For example, if you are holding a teddy bear and a car in visual WM, it would not be matter whether there are a teddy bear and a car in search display or two instances of the teddy bear (car) - search should be guided equally well in these two cases. We would therefore predict no difference in guidance between both T and same T conditions. Guidance, however, should be better in the both T and same T conditions compared to the one T condition for purely probabilistic reasons; because there are two targets and six distractors in the two-target conditions (both T and same T conditions) while there is only one target among seven distractors in the one T condition, the probability of finding the target simply by chance is higher when there are two targets.

In contrast, if only one target is used to guide search, better guidance is expected in the both T condition than in same T condition, because there should be no guidance in the same T condition when the subject picks the wrong target representation to guide their search. In the Both T condition, regardless of which object the subject picks, guidance should still be good because both targets are present in the search display. For example, if the car was selected to guide search, guidance to the car target in the search display should be strong. On average, guidance should only be half as strong in the Same T condition because, on half of the trials, the wrong target will be selected. If the car is selected, there will be no guidance on trials when the search display depicts two teddy bears. Again, should subjects correctly guess the target, better guidance would be expected in the same T condition than in one T condition as a result of the increased probability of finding one of the two targets.

2.3.2. Method

Participants

Seventy-two students from Stony Brook University participated in the experiment for course credit. All had normal or corrected to normal visual acuity and were naive with regard to the experimental hypotheses.

Stimuli and Apparatus

Teddy bears from Cockrill (2001) and cars from the Hemera Photo Objects Collection and random web sites were modified and used as targets. Distractors were real-world objects randomly selected from the Hemera Collection. The apparatus were identical to what was described for experiment 1. None of the distractors belonged to the two target categories. All objects were presented in color and resized such that each would just fit into a 5,000 pixel bounding box. None of the stimuli were used more than once.

Procedure

The experiment began with an initial calibration procedure, followed by practice trials and another calibration. Each trial began with a button press after fixation on a central dot. Target previews consisting of a teddy bear and a car were presented for 3 seconds, one 4.5° to the left of central fixation and the other 4.5° to the right of central fixation. Each category of object was equally likely to be presented on either side of the display. After a 1 second delay with a central fixation dot, an 8-object search array was presented, with these objects arranged into a ring with a radius of 8.9° surrounding central fixation. There were three different search display conditions, manipulated between subjects. In the both T condition, both previewed targets appeared in the search display, along with six distractors. In the same T condition, two identical teddy bears or two identical cars from the preview appeared with six distractors in the search display. In the one T condition, one of the previewed targets and seven distractors appeared in the search display. In order to determine whether search behavior would be affected by the separation between the two targets in the both T and same T conditions, in one quarter of the trials the two targets appeared next to each other (no distractors between them), and in the remaining three quarters of the trials the two targets would have either one, two or three distractors between them, with equal probability. Targets in one T condition were randomly assigned to one of the eight possible display positions, with the constraint that a target appeared equally often in each location. The target in the one T condition and one of the two targets in the both T and same T conditions were accompanied by a '+' or an '×', and the insertion of this character was evenly divided between teddy bears and cars. The task was to report whether this character was a '+' or an '×'. Subjects pressed the left trigger of the Game Pad with their left index finger when the target was accompanied

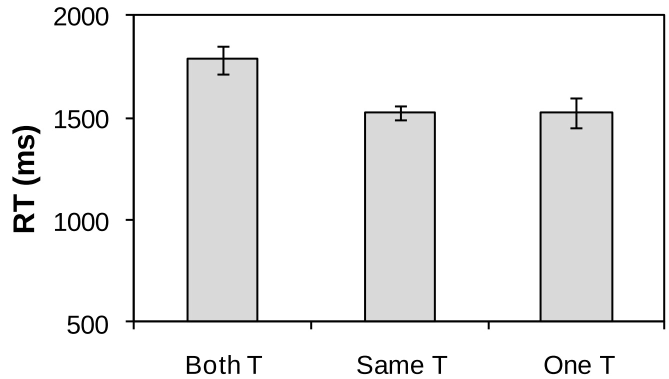

Figure 5. Mean manual RTs in Experiment 2. Error bars indicate standard error.

by a '+', and they pressed the right trigger with their right index finger to indicate an '×'. Accuracy feedback followed the response. Each condition had 128 trials.

2.3.3. Results and Discussion

Error rates were less than 3% in all conditions, and error trials were excluded from all subsequent analyses.

Analysis of RTs revealed significant differences among the three conditions (Figure 5), $F(2,69)=6.4$, $p<.01$. Through LSD post-hoc tests we found that RTs in the both T condition were longer than those in the same T and one T conditions (both $p<.01$). RTs did not differ between the same T and one T conditions ($p>1.0$). These differences were not due to differential times needed to execute the initial saccades. Initial saccade latencies were 198 ms, 195 ms, and 187 ms for the both T, same T and one T conditions, respectively, with these latencies failing to reliably differ $F(2,69)=1.2$, $p>.3$.[4] We

[4] We think it is interesting that the latencies of initial saccades directed to targets did not differ between the one T condition and the two-target conditions (both T or same T). Although recent studies have shown that initial saccade latencies are largely immune from Hick's Law (Kveraga, Boucher, & Hughes, 2002; Lawrence, St. John, Abrams, & Snyder, 2008), meaning that increases would not be expected due strictly to an increase in the number of alternatives, it is also fairly well established that initial saccade latencies increase with target-distractor similarity (e.g., Ludwig & Gilchrist, 2002). We therefore expected latencies to increase in the two-target conditions simply due to the competition created by two target similar patterns

therefore attribute the longer RTs found in the both T condition to later decision processes associated with the manual response, and not to search guidance. We also failed to find any effect of distance between the two targets on initial saccade latencies in either the both T condition, $F(3,69)=1.7$, $p>.2$, or the same T condition, $F(3,69)=.16$, $p>.9$. There was also no effect of target distance on immediate target fixations in the both T condition, $F(3,69)=2.2$, $p>.1$. Although there was a small effect of target separation on immediate target fixation in the same T condition, $F(3,69)=3.5$, $p<.05$, the effect appeared for only one level of target separation. Rather than showing a linear decrease in guidance with increasing distance between the two targets, post-hoc tests revealed less guidance to targets separated by two distractors compared to targets in the other conditions ($p<.05$ for all comparisons; mean immediate target fixations were 62.7, 61.6, 55.7, and 62.6 in the 0-3 separation conditions, respectively). Regardless of whether the two targets were the same or different objects, search behavior was largely unaffected by their separation in the search display. For all subsequent analyses we therefore collapse across the target separation condition.

To determine whether there were differences in guidance among the three conditions, we again compared the percentages of first object fixations on targets. These results are shown in Figure 6a. Immediate target fixations occurred far more frequently than what would be expected based on their respective chance baselines, $t(23)\geq10.2$, $p<.001$, indicating that search guidance was quite pronounced in this task. We also found a significant difference in guidance among the three conditions, $F(2,69)=12.4$, $p<.001$.

(indeed, actual targets) appearing in the search display, with each ostensibly vying for attention. The fact that we did not observe this increase raises the intriguing possibility that the process of selecting a target from a target-similar distractor may require competition, but deciding which of two targets to look at (essentially selecting one of two targets) does not.

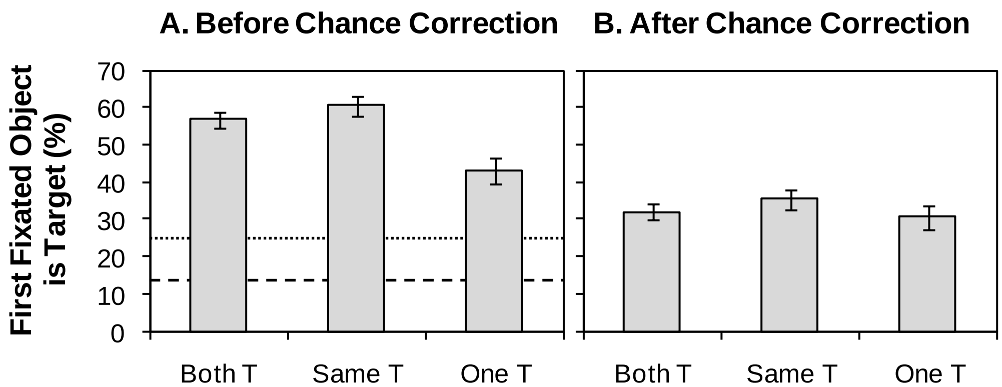

Figure 6. Percentage of trials where the target was the first fixated object. Error bars indicate standard error. The dotted line indicates chance level of guidance for the Both T and Same T conditions; the dashed line indicates chance level of guidance for the One T condition.

However, post-hoc analyses revealed that there were no differences in immediate target fixations between the both T and same T conditions, $p>.3$. Indeed, contrary to the prediction of better guidance in the both T condition, there was a numerical trend towards worse guidance in this condition compared to same T. As expected, we did find that the percentage of immediate target fixations was significantly lower in the one T condition than in the other two conditions, *p*s<.001, which we can attribute to the different chance baselines in these conditions; the one T chance baseline was 12.5% (=1/8*100) whereas the both T and two T chance baseline was 25% (=2/8*100). However, after correcting for these different chance levels by subtracting each condition's baseline from the observed guidance level, we found no differences in guidance between any of the three conditions, $F(2,69)=.98$, $p>.3$ (Figure 6b).

These findings suggest that multiple target representations are held in visual WM and used to guide search. Had subjects simply picked one of the two target preview representations to guide their search, guidance should have been stronger in the both T condition, because regardless of which guiding representation was chosen it would have matched a search target. This would not be the case for the same T condition, in which half of the time the chosen representation would not have matched the search target. The fact that guidance did not differ between these conditions, and indeed was numerically stronger in the same T condition, allows us to rule out this target selection possibility. We are therefore left with our original formulation of the feature-mismatch hypothesis; multiple target search is less guided than single target search due to the inevitable feature mismatch that results from having multiple target representations guiding search, but only having a single target appearing in the search display.

2.4. General Discussion

Previous studies provided mixed results on whether multiple target search is less efficient than single target search. Most of them used simple stimuli such as letters or colored bars. However, searching for two letters or two bars at the same time is far from what we do in our everyday lives. We typically search for more complex objects that consist of many different colors, shapes, and textures. Also, there is no limit on the number and kinds of other objects that might serve as distractors in real-world contexts. These distractors can also be very different from each other, and at the same time they could be very similar to targets. All of these factors could affect search performance, thereby making it hard to generalize from simple stimuli to our everyday search.

In this study we wanted to see if the search for multiple real-world targets is less guided and less efficient than the search for a single real-world object. We found that this is indeed the case, and in fact further increasing the task load by adding more targets further decreased search efficiency and guidance. Additionally, we wanted to explore the causes of this load effect; why is multiple target search less guided? We considered two hypotheses, one in terms of a capacity limitation on visual WM encoding, and the other in terms of mismatching features during guidance. Although our RTT data partially supported both explanation, analysis of the first fixated objects, a conservative measure of search guidance, strongly supported only the feature-mismatch hypothesis. Having more previews, and thus increasing the potential for WM capacity limitations, did not affect search guidance; decreased guidance resulted from too many target features (from multiple targets) weakening the guidance signal to the single target in the search display and generating spurious matches to the distractors. Although we cannot rule out any factor influencing WM after the retro cue, this is nevertheless a new and important finding, as most previous work had just assumed that load effects on search were due to capacity limitations on WM (e.g., Menneer, Cave, & Donnelly, 2009).

Might our conclusions be different had we presented more than four objects in the preview? Certainly, if 6, 8, 12 or more objects appeared in the preview we would have observed more pronounced effects of WM encoding limitations on search. However, this would not be terribly interesting. There is now quite a large literature suggesting that when WM is taxed beyond about four things, regardless of whether these things are features or objects, or whether the reason for the limitation is slots or graded resources, performance will suffer. The goal of our study was not to speak directly to the reason for this WM limitation, but rather to explore whether such a WM limitation is responsible for load effects commonly observed in the context of search. Despite having a WM load clearly within most accepted limits of WM capacity (2-4 objects), we nevertheless observed a pronounced load effect in our task. Moreover, given the visually complex objects used in this study it was possible that this was due to a WM encoding limitation. As already summarized, this was not the case. The clarity of this conclusion for feature-mismatch would not have been possible had we imposed a WM constraint, as we would not have been able to separate WM effects from feature-mismatch effects as cleanly.

In experiment 2 we clarified the effect of feature mismatch on search that we found in experiment 1. Although experiment 1 ruled out a WM encoding limitation as a source of the load effect, it did not provide conclusive proof for the feature mismatch hypothesis; subjects may have been simply choosing not to use all of the target representations to guide their search. To test for this possibility we presented two targets in the search displays. We found that guidance was the same regardless of whether subjects were searching for two different targets or two instances of the same target. This pattern would only be expected if subjects were holding representations of both targets in visual WM, and using both to guide their search. When combined with experiment 1, we can therefore definitively conclude that effects of target load on search guidance in our task were due to mismatching features between the target representations in WM and the target appearing in the search display.

In some sense, our decisive failure to find evidence for an effect of WM capacity limitations on search guidance is counterintuitive. Given our use of visually complex real-world objects, one might think that resource limitations on visual WM (e.g., Bays & Husain, 2008) would prevent the additional targets in the 2P and 4P conditions to be less precisely represented than the single target in the 1P condition. This would be particularly true if features, and not objects, were the determining factor in visual WM capacity, as has been suggested (Alvarez & Cavanagh, 2004; Eng, Chen, & Jiang, 2005).

However, it may be possible to reconcile a capacity-limited encoding process with the fact that we found no evidence for such a limitation on search guidance. A WM capacity limitation at encoding should only affect search guidance to the extent that the excluded features are important to the guidance process. If the process of creating a target representation is intelligent enough to select only those features that best discriminate between the target and expected distractors, then only these features would need be encoded into visual WM, allowing the rest to be discarded. Assuming that the total number of these discriminative target features does not exceed visual WM capacity, it may be that efficient guiding representations can therefore be formed for multiple targets even with a limited WM capacity. This possibility will be considered more fully in Chapter 3, where we attempt to distinguish between the use of discriminative features and common features to guide multiple target search.

3. Is multiple target search guided by distinctive features or by common features?

3.1. Introduction

In the previous experiments, we showed that multiple target search for real-world objects is less efficient and less guided than single target search. To determine what factors affected efficiency and guidance for multiple target search, we used retro cues to separately explore the effect of WM capacity limits during preview encoding and the effect of feature mismatch during actual search. We found that only mismatching features, and not an encoding limitation on visual WM, was responsible for the reduced search guidance. We also argued that this decreased guidance for multiple targets was indeed due to mismatching features from multiple targets rather than the hit-or-miss selection of only one of the viable targets to guide search. The features of targets held in visual WM but not appearing in the search display would create mismatch signals that weaken guidance and reduce search efficiency.

The present series of experiments were motivated by two goals. First, our evidence against the use of a “pick one target and discard the other” guidance strategy was based on a task in which guidance was compared between a condition in which both targets appeared in the search display and a condition in which two instances of the same target appeared in the search display. Although we believe that this experiment produced compelling evidence against this strategy, we acknowledge the fact that having two identical targets appearing in a search display is an unorthodox manipulation that may have affected search behavior in some unknown way. For this reason we sought to test this same possibility using a completely different method, thereby providing converging evidence for our conclusion in favor of the feature-mismatch hypothesis. Second, given our finding that the features from multiple targets are held in memory and used to guide search, one logical next question asks how the targets in multiple target search are represented. Specifically, are they represented by the common features shared by all the targets or by distinctive features unique to each of the targets? As a simple example of this distinction (focusing just on color), suppose subjects were searching for a brown teddy bear with a red bow and a yellow race car with a red rear foil. Do subjects extract the common feature of red and guide their search using this unitized target representation, or do they extract the distinctive features of brown and yellow and guide their search using these object-specific representations? Either method of target representation would be expected to produce worse guidance relative to single target search, but for very different reasons. In the case of common features, the target would be defined too imprecisely to generate strong guidance (Schmidt & Zelinsky, 2009); in the case of distinctive features, reduced guidance would again be expected due to the feature mismatch problem. The following series of experiments were designed to address the nature of the target representation underlying multiple target search, and how this representation might change with the similarity of the previewed targets.

3.2. Experiment 1

3.2.1. Introduction

To address both of our research goals, we devised an estimate of multi-target search guidance that we refer to as the *guidance index* (GI). Stately simply, the GI is the proportion of two target search guidance to single target search guidance, where guidance is corrected for chance. Stated more formally:

$$GI = (G_{2P} - c) / (G_{1P} - c),$$

where G_{2P} and G_{1P} are the levels of guidance observed in two-preview and one-preview search conditions, respectively, and c is the level of guidance expected by chance, which was a constant 11.1% in the following experiments given a set size of 9.

A simple example will help to illustrate how the GI can be used to tease apart representational issues important to understanding multiple target search. Let us assume that a person picks only one of two possible targets to guide their search (the “pick one and discard the other” hypothesis). If this chosen target then actually appears in the search display, guidance to this target should be as strong as when only this single target was presented at preview, given our previous evidence for the absence of encoding limitations on visual WM. However, if this person picks the wrong target (the one that does not appear in the search display), guidance should be at chance. Assuming that these two scenarios occur equally often, half of the trials should show relatively strong guidance (as good as single target search), while the other half of the trials should show no guidance. The overall guidance level in two target search, assuming a “pick one and discard the other” target representation, should therefore be half of the guidance level found in single target search, yielding a proportion of two-target to one-target guidance (GI) of .5.

One of our goals in formulating the GI is to define a theoretical upper bound on the level of guidance in a multiple target search task that might exist if only one of the designated targets was actually used to guide search. In the above formulation, our intension was to have this upper bound be .5, with any GI above .5 indicating proof positive for the use of features from both targets to guide search. However, there is one scenario that might produce a GI exceeding .5 even when a person is using only one target to guide their search. Let us assume that a yellow teddy bear and a yellow car were designated as targets during preview. If a subject decided to represent only the yellow teddy bear in their visual WM (and not the yellow car), but the yellow car appeared as the target in the search display, there might be some above-chance level of guidance to the yellow car due to the fact that both targets had in common the yellow feature. Consequently, the GI would be greater than .5 even though the subject would have picked only one target to guide their search. Formulating a theoretical upper bound on the guidance contribution from one target in multi-target search task therefore requires factoring out the guidance expected due to the targets having common features.

To estimate the level of guidance due to common features we conducted two single target control experiments, one for each of the two target sets used in the multiple target experiment. For example, if a trial in the multiple target experiment designated a yellow teddy bear and a yellow car as targets, then the corresponding trial in the teddy bear control experiment would show only the yellow teddy bear target at preview, but would have the yellow car appear as a lure in the target absent search display. Likewise, the corresponding trial in the car control experiment would have only the yellow car appear at preview, and the yellow teddy bear appear as the search lure. By doing this, we can determine the degree of guidance to the yellow car that results from searching for only the yellow teddy bear, and conversely, the degree of guidance to the yellow teddy bear that results from searching for only the yellow car. These levels of guidance can then be subtracted from the two-target level of guidance for that trial, thereby correcting the common feature confound and reinstating a GI of .5 as a theoretical upper bound. Of course this correction for common feature guidance requires amending the GI, which now becomes:

$$GI = \{(G_{2P} - c) - (G_L / 2)\} / (G_{1P} - c),$$

where, G_L is the level of guidance to the lures obtained from the control experiments. Note that we divide by two this lure-related guidance because it should arise on only half of the trials, specifically those in which subjects picked the wrong target preview to represent and guide their search.

By including a control for common features in the GI, this measure also allows us to address our second goal, determining whether the guiding representation for multiple targets consists of common or distinctive features. Because we explicitly estimate and subtract the contribution of common features from the two-target level of guidance, a GI > .5 would specifically indicate guidance due to distinctive features from both targets. Similarly, if subjects were consistently choosing to represent the targets in terms of common features, this would be expressed by a GI < .5. Although the GI does not rule out the possibility that both distinctive and common features are used to guide search, it does allow us to identify the dominant contributing factor; a GI greater than .5 indicates a net guidance contribution by distinctive features, a GI less than .5 indicates a net guidance contribution by common features. A GI of .5 would be less informative with respect to multiple target search, indicating either the roughly equivalent use of both common and distinctive features, or perhaps the use of only one of the two targets to guide search (i.e., the "pick one and discard the other" strategy).

3.2.2. Method

Participants

Forty students from Stony Brook University participated in the experiment for course credit. Twenty four of these students participated in the main experiment, and sixteen participated in the control experiments. All had normal or corrected to normal visual acuity, by self report, and were naive with regard to the experimental hypotheses.

Stimuli and Apparatus

The targets were images of teddy bears and cars. Teddy bears were obtained from Cockrill (2001), and the cars were selected from the Hemera Photo Objects Collection (Gatineau, Quebec, Canada) and random web sites and were modified for use in this experiment. The distractors were again images of real-world objects selected from the Hemera Photo Objects Collection, and none were members of the teddy bear or car categories. All of the objects were resized such that the smallest bounding box enclosing each consisted of exactly 5000 pixels. This manipulation resulted in the widest object subtending ~4° and the tallest object subtending ~3.6°. All objects were presented in color, and no object, target or distractor, was presented more than once.

The experiment was designed and controlled using Experiment Builder (SR Research Ltd, version 1.5.1), and eye movements were recorded using an EyeLink II eye tracking system (SR Research Ltd). Eye position was sampled at 500 Hz, and a calibration was not accepted unless the maximum spatial error was less than 1.0°. We used a chinrest to prevent head movement and to keep a fixed viewing distance between the monitor and subjects (72cm). Stimuli were displayed using a ViewSonic 19" flat-screen CRT monitor at a refresh rate of 100 Hz. The viewing angle of the display was 26° horizontally and 20° vertically. Manual responses were collected using a Game Pad controller interfaced through the computer's USB port.

Procedure

Main experiment. Following calibration and practice each experimental trial began with a button press after fixation on a central dot. A teddy bear, a car, or both objects were then presented as previews, depending on the condition. In the 1P-teddy bear condition, an image of a teddy bear was presented in the center of the screen for 1.5 seconds, and in the 1P-car condition, an image of a car was presented in the center of the screen for 1.5 seconds. In the 2P condition, a teddy bear and a car, one 4.5° to the left of center and the other 4.5° to the right of center, were presented for 3 seconds. After a 1 second delay, during which the subject was fixating a dot at the center of the screen, a search display with one target and eight distractors was presented. The target was the previewed teddy bear in the 1P-teddy bear condition, the previewed car in the 1P-car condition, and either the previewed teddy bear or car in the 2P condition. In the case of the 2P condition, half of the trials had a teddy bear target in the search display, and half had a car target in the search display. The target had a '+' or an '×' character inserted near it, and the subject's task was to find the target and to report if this character was a '+' or an '×'. Subjects pressed the left trigger of the Game Pad with their left index finger to indicate a '+' and they pressed the right trigger with their right index fingers to indicate an '×'. Accuracy feedback was provided following this response. Conditions were blocked, and block order was counterbalanced. Each block had 64 trials, for a total of 192 experimental trials per subject.

Control experiment. A separate group of subjects participated in two control experiments. Their task was target present-absent search for only a single target category. Half of the subjects searched for a teddy bear (bear control condition); the other half searched for a car (car control condition). The 32 target-present trials of the bear control condition were randomly selected from the 1P-teddy bear trials from the main experiment. In these trials, one bear was presented as a preview for 1.5 seconds. After a 1 second delay, again with a central dot, a search display with one target and eight distractors was presented. The target was the previewed teddy bear as in the 1P-teddy bear condition. The only change relative to the corresponding 1P teddy bear condition trial was the appearance of a car in the search display. Because these subjects would only be searching for a teddy bear, and never a car, the car would be just another distractor appearing in the search display. The 32 target-absent trials of the teddy bear control condition were the 32 trials from the 2P condition of the main experiment where the actual target in search display was a car. This car was the car presented at preview in the corresponding 2P condition of the main experiment. By having this car now appear as a distractor in the search display, we can quantify the degree of guidance to this car due to its similarity to the teddy bear target, and we do this using the actual car and teddy bear objects designated as targets in the 2P condition from the main experiment.

The procedures for the car control condition were the same as those for the teddy bear control condition, with the predictable exceptions. The 32 target-present trials of the car control condition were randomly chosen from the 1P-car condition from the main experiment. In the search display, this car was presented as the target along with eight distractors. One of the distractors would now be a teddy bear. The target-absent trials of the car control condition were the 32 trials from the 2P condition in the main experiment where the actual target in search display was a teddy bear, meaning that this was the same teddy bear presented at preview in the corresponding 2P condition of the main experiment. Again, by having this teddy bear appear as a distractor, we can quantify the degree of guidance to this object due to its similarity to the car target.

Using these control experiments, for each bear-car pair from the 2P condition of the main experiment, we can quantify whether search is preferentially guided to the car (relative to the other distractors) when subjects were searching for the teddy bear target, and whether search is preferentially guided to the teddy bear (relative to the other distractors) when subjects are searching for the car. We then subtract this level of guidance due to any feature overlap between the specific bear and car objects from the 2P level of guidance on the corresponding trial from the main experiment. As formally captured by the GI, this adjustment gives us a pure estimate of guidance by features distinctive to the two targets in a multiple target search task. Note also that our method of creating these control conditions from the corresponding 2P trials means that the other distractors in the search scene would be identical between the 2P and control conditions on a trial-by-trial basis. This is important, as these distractors would be expected to influence the degree of guidance to the car and bear lures.

3.2.3. Results and Discussion

Errors were less than 2% in all conditions and were excluded from all analyses. Table 2 shows the RT data (Experiment 1). As in the case of the experiments reported in

Table 2
Mean manual RTs in Experiment 1 and 2. Numbers in parentheses indicate standard error.

		Teddy Bear	Car
Experiment 1	1P	1369.2 (57.1)	1433.5 (80.1)
	2P	1592.6 (84.8)	1821.2 (114.6)
Experiment 2	1P	1310.6 (42.1)	1298.3 (33.6)
	2P	1399.2 (45.6)	1503.3 (59.2)

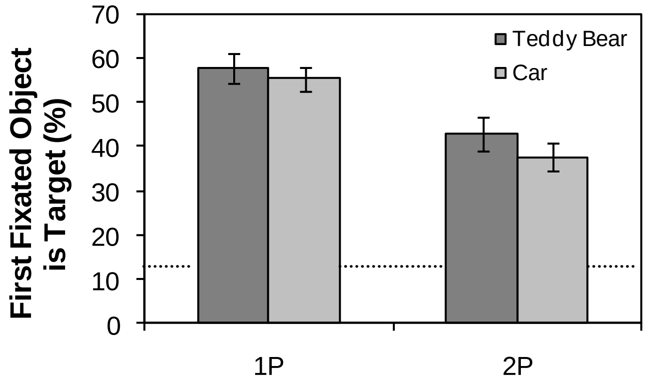

Figure 7. Percentage of trials where target was the first fixated object. Error bars indicate standard error. Dotted line indicates chance level of guidance.

Chapter 2, these manual RT data include the time needed to find the small + or × character and to make this discrimination judgment, in addition to the delays involved in actually searching for the target. However, for completeness, we analyzed these data and found that RTs in the 1P condition were shorter than those in the 2P condition, $F(1,23)=34.9$, $p<.001$, and RTs for teddy bear targets were shorter than those for car targets, $F(1,23)=11.1$, $p<.01$. The interaction between target type and number of targets was also significant, $F(1,23)=5.1$, $p<.05$.

To quantify guidance in our multiple target search task we again analyzed the percentage of trials in which the target was the first object fixated after search display onset. These results are shown in Figure 7. The observed levels of guidance were far more pronounced than the 11.1% level expected by chance, $ts(23)\geq 8.2$, $ps<.001$. Guidance was also better in the 1P condition than in the 2P condition, $F(1,23)=76.1$, $p<.001$, and was stronger for teddy bears targets than car targets, $F(1,23)=4.9$, $p<.05$. However, we found no significant interaction between target type and number of targets,

$F(1,23)=.59$, $p>.4$. Given this absence of an interaction, we therefore collapsed across target type in our following calculation and analysis of the guidance index.

We calculated the GI using the equation: $GI = \{(G_{2P} - c) – (G_L / 2)\} / (G_{1P} - c)$. The terms G_{2P} and G_{1P} were obtained by averaging the percentages of immediate target fixations over the two target categories, giving us 2P and 1P guidance values of 40.3% and 56.5%, respectively (from Figure 7). To calculate the value for G_L, we first obtained the percentages of immediate fixations to the critical lures in the control experiments, which were 9.4% for cars (when subjects were searching for teddy bears) and 7.1% for teddy bears (when subjects were searching for cars). The guidance differences between each critical distractor and the averaged guidance level from all the non-critical distractors were -1.9 and -4.5 for cars and bears, respectively, and the estimated guidance due to visual similarity (common features) between the two target categories (G_L) was -3.2, obtained by averaging these two differences. The final term from the GI equation is a constant, c, which is determined by the set size and indicates the percentage of immediate target fixations expected by chance. This was 11.1% in the current experiment.

Plugging in these values from the main and control experiments, we obtained a GI of .65 (Figure 8). Recall that a GI > .5 would represent true multiple object search guidance; that is, guidance due to the representation in visual WM of features distinct to both of the targets. A GI < .5 would indicate a guiding representation based on the features common to both of the previewed targets. Our obtained value of .65 was significantly greater than .5, $t(23)=3.4$, $p<.01$, which was the theoretical upper bound on the use of only one target to guide two-target search.

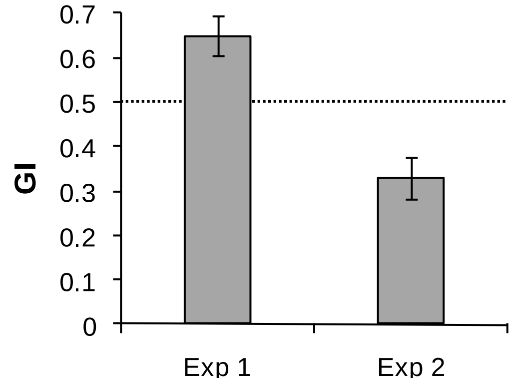

Figure 8. GI for experiment 1 and 2. Dotted line indicates a theoretical GI under the assumption that people use only one target to guide two target search.

Our analysis so far indicates that features distinctive to the two targets in a multiple target search task are extracted and used to guide search, with the obtained GI being our evidence for the existence of these guiding features. To the extent that the GI is a valid estimate of these distinctive target features, we would expect that subjects who tended to use distinctive features should show better guidance during multiple target search than subjects who did not (those who either picked one of the two targets to guide their search, or who used common features), and subjects who relied more on distinctive features to guide their search would show stronger guidance than subjects who relied less on distinctive features for search guidance. To assess both of these patterns, we correlated the GI for each subject with their level of two-target guidance obtained from the main search experiment (Figure 9). This analysis revealed a very strong relationship between a subject's GI and the percentage of immediate fixations that they make on a target during two-target search, $r^2 = .72$, $t(23)=7.4$, $p<.001$. Subjects who are able to extract and use distinctive features from both targets to guide their search are far more successful in looking directly to the target in a multiple target search task than subjects

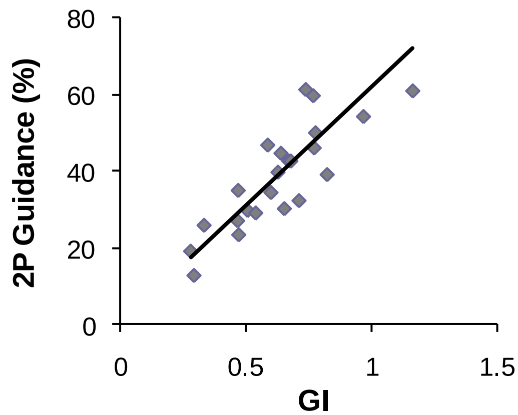

Figure 9. Individual subject's guidance level in 2P condition as a function of GI.

who may be simply picking one of the two targets to guide their search. Moreover, guidance improves with the recruitment of these distinctive features, with the GI accounting for up to 72% of guidance variability in our task. However, while validating the GI as a measure of useful features, this analysis also highlights the fact that not all of our subjects used distinctive features to guide their search. Based on a confidence interval of .95, 8 out of the 24 subjects from our multiple target experiment exhibited GI values falling outside the lower bound of this interval, indicating the potential use of common features to guide search. This variability begs the question of what factors might contribute to the differential use of common and distinctive features in multiple target search guidance, and it is this question that we attempt to answer in the following experiment.

3.3. Experiment 2

3.3.1. Introduction

In experiment 1 we determined that distinctive features can be extracted from multiple targets and used to guide search, but we also found that subjects varied considerably in their reliance on these features. In the present experiment we ask how the practice of using distinctive features depends on the similarity between the two targets in a multiple target search task. The classes of target images used in experiment 1 were highly dissimilar. In general, cars look nothing like teddy bears, and perhaps this high degree of visual dissimilarity is what caused the majority of our subjects to use distinctive features to guide their search. Would distinctive features still be used if the two targets were more similar, as would be the case if they both came from the same category?

Our working hypothesis is that the types of features used in the representation of multiple search targets may depend on the feature similarity between these targets, and the degree to which they differ from the distractors. When two targets are very dissimilar, it may not be possible to find features that are common to both objects while still being different from random distractors. In the absence of these common features, search may be forced to use features that are (relatively) unique to each target, even though this increases the potential of feature mismatch to the distractors. Feature selection may be very different for objects from the same category. Although two teddy bears may differ with respect to their specific color or shape, with some brown and tall and others yellowish and thin, these differences are likely to be small compared to the differences between teddy bears and cars, or between either target class and random objects. A similar argument might be made for cars. These smaller within-category differences may make it easier to find features that are common to the two targets (e.g., both have fur, or both are metallic), yet still relatively distinct from the distractor set. This representation in terms of common features may be desirable, as both targets could be coded with minimal feature variability; as the feature variability increases, so does the opportunity for feature mismatch. We will test this hypothesis using the same GI measure from experiment 1. If the selection of common features is the preferred method of representing multiple targets, then the search for multiple targets from the same category should produce a GI well below the .5 level indicating pure single target search. However, if multiple search targets are represented in terms of discriminative features, irrespective of the feature overlap between the targets, then we may again find a GI above .5, as we did in experiment 1.

3.3.2. Method

Participants

Forty students from Stony Brook University participated in this experiment for course credit. Twenty four of these students participated in the main experiment, and sixteen participated in the control experiments. All had normal or corrected to normal visual acuity, by self report, and were naive with regard to the experimental hypotheses.

Stimuli and Apparatus

The stimuli and apparatus were identical to those used in experiment 1.

Procedure

<u>Main experiment</u>. The 1P-teddy bear and 1P-car conditions were the same as those described for experiment 1. The 2P condition, however, differed in one important respect; the two targets presented at preview were either both teddy bears or both cars. On 2P-teddy bear trials, two different teddy bears were presented as previews, with one of the two teddy bears appearing as the target in the search display. On 2P-car trials, two different cars were presented as previews, with the target in the search display now being

a car. The two 1P conditions and the two 2P conditions (bear or car trials) were blocked (32 trials per block), and block order was counterbalanced across subjects.

Control experiment. The control experiments used a similar procedure, and followed the same logic, as was described for experiment 1. As before, a new group of subjects participated in two control experiments; half of these subjects searched for a bear (teddy bear control), and the other half searched for a car (car control). The one salient difference from the previous description is that the 32 target-absent search displays now contained a categorically similar lure as the previewed target. With respect to the corresponding 2P-teddy bear trials from the main experiment, one of these bears would be presented at preview, and the other bear would be presented as a lure in the target-absent search display. The same contingencies existed for the car trials. From these control experiments, we will be able to determine the degree of guidance to a lure when a subject is searching for a categorically related specific target.

3.3.3. Results and Discussion

Errors were less than 3.5 % in all the conditions and were excluded from all analyses. Manual RTs in this task, provided in Table 2, were again of marginal interest in this task due to their inclusion of non-search delays related to finding the +/× character and making this discrimination judgment. More informative are percentages of the first fixations on the target (Figure 10). As in experiment 1, Figure 10 shows above chance levels of guidance in all the conditions, $ts(23)>=12.4$, $ps<.001$. As expected we also found stronger guidance in the 1P condition than in the 2P condition, $F(1,23)=32.5$, $p<.001$, but guidance did not differ between the teddy bear and car conditions,

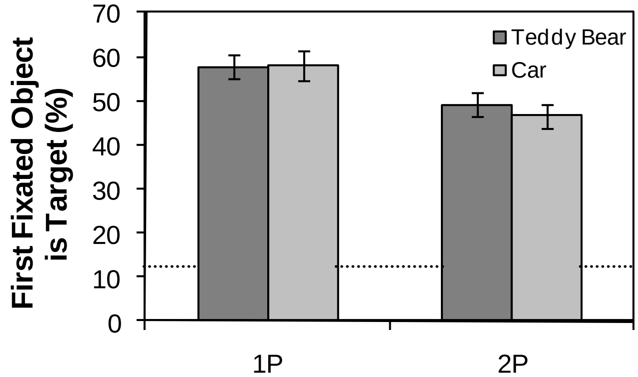

Figure 10. Percentage of trials where target was the first fixated object. Error bars indicate standard error. Dotted line indicates chance level of guidance.

$F(1,23)=.12$, $p>.7$. We therefore collapsed across teddy bear and car trials in the GI calculation.

We calculated the GI using the same equation that we used for experiment 1: GI = $\{(G_{2P} - c) - (G_L / 2)\} / (G_{1P} - c)$. The terms G_{2P} and G_{1P}, again obtained by averaging the percentages of immediate target fixations over the two target categories, yielded 2P and 1P guidance values of 47.82% and 57.34%, respectively (from Figure 10). To calculate G_L, we obtained the percentages of immediate fixations to the critical lures in the control experiments, which were 61.0% for teddy bears (when subjects were searching for a different teddy bear) and 32.9% for cars (when subjects were searching for a different car). The guidance differences between each critical distractor and the averaged guidance level from all the non-critical distractors were 56.2 and 24.5 for bears and cars, respectively, and the estimated guidance due to common features between the two target categories (G_L) was 40.3, obtained by averaging these two differences. Chance guidance, c, was again 11.1%.

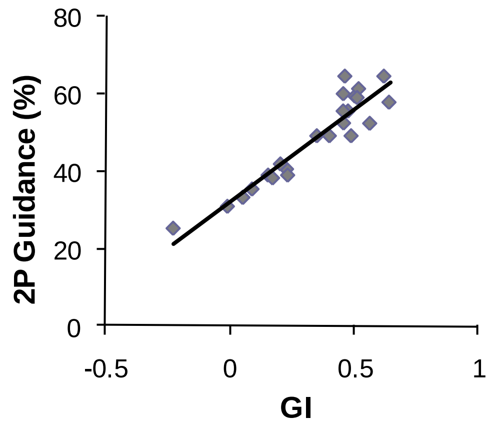

Figure 11. Individual subject's guidance level in the 2P condition as a function of GI.

Inserting these values into the equation produces a GI of .33, an estimate that is significantly lower than the .5 value expected if subjects were just using one of the two targets to guide their search (Figure 8), $t(23)=-3.7$, $p<.01$. This swing in the direction of the GI, from significantly above .5 to significantly below, indicates the predicted shift in the use of common features when multiple search targets come from the same category. This movement towards the representation of common features is a reasonable alternative search strategy in this task, as punctuated by the very strong level of guidance found in immediate target fixations. This is intuitive, if one is able to extract the common features of the multiple targets, for example the distinctive pose of the teddy bears or the metallic texture of the cars, such a compact representation should be efficient in guiding search, at least insofar as the distractors are not metallic or shaped like teddy bears.

But does this use of common features to guide multiple target search mean that distinctive features are unimportant in the case of categorically similar targets? Figure 11 answers this question by again plotting the GI as a function of 2P guidance. The GI is a measure of the proportion of 2P guidance that can be attributed to the use of distinctive features; as the GI increases, so does the reliance on these features to guide search. What is clear from this figure is that there is a very strong positive correlation between multiple target guidance and distinctive features, $r^2 = .89$, $t(23)=13.1$, $p<.001$. Although distinctive features may be difficult to find when targets are from the same category, to the extent that they are isolated, their use can dramatically improve search guidance. In a sense, a subject searching for multiple targets from the same category has more representational choices available; they can choose to represent the targets in terms of either common or distinctive features, with perhaps the best guidance resulting from a combination of the two. Indeed, when only common features were used, as represented in Figure 11 by the subject with the negative GI, we see that 2P guidance was relatively poor. This too is intuitive, and perfectly consistent with previous work on categorically guided search (Schmidt & Zelinsky, 2009; Yang, Chen & Zelinsky, 2009; Yang & Zelinsky, 2009). When the GI is very low, the subject may essentially be searching for the target category (the features common to any teddy bear or any car), and not the specific targets themselves. Such a categorical search would be expected to yield low, but still above chance levels of guidance.

3.4. General Discussion

The present series of experiments investigated the target representation underlying multiple target search, and how this representation changes with the similarity of the previewed targets. Our first goal was to provide independent and converging evidence in support of our earlier claim that feature mismatch, and not a "pick one target and discard the other" strategy, was responsible for our observed load effects. We did this by developing a measure, the guidance index, which shows the proportion of two-target search that can be accounted for by one-target search, after appropriate correction for chance and feature similarity between the two targets. This measure provides a theoretical upper bound, a GI of .5, on the level of guidance that would be expected under a "pick one and discard the other" multiple target search strategy. We found a mean GI of .65, significantly higher than this upper bound, and can therefore conclude that subjects were indeed basing their guiding representation on information from both of the previewed targets, not just one. When considered together with the findings from Chapter 2, we now have considerable evidence suggesting that feature mismatch is responsible for the worse search guidance found for multiple target search. Given that only one target appears in the search display, guiding search based on two targets will inevitably create a feature mismatch that will degrade guidance relative to single target search.

Using the GI, our second goal was to determine whether this multiple target representation was dominated by common features or distinctive features. Given that either type of representation might result in decreased guidance due to a mismatch between the features in WM and the actual target features from the search display, distinguishing between these two representational schemes is essential to the understanding of multiple target search. We were able to make this distinction using the same GI measure. By using control experiments to determine the level of guidance to a target lure, and subtracting this level from the GI, the GI provides a pure estimate of multiple target search guidance arising from the use of distinctive features. GIs in experiment 1 were found to be consistently greater than .5, suggesting not only that both previewed targets were represented and used to guidance search, but that this guidance was accomplished using features that were largely distinct to both targets. We also found a very strong linear relationship between GI and two-target search guidance; the greater the GI the better the guidance. To the extent that subjects are successful in extracting features that are unique to the two targets, they are better able to guide their search to these targets. This benefit linked to the use of distinctive features makes intuitive sense, and is consistent with current machine learning methods that also rely on discriminative features for object detection (e.g., Zhang, Yang, Samaras, & Zelinsky, 2006). Given that the two targets in experiment 1 were random objects, the features that are unique to these targets (i.e., discriminative) will tend also to be those features that best enable these targets to be discriminated from random object distractors, thereby improving search guidance.

In experiment 2 we asked whether the reliance on distinctive features to guide two-target search depends on the similarity between the two targets. We did this by having subjects search for objects from the same target category, either two teddy bears or two cars. If search under these conditions remained dominated by distinctive features, we would have expected to find GIs greater than .5, as we did in experiment 1. However, if this within-category multiple target search caused features common to the two targets to be used for guidance, then GIs should have been consistently lower than .5. We found a decisive shift in GIs to values less than .5, suggesting guidance from common features. However, we again also found a strong linear relationship between GI and two-target search guidance; guidance improved with larger GIs. This relationship suggests that the recruitment of distinctive features helps guidance even when the two search targets come from the same category.

If search guidance always benefits from the use of distinctive features, why then did our experiment 2 subjects often elect to use common features to guide their search? There are at least two possible explanations. One reason may be simply that it is difficult to find features that distinguish between categorically similar targets. Whereas finding a set of guiding features unique to a given teddy bear and car may be straightforward (the car may be rectilinear, the teddy bear may be curvy), there are likely far fewer features that are unique to two teddy bears or two cars, potentially resulting in an insufficient number of features to guide search. Another reason may be that we are used to relying on common features for search guidance, as exemplified each time we look for a categorically defined target. In previous work we demonstrated that categorical search is guided (Yang & Zelinsky, 2009), meaning that subjects were able to use the features common to a target category to preferentially fixate never before seen exemplars. It may be that some of our subjects treated the experiment 2 multiple target task as categorical search, deciding to search for the *category* of teddy bears or cars despite being shown the two specific exemplars (the previews) from which the actual target would be selected. The problem with this strategy, however, is that although categorical search yields an above chance level of guidance, this guidance is not as strong as when one knows the specific appearance of a target (Schmidt & Zelinsky, 2009; Vickery, King, & Jiang, 2005; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004; Yang & Zelinsky, 2009). It may therefore be possible to perform the experiment 2 task using common features, but this strategy, by ignoring information about the targets' actual appearance, would tend not to be as efficient as one relying on distinctive features.

One final question might be raised here regarding multiple target search and its relationship to common and distinctive features; if common features are used to guide the search for two targets, is this multiple target search or single target search? In one sense, common features are, by definition, indeed representative of both targets. In another sense, the fact that they are common means that these features might be aligned with one target or the other, but not necessarily both. Following this reasoning, a common feature search, because it is representationally equivalent to the search for a subset of a single target's features, might therefore not be considered a multiple target search task at all. Consistent with this suggestion, Menneer et al. (2009) found that subjects searching for a knife and a gun, both of which appeared blue in an x-ray image, showed a level of accuracy comparable to that of a search for the knife target alone. This suggests that subjects may have redefined this gun-knife multiple target task as a single search task—one in which the target was simply a blue thing.

We believe that this potential for target redefinition, stemming largely from the extraction and use of common features, is one reason for the discrepant findings in the literature regarding the efficiency of multiple target search. Future work will attempt to clarify target redefinition behavior by better specifying the conditions under which common feature search might be, and might not be, useful. For example, a common feature search might be beneficial in that it compresses the target representation, thereby reducing the feature variability that might result in extraneous matches to distractors; however, it might also be detrimental to the extent that this representation becomes too compressed, thereby degrading the task into a form of categorical search. We hope to exploit methods from the machine learning community to help find this delicate balance, and to estimate the optimal mixture of common and distinctive features in a multiple target representation. We will also attempt to specify whether search guidance depends on the guiding target representation consisting of features from a single object (as in the case of a redefined target using common features) versus a true multiple target search task in which the guiding features are aligned with different targets. Does the number of objects over which these features are spread affect search guidance, and if so, does a common feature search behave more like a single target search or a multiple target search? Addressing this question in the context of real-world objects will be another important direction for future work.

4. Conclusion

This study ascertained how the representation of multiple targets in visual WM affects the efficiency and guidance of visual search. Using images of real-world objects, the first part of this study asked the most basic question regarding the relationship between visual search and WM; does searching for multiple targets degrade search efficiency? We found clear evidence for such a load effect; search time increased with the number of targets. We then ask *why* this load effect appeared. Two possibilities were considered; one appealing to a capacity limitation on WM encoding, and the other appealing to feature mismatch between the multiple targets in WM and the single target in the search display. We used a retro cue paradigm to tease apart the effects of these two factors, and found that both played a part in lengthening the time needed for subjects to first fixation the target (RTT). However, the percentage of initial fixations on targets, a conservative measurement of search guidance, decreased only with an increase of guiding features. We therefore conclude that feature mismatch is primarily responsible for the decreased guidance observed to search targets when there is a WM load. To the extent that capacity limitations on WM contribute to the load effect, this contribution seems limited largely to the decision stage of search and not actual search guidance. In a final experiment we showed that subjects hold and use all the designated targets to guide their search; they don't just pick one and search for it.

The second part of this study sought to obtain converging evidence for the previous conclusions, as well as to determine whether multiple targets are represented in terms of common features or distinctive features. We satisfied both of these goals by analyzing the chance-corrected and feature-overlap-corrected proportion of two-target to one-target guidance, a measure that we referred to as the guidance index. Not only did this analysis support our previous claims that multiple search targets are being represented in visual WM (subjects aren't simply picking one or the other to guide their search), it also suggested the nature of these representations; subjects searching for teddy bear and car targets overwhelmingly chose to represent these target classes using distinctive features. However, in a follow-up experiment we showed that this representation depended on the categorical similarity of the targets; when the targets were from the same category (two teddy bears or two cars), subjects now choose to represent these targets using common features. Still, despite this similarity-dependent shift in representation, we found that search guidance improves to the extent that subjects code multiple targets using distinctive features. These results, combined with those from the first series of experiments, support our claim that the search for multiple targets degrades guidance by introducing too many guiding features; the target features in WM that don't match the target in the search display weaken the guidance signal and ultimately reduce search efficiency.

Our finding that load effects impair search guidance due to mismatching features is new, and the suggestion and confirmation of this alternative explanation is the primary contribution of this study. However, some may be equally surprised that we found little evidence for a WM capacity limitation degrading search guidance for multiple targets. This finding is broadly consistent with the findings that the units of WM that determine its capacity are features rather than objects (e.g., Alvarez & Cavanagh, 2004; Eng, Chen, & Jiang, 2005); it may simply be the case that the number of features selected in our multiple target conditions were still within the limits of this feature WM capacity.

However, this observation in itself is interesting in that it means that the features used to code up to four objects for search, the maximum number of previews used in this study, is still within this capacity limit. This further implies that this feature capacity may be somewhat larger in a search task than what had been found using a change detection task, or that the number of features needed to code search targets is smaller than what had been assumed. It may also be the case that all (or most) of the target features are initially represented (as assumed by search theories), but that only a subset of critical features are selected and maintained in WM (Ko & Seiffert, 2009; Woodman & Vogel, 2008), and ultimately used to guide search. Answering these questions will be important directions for future work. Another difference between our study and the WM literature is that subjects in WM studies are typically instructed to selectively focus on a specific feature of an object (e.g., Ko & Seiffert, 2009; Woodman & Vogel, 2008), while subjects in the current study were not. Our results therefore extend previous work by showing that the features of an object can be selectively encoded in visual WM without explicit effort to do so. One intriguing possibility is that it is this intentionality that is responsible for the previously observed evidence for WM limitations in the context of non-search tasks. Future work will also explore this possibility.

References

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.

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.

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

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* (4), 753-763.

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

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

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

Downing, P. E., & Dodds, C. M. (2004). Competition in visual working memory for control of search. *Visual Cognition*, *11*, 689-703.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96* (3), 433-458.

D'Zmura, M. (1991). Color in visual search. *Vision Research, 31*, 951-966.

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, *12*, 1127-1133.

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

Han, S.-H., & Kim, M.-S. (2004). Visual search does not remain efficient when executive working memory is working. *Psychological Science, 15* (9), 623-628.

Hollingworth, A., & Luck, S. J. (2009). The role of visual working memory in the control of gaze during visual search. *Attention, Perception, & Psychophysics, 71* (4), 936-949.

Houtkamp, R., & Roelfsema, P. R. (2006). The effect of items in working memory on the deployment of attention and the eyes during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *32* (2), 423-442.

Kaplan, I. T., & Carvellas, T. (1965). Scanning for multiple targets. *Perceptual and Motor Skills*, *21*, 239-243.

Ko, P. C., & Seiffert, A. E. (2009). Updating objects in visual short-term memory is feature selective. *Memory & Cognition*, *37* (6), 909-923.

Kveraga, K., Boucher, L., & Hughes, H. C. (2002). Saccades operate in violation of Hick's law. *Experimental Brain Research*, *146*, 307–314.

Lawrence, B. M., St. John, A., Abrams, R. A., & Snyder, L. H. (2008). An anti-Hick's effect in monkey and human saccade reaction times. *Journal of Vision*, *8* (3), 1-7.

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

Ludwig, C. J. H., & Gilchrist, I. D. (2002). Stimulus-driven and goal-driven control over visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, *28* (4), 902-912.

Luria, S. M., & Strauss, M. S. (1975). Eye movement during search for coded and uncoded targets. *Perception & Psychophysics*, *17*, 303-308.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9* (11), 1-13.

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: Effect of practice and target similarity. *Journal of Experimental psychology: Applied*, *15*, 125-139.

Metlay, W., Sokoloff, M., & Kaplan, I. T. (1970). Visual search for multiple targets. *Journal of Experimental Psychology*, *85*, 148-150.

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

Neisser, U., Novick, R., & Lazar, R. (1963). Searching for ten targets simultaneously. *Perceptual and Motor Skills*, *17*, 955-961.

Oh, S.-H., & Kim, M.-S. (2004). The role of spatial working memory in visual search efficiency. *Psychonomic Bulletin and Review*, *11*, 275-281.

Quinlan, P. T., & Humphreys, G. W. (1987). Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints on feature and conjunction searches. *Perception & Psychophysics, 41*, 455- 472.

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology*, *62* (10), 1904-1914.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1-66.

Scialfa, C. T., & Joffe, K. M. (1998). Response times and eye movements in feature and conjunction search as a function of target eccentricity. *Perception & Psychophysics*, *2*3, 244-262.

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* (2), 248-478.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, *40A*, 201-237.

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

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

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.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception & Psychophysics*, 56, 495-500.

Wheeler, M. E., & Treisman, A. M. (2002). Biding in short-tem visual memory. *Journal of Experimental Psychology: General*, *131*, 48-64.

Williams, D. E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks: the effects of feature discriminability and saccadic amplitude. *Psychonomic Bulletin and Review*, *8*, 476-488.

Williams, L. G. (1967). The effects of target specification on objects fixated during visual search. *Acta Psychologica*, *27*, 355-360.

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

Wolfe, J. M., Horowitz, T., 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.

Woodman, G. F., & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. *Psychonomic Bulletin and Review*, *11*, 269-274.

Woodman, G. F., & Luck, S. T. (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. T., & Schall, J. D. (2007). The role of working memory representations in the control of attention. *Cerebral Cortex*, *17*, 118-124.

Woodman, G. F., & Vogel, E. K. (2005). Fractioning working memory: Consolidation and maintenance are independent processes. *Psychological Science*, *16* (2), 106-113.

Woodman, G. F., & Vogel, E. K. (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin & Review*,*15* (1), 223-229.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science*, *12*, 219-224.

Xu, Y. (2009). Distinctive neural mechanisms supporting visual object individuation and identification. *Journal of Cognitive Neuroscience*, *21* (3), 511-518.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91-95.

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

Yang, H., Chen, X., & Zelinsky, G. J. (2009). A new look at novelty effects: Guiding search away from old distractors. *Attention*, *Perception*, *& Psychophysics*, *71* (3), 554-564.

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

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233-235.

Zhang, W., Yang, H., Samaras, D., & Zelinsky, G. J. (2006). A computational model of eye movements during object class detection. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in neural information processing systems* (Vol. 18, pp. 1609-1616). Cambridge, MA: MIT Press.