CORRESPONDENCE COMPUTATIONS IN VISUAL COGNITION

by

Gi-Yeul Bae

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

March, 2014

Abstract

15 behavioral experiments were conducted to investigate the role of object correspondence computations in visual cognition. Correspondence computations refer, here, to algorithms that identify relationships between objects in temporally separate encounters.

In Experiment 1-5, I hypothesized that tracking failures occur because of correspondence failures during close encounters of targets and nontargets. To test this idea, I provided observers with different surface feature information to nontargets whenever they approached within 4° of a target (Experiment 1). This manipulation significantly improved performance by alleviating correspondence challenges. Two control experiments showed that this color change benefit is not merely due to target recovery (Experiment 2 and 4). A follow-up experiment measured the distance at which objects correspondence becomes challenging (Experiment 3). And an additional experiment demonstrated that the overall frequency of target-nontarget close encounters predict human performance (Experiment 5).

Experiment 6-10 explored the role of object correspondence in the context of spatial working memory. Experiment 6 supplied evidence of object correspondences in a typical spatial working memory task through a trial specific analysis. In addition, a model that implements correspondence algorithms successfully predicted human performance without assuming any independent memory-related limits. Experiments 7 and 8 employed a preview display that indirectly provided information about memory location to be tested. This manipulation improved SWM performance dramatically (e.g. performance with 8 objects were comparable to 2 objects). A control experiment showed that the improved performance is not due to mere reactivation of memory representations

(Experiment 9). Additional experiment showed that object colors do not support correspondence computations in this context.

Experiment 11-15 employed integral features to prevent correspondence failures in a visual working memory task. I reasoned that integral features can be used to solve correspondence problems by preventing confusions between objects. Experiment 11 and 12 independently identified integral features using perceptual sorting experiments. When these features were used in change judgment tasks, working memory with two objects produced performance as precise as with one (Experiments 13-15).

Taken together, these results suggest that object correspondence play a crucial role in the constraints typically observed in visual cognition.

Readers:

Jonathan Flombaum (advisor)

Lisa Feigenson

Marina Bedny

Colin Wilson

Steven Gross

Acknowledgements

Although they may not know what I am doing in the opposite side of the earth, my mom and dad always told me that they are so proud of me. Like any other parents do, they give me endless supports, love and encouragements no matter what I do. Their job is simply to love me, not getting the outcome of what I did. Thank you very much for your endless support. I cannot say enough how much I love you. I'd like to thank Se-Yeul, my younger brother. He is a part of my life as a family member and also he is my best friend. I used to think that he is my 'younger' brother whom I need to take care of, but that's simply not true any more. It's opposite. He took care of me and always gave me positive feelings. Thank my brother. Seojin and Yumin, my lovely son and daughter, your existence is the most powerful source of energy for me. I love you so much and thank for growing up really well.

There are so many Hopkins people who helped me to achieve Ph.D. I must begin with, of course, my advisor, Dr. Jonathan Flombaum. Jon, you are the best mentor ever. We were really a good team, aren't we? Like I said all the time, you are not merely an academic mentor. You are a good mentor as well as a good friend. In fact, I cannot imagine how my grad school life would look like if you were not my advisor. Thank you so much! Dr. Justin Halberda, your sense of humor always made me think of Vision lab as a family. Thank you very much for creating nice and cozy atmosphere. Dr. Colin Wilson, I am so lucky that I had an opportunity to work with you. Everything worked out nicely after I started to work with you. That would be highly unlikely if it was a mere coincidence. And, doubtlessly, I've learned a lot from you. I would like to thank Dr.

Howard Egeth. He is a great person in general but here I want to focus on his role as a senior faculty in the department. Simply, he is a kind of library. His ideas always inspired me to think about problems with a different perspective. Dr. Steve Yantis, I still remember what you told me after my first research presentation in the department — "This is a great idea. You can do a lot of things with that!" Thank you very much for this encouragement! There are numerous people who helped me directly/indirectly in the Psychological and Brain Sciences department including faculty members, my colleagues in the vision lab and administrative staffs. Thank you all! It is really bittersweet to leave this great department.

Lastly, I would like to mention a quote that raised me up whenever I had difficult moments in my journey toward Ph.D.

Whoever wants to be born, must destroy a world.

— *Demian* (Hesse, H., 1919)

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Correspondence computations

Though we rarely really think about it, a central challenge faced by the visual system is to identify correspondences between images and their parts. A simple example has to do with the fact that the visual system takes inputs from two eyes to construct an internal representation (Julez, 1971). Because the inputs from two eyes are not identical, the visual system must identify correspondences between them. In fact, this is necessary for using those inputs to infer 3D depth relationships.

A more intuitive kind of correspondence that the visual system faces regularly involves occlusion, and what is often called 'amodal completion'. Consider, for example, the image in Figure 1A. It can be interpreted in at least three different ways. What should the visual system make of the relationship between the black lines? It could construe them as parts of a single object, attached to the box in the middle (i.e., The object has a box and two lines on each side, Figure 1.01.A), or as three independent objects (i.e., two short lines and a box, Figure 1.01.C). But what you probably see is an inferred correspondence between the lines — that they correspond to a single object, partially occluded by the box (Figure 1.01.B). Note that these really are very different interpretations. They concern the number of objects in the scene.

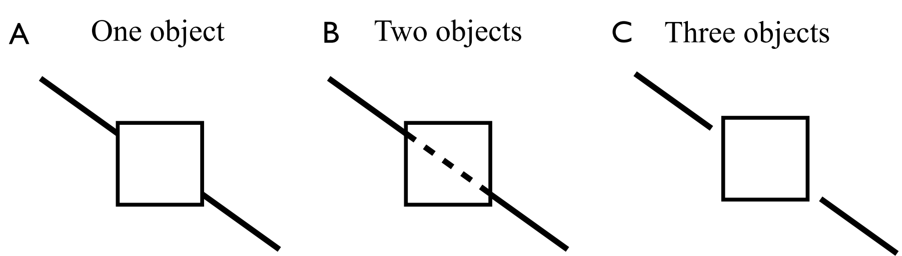

Figure 1.01. Amodal completion. Despite the fact that image A can be interpreted in multiple ways, the visual system tends to take the 'Two objects' option more frequently.

There is another very ubiquitous kind of correspondence that the visual system makes use of frequently, but that is discussed less often than amodal completion. Object correspondences need to be addressed in the temporal domain. The visual system needs to infer relationships of object identity between temporally separated visual encounters. Imagine a scene where multiple objects are moving around haphazardly (e.g. birds in the sky). As they move, they change position and size, and they occasionally disappear from view becoming occluded by others. How does the visual system determine which object went where? Is that particular object I see now the same object I saw a moment ago? If it is, which property of the object leads to the perceptual judgment of the sameness? If not, where was the object a moment ago? Answering these questions is crucial for the visual system because object identity and its relationship to others are the basis over which further visual processes operate, including guiding attention and remembering properties of objects (Flombaum, Scholl, & Santos, 2009; Scholl & Flombaum, 2010; Scholl, 2001).

In the remainder of this introduction, I will discuss two psychological phenomena, the tunnel effect and the apparent motion, which have been studied extensively with respect to correspondence computations. The purpose of the discussion will be to explain how these examples reveal general rules underlying correspondence computations, particularly with respect to the use of spatiotemporal information versus surface featural information when determining object correspondences. Following this discussion, I turn to other areas in visual cognition that are also likely to involve correspondence problems, and the same kinds of rules, but where the role of correspondences has not been addressed explicitly or empirically. The empirical portion of the dissertation will then report a set of experiments that reveal the role of correspondence computation in these areas.

Correspondence computations in The Tunnel Effect

Object correspondence has been studied extensively in the context of the tunnel effect. Imagine a scene where an object moves toward an occluder, it becomes occluded, and then another object emerges from the occluder (Figure 1.02.A). Clearly, interpreting this simple visual event requires a correspondence computation. Is the object after the occlusion the same individual as the one before the occlusion? How we see this visual event depends on the answer to this question. What should one think or see if the two objects look featurally different from another on either side of the occluder? What influence should contextual events have? How important should the time and place of disocclusion be?

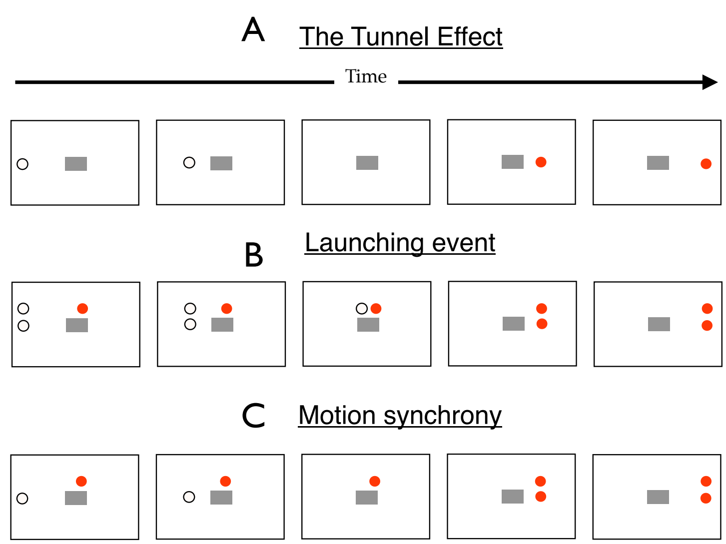

Figure 1.02. The Tunnel effect. A. Naïve observers tend to see the event as a single object continuously passing through the occlusion despite the featural change. Namely, the visual system identifies the post-occlusion object as the same object as the pre-occlusion object. B. The correspondence decision about the tunnel event can be changed. When a launching context is added nearby the tunnel event, observers are more likely to see the pre-occlusion object as causing motion of the post-occlusion object – a phenomenon called amodal causal capture. C. Observers see amodal causal capture even when the context event does not involve a launching event. Motion synchrony is sufficient to alter the perception of the tunnel event.

Studies have shown that the event is more likely to be seen as a single object persisting behind the occluder whenever the emerging object appears at the right time and space, and this is so even when the two objects look very different from one another (e.g. different colors or shapes) — a phenomenon called the tunnel effect. If a temporal delay intervenes between the pre- and post- occlusion views and/or the emerging object appears at an unexpected position, the tunnel effect is greatly weakened. These phenomena suggest that correspondence computations involved in processing motion through occlusion depend greatly on spatiotemporal display properties, a principle called spatiotemporal priority (Flombaum, Scholl, &Santos, 2009). For current purposes, then, the point is that we may expect correspondence computations in many contexts to rely heavily on spatiotemporal properties, and less so on other features.

What about the local context? When the basic tunnel event is accompanied by temporally synchronized events that imply either a one object or a two object interpretation, these can strongly bias perception. Indeed, demonstrating this was the focus of a project preceding this dissertation (Bae & Flombaum, 2010). For instance, when a clearly causal, two-object event is presented nearby the basic tunnel event, naive observers are more likely to see a two-object (and causal) event in the tunnel (Figure 1.02.B). Similarly, an extra object that starts to move together with the emerging object in the tunnel event also produces strong impression of "launching" in the tunnel event (Figure 1.02.C). The effects of context on the tunnel effect reveal another general heuristic that we may expect to find elsewhere: that the relevant computations seek correspondence decisions that appear the most likely given the surrounding context.

Correspondence computations in Apparent Motion

Object correspondences also play an important role in the perception of unoccluded object motion, especially what is known as apparent motion. In fact, motion is the most obvious case in which object correspondence should play an important role because tracking object motion cannot occur without moment to moment correspondence.

A basic apparent motion display involves the sequential presentation of an object that appears in one location, disappears, and then reappears in another location (Figure 1.03.A). In this display, we readily perceive motion in spite of the fact that there is no visible motion trajectory. That is, the visual system tends to identify that the object in the second frame is the same individual as the one in the first frame.

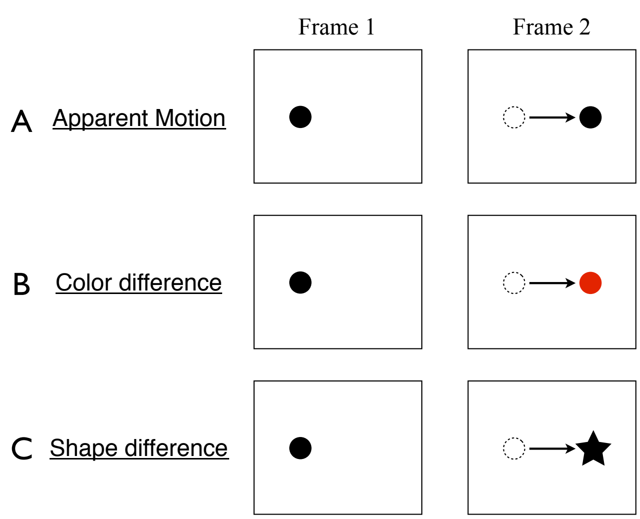

Figure 1.03. Apparent motion. Observers perceive motion of a single object (A) even when the two objects look different (B, C).

As in the tunnel effect, spatiotemporal properties primarily determine apparent motion percepts. Apparent motion is thus seen even when the objects in the two flashes look very different (e.g., different shape and color, Figure 1.03B and C). But if the temporal delay between the two frames is long enough (>1s), observers are more likely to see the object in the second frame as a different individual from the one in the first frame.

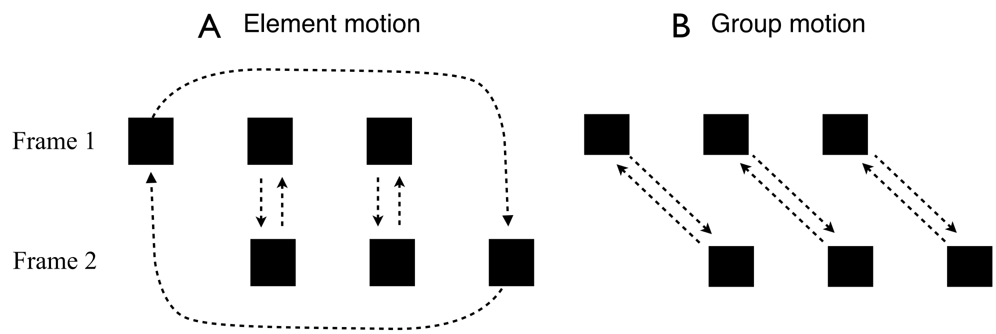

Figure 1.04. Ternus illusion. The identical visual event can be seen differently depending on a temporal delay between the two frames. A. Observers see element motion when the delay is short enough (e.g., <100ms). B. When the delay is longer, however, they tend to see a group motion.

Apparent motion thus reinforces the expectation that spatiotemporal rules may generally control correspondence computations. And they can help us to understand those rules in more detail. This can be seen in displays with more than one object in each of the relevant frames, displays in which several different kinds of correspondences are possible. Imagine, for example, something called a Ternus display, where three horizontally arranged objects appear in the first frame and another set of three objects appears in the second frame (Figure 1.04). How we see the display depends on the timing between the two frames. We tend to see that the left-most object in the first frame jumps over the two objects on the left in the second frame (called element motion, Figure 1.04.A) when the temporal delay between the two frames is short enough. In contrast, we are more likely to see that all three objects in the first frame move together to new positions (called group motion, Figure 1.04.B), when the temporal gap is long enough. Crucially, although two different kinds of interpretation are possible, we do not see both simultaneously, or partially. Instead, we perceive the events in one way or the other, even though the visual system is clearly capable of producing the alternative option. The general lesson, then is that correspondence computations provide just one of a set of mutually exclusive interpretations on any given occasion.

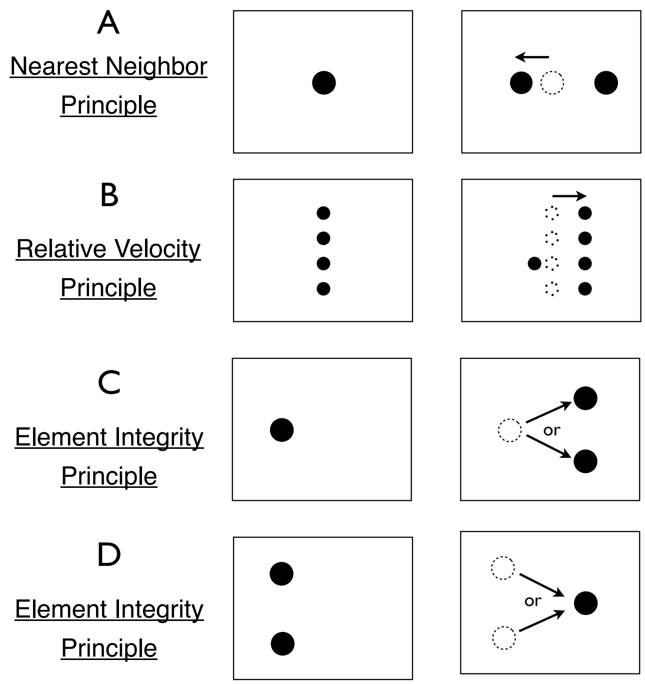

Figure 1.05. Principles used in Apparent motion perception (Dawson. 1991). Left column of the figure represents the first frame and the dotted objects in the right column of the figure represents the positions of objects in the first frame.

Studies of apparent motion with multiple objects can also help us to elaborate the principles that likely guide correspondence computations in general. Figure 1.05 depicts a variety of displays that have been used in this way, resulting in taxonomy of rules that seem to guide correspondence computations (Dawson, 1991). For example, the visual system tends to find correspondences by minimizing the distance any given object is though to have moved — a rule sometimes called ‘the nearest neighbor principle’. But in some instances (e.g., 1.05.B) the visual system seeks to minimize the implied relative differences in how objects moved under the assumption that all the objects in a display should tend to move in similar ways, sometimes called ‘the relative velocity principle.’ We also know that the visual system tends to avoid interpretations in which an object moves to two new locations at once, and in which multiple objects converge into one location, revealing what has been called, the element integrity principle, Figure 1.05.C and D). We may expect that all these principles generalize to other areas of visual cognition where object correspondences play a role.

<u>Object correspondence in other areas of visual cognition</u>

So far, we have seen two cases where object correspondences play an important role. In both cases, spatiotemporal factors are taken as the primary inputs for correspondence judgments. This gestures towards the expectation that object correspondences should play an important role in any scenes involving spatiotemporal dynamics. In fact, there are several important areas in visual cognition where object correspondences are likely to play an important role, but where they have been relatively understudied.

One of these is Multiple Object Tracking (MOT). In typical MOT, observers attempt to mentally track a set of continuously moving objects (Figure 1.06.A). So, obviously, MOT requires identifying object correspondences sequentially. One important aspect of MOT is that one should also discriminate targets from nontargets. That is, one should know which objects are targets and which are not, otherwise the observer would track a nontarget while thinking it is a target.

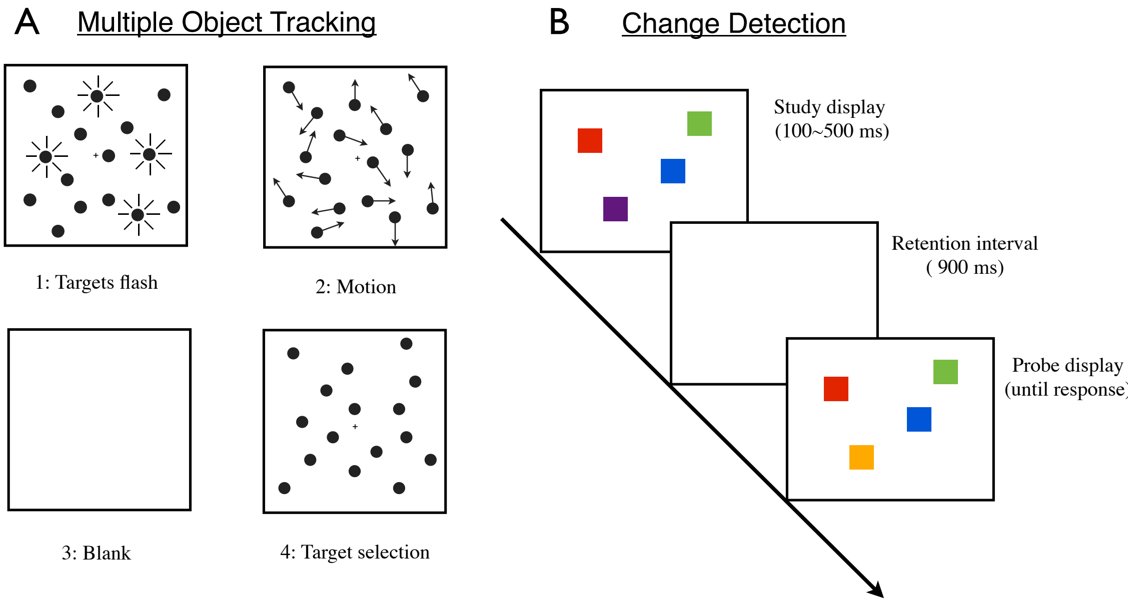

Figure 1.06. Two common tasks employed in visual cognition research. A. Multiple Object Tracking. Observes attempt to mentally track a subset of pseudo-randomly moving objects. At test, they are asked to identify the target objects. B. Change Detection. Observers attempt to remember a set of objects presented in the study display and make a decision if any object in the probe display is different from its corresponding memory object.

Similarly, the tasks utilized in many Visual Working Memory (VWM) studies also involve potential challenges of object correspondence. For instance, in a typical change detection task where observers are asked to detect any differences between objects in memory and those in view, one needs to know which memory object should be compared to which object in view, making some kind of correspondence decision necessary (Figure 1.06.B). Of course, failures to find a corresponding memory object may lead to a wrong decision, which would influence performance, and should thus play a role in how we characterize the limits of performance in this context.

But very little research has explored how object correspondences are computed in these contexts. I speculate that the reason for this is that researchers have assumed that object correspondences can be computed without difficulty. But, as we've seen in the tunnel effect and apparent motion the relevant computations may be context-dependent and non-trivial. Determinations need to be made about the kind of information to use in making correspondence judgments, and simple manipulations can change the judgments made, demonstrating the tentative nature of any correspondence decisions.

Whatever the reasons are for the neglect of object correspondences in MOT and VWM, they have been under investigated. In what follows, I will review current studies on MOT and VWM in more details, identifying important places where correspondence computations might play a role. I will then, introduce experiments that can provide empirical evidence for the role of correspondence computations in MOT and VWM.

1.2.Correspondence computations in Multiple Object Tracking

In a typical MOT trial, several identical objects are presented on a screen and move randomly for a certain period of time (Figure 6A). Observers attempt to mentally track a subset of objects and then to identify them at the end of the motion sequence. Most of the time, observers fail to track more than three or four objects at once, and sometimes with fast movement speed, they fail to track even as few as one object (Alvarez & Franconeri, 2007; Holcombe & Chen, 2012). This limit in tracking is typically thought to reflect the capacity of attention and/or other mental resources such as working memory.

Current understanding of the limits in MOT

When Pylyshyn and Storm (1988) first developed the MOT task, they were interested in the kind of mechanisms that allows the visual system to represent object features in the world. According to Pylyshyn's FINST (Fingers of INSTantiation) theory, the visual system employs a FINST mechanism that can be used to connect internal representations and the world. Specifically, FINSTs play a role as a pointer that is stuck to object representations, enabling the visual system to maintain the representation over the course of tracking. Later, Pylyshyn stipulated that the reason for the limited MOT performance is because only limited quantities of FINSTs are available in the visual system. This conception about the limited number of FINSTs became a theoretical basis of a recent theory on MOT, known as a fixed resource theory. According to this theory, the visual system can process only a limited quantity of individual objects at once because it has only small number of discrete slots, buffers or FINST-like representations. These capacity limits become evident when an observer attempts to track more than the theoretical upper bound, which is typically 3 or 4 objects (Cowan, 2000; Drew & Vogel, 2008; Luck & Vogel, 1997).

But in recent research, it appears that tracking performance is not fixed at a particular level, but rather varies depending on numerous display factors. Studies have found that MOT becomes harder when objects move fast (Alvarez & Franconeri, 2007; Holcombe & Chen, 2012; Liu et al., 2005; Pylyshyn & Storm, 1998), when spacing between objects is tight (Alvarez & Franconeri, 2007; Franconeri, Jonathan, & Scimeca, 2010), and when tracking duration is longer (Franconeri, Jonathan, & Scimeca, 2010). Sometimes observers fail to track even one object moving very fast (Holcombe & Chen, 2012), while they can track up to 7 or 8 slowly moving objects (Alvarez & Franconeri, 2007). These results suggest that tracking capacity is a dynamically determined limit that depends on several display parameters.

Why and how do these display factors limit MOT performance? One possible reason is that attention has limited spatial resolution. Attention cannot always select two objects successfully when they are close enough. Instead, there must be limits in how finely attention can discriminate between the two nearby objects. In order to test the effect of limited attentional resolution on MOT, Intriligator and Cavanagh (2001) manipulated the visible density of the tracking display by simply varying the viewing distance. The idea was that the identical tracking display can be seen as crowded or uncrowded when the viewing distance is longer or shorter. Without any direct manipulation of display parameters, Intriligator and Cavanagh could test how tracking performance varies as a function of the crowdedness of the display. Results showed that tracking performance was poorer when the viewing distance was farther, and thus more crowded, suggesting attentional resolution limits MOT performance.

In a similar vein, a recent study proposed that, among many display parameters, inter-item proximity maybe the only limiting factor in MOT (Franconeri, Jonathan, & Scimeca, 2010). According to Franconeri and his colleagues, many display factors limit MOT performance by introducing more frequent moments where objects are in close proximity such that attention becomes more likely to fail to discriminate them. For instance, object speed may not independently limit MOT, but rather increase the frequency of inter-item proximity that challenges the resolution of spatial attention. In order to test this possibility, they carefully manipulated object speed and tracking duration. In one condition, objects moved fast but for short duration. And, in another, they moved slowly for longer duration. On average, the frequency of inter-item interaction was equivalent between the two conditions. Results showed that MOT performance for faster speed conditions was not distinguishable from slower speed conditions, suggesting speed limits MOT through its impact on inter-item proximity.

This work on the effects of inter-item proximity has been interpreted in a number of ways, often with respect to the resolution of attention, and with a particular focus on how limited mental resources may reduce the resolution of attention as tracking increases. But there is perhaps a simpler explanation. In particular, tracking errors are by definition correspondence errors, responses to targets instead of nontargets. But perhaps they are also correspondence errors in a local sense —confusions between targets and nontargets that take place at the moments that those objects incidentally approach one another? In other words, perhaps MOT errors can reveal the limitations of correspondence computations from moment to moment?

Some research supports this point of view indirectly. Sears and Pylyshyn (2000), for example, combined MOT with probe detection in order to test which objects observers track from moment to moment. Specifically, in their task, observers tracked the pre-designated targets and, at the same time, they detected probes that appeared on either targets or nontargets. The rationale behind this task is that observers should detect probes better when the probe appears on object they are tracking. Thus, if they track a nontarget by mistake, then they should detect better when a probe appears on that nontarget. Overall, probe detection was better when probes appeared on targets than non-targets. More importantly, when they increased the number of nontargets while maintaining the same number of targets, observers more frequently detected probes that appeared on nontargets. Further, on perfect tracking trials, probes that appeared on nontargets were not detected, suggesting that on those trials nontargets were not tracked. In addition, O’Hearn, Landau, and Hoffman (2005) found that the distance between incorrectly identified objects and missed targets were shorter than successfully rejected nontargets and missed targets, suggesting targets and nontargets are more confusable when they are in close distance.

<u>Finding direct evidence for object correspondence in MOT</u>

Converging evidence that inter-object distance constrains MOT directly suggests that object correspondence does play an important role. This leads to more specific predictions about how and when MOT becomes challenging. First, we know from previous studies that increasing inter-object distance improves MOT performance (Alvarez & Franconeri, 2007). However, it is unclear whether the improved performance is mainly due to the minimum distance between target and non-targets or to the distance between every object in the display. Namely, it is possible that the improved performance is due to the infrequent interactions among objects in a global level rather than local close encounters of targets and non-targets. If target-nontarget confusions are the main causes of tracking errors, then we can predict improved MOT performance when the object confusion is prevented. If this is the case, then it would provide direct evidence that the correspondence failures are the important causes of errors in MOT.

In relation to this, we can also test the degree to which the correspondence computation plays a role in MOT. How much errors are accounted for by the failure of correspondence computations? Several predictions are possible. The most extreme prediction would be the case where the correspondence computations are the only causes of errors. If this is the case, then preventing correspondence failures should produce near perfect performance. On the other hand, if correspondence computations play a minor role, then preventing correspondence error would not improve performance that much. It is also possible that correspondence computations may interact with other display factors such as object speed. For example, it plays an important role when objects move slowly but not when they move rapidly, or vice versa.

Furthermore, at which distance do objects become confusable from one another? In a static display with crowded objects, numerous studies have shown that an object becomes difficult to recognize when they are placed with other objects nearby, and studies have described the critical spacing between objects that allows us to recognize objects (for review, see Whitney & Levi, 2011). However, in a display with motion, like an MOT display, no study has supplied a direct measurement of what the critical spacing is. What has been reported is that increasing object spacing produces better MOT performance (Alvarez & Franconeri, 2007), but, again, manipulating distance between all the object in a tracking display cannot provide a direct answer for when targets become confusable with nontargets.

In sum, I have discussed current understanding of MOT performance with an eye toward finding the role of correspondence computations. Converging evidence suggests that correspondence computation play an important role in MOT, but no direct evidence has been provided. And important questions concerning the role of correspondence errors, the distances and speeds at which they are most likely have not been answered. Chapter 2 will therefore present experimental work done to address some of these outstanding questions.

1.3.Correspondence computations in Visual Working Memory

Visual working memory (VWM) allows us to temporarily maintain visual information that can be used for future behavior such as guiding attention, programming saccades and limb movements, tracking objects, perceiving motion, and supplying inputs to a longer-term memory. Despite its importance in numerous human behaviors, VWM seems to be severely limited. Numerous studies have reported that VWM can store only three or four visual objects at once and that the quality of VWM representation declines with memory loads even for two objects compared to one. Why is VWM so limited?

Current understandings of the limits in VWM

A seminal study by Luck and Vogel (1997) proposed that VWM is a capacity limited system where only a limited number of objects can be stored —a magic number of 4 (Cowan, 2000), similarly to the notion of 4 FINSTs, as described in the context of MOT. Some neuronal evidence supports the idea by showing that the relevant electrophysiological signal (contra-lateral delayed activity) reaches a plateau at the theoretical upper bound of object limits (Vogel & Machizawa, 2004).

But other studies have suggested that a more continuous kind of resource limits the quality of VWM representations, and thus the ability to use those representations effectively as memory load increases. Using a signal detection approach, Wilken & Ma (2004) have shown that measured working memory noise increases continuously with memory loads. Later, Bays & Husain (2008) investigated how limited working memory resources are allocated to objects. They found that resource allocation is highly flexible such that it can be divided among objects equally (e.g. 1/n), or it can be allocated to a particular object more or less, depending on cognitive factors such as eye movements and covert attention. These results became the basis of a theory of working memory called a flexible resource theory. A recent computational model also supports the idea that VWM is limited by a mental resource that can be flexibly used on a trial-by-trial basis (van den Berg at al., 2012).

Other studies have approached the limits of VWM by investigating the contents of VWM representations. What kind of visual information is stored in VWM? An influential theory on attentional mechanism suggested that VWM stores the individual feature values of objects. Complete object representation is only possible when separate attentional processes integrate features together (Treisman & Gelade, 1980). But other studies have suggested that VWM may hold a complete object representation rather than individual feature values (Luck and Vogel, 1997).

And some of have proposed hybrid sorts of accounts, including the possibility that memory represents individual objects, but that the number depends on the information load that some object category demands (Alvarez & Cavanagh, 2004). For example a simple colored square requires less information than more complex random polygons. In this study, they first measured the information load using a visual search. Different search times were obtained for different classes of stimuli, allowing the authors to operationalize information in terms of search time. They then used those stimuli classes in a working memory task and they found a linear relationship between measured working memory capacity and search time— stimuli that produced longer search times produced lower capacity. With a careful inspection of the data they also found evidence for a discrete capacity limit of about four or five objects. These results suggest that VWM is not solely limited by the quantity of contents in VWM but also limited by the amount of information that each object requires (but see, Awh, Barton, & Vogel, 2007; Barton, Ester, & Awh, 2009).

Some studies have also suggested that the contents of VWM interact with one another, creating hierarchically organized representations. It is known that the visual system can compute the statistical properties of multiple objects in a scene, such as the mean size of objects (Ariely, 2001; Chong & Tresiman, 2003), mean location (Alvarez & Oliva, 2008) and mean orientation (Parkes et al., 2001). The idea is that when observers encounter a scene with multiple objects, they can easily extract a higher order representation of the scene by computing the statistical properties of multiple objects. Likewise, when observers attempt to memorize objects in a working memory task, they tend to encode statistical information in VWM. Researchers tested this hypothesis in a size working memory task where observers remember a set of objects in different colors and estimated a size of a single object (Brady & Alvarez, 2011). They found that individual estimates were biased toward the mean of objects presented on the test display (for extensive review, see Brady, Konkle, & Alvarez, 2011).

Similarly, studies on spatial working memory have shown that memory for the position of an object is highly dependent on others. Jiang, Olson, and Chun (2000) had participants remember positions of a set of object in a study display and judge if a position of a target object in a test display has changed compared to the position in the test display. They had three critical conditions. In one condition, the target object appeared alone in the test display. In the second condition, nontarget objects in the test display were in identical locations to those in the study display. In the third condition, nontarget objects were displaced in random location. Results showed that performance declines severely when nontarget objects were displaced and improved when they were in the same positions. Jiang and colleague have concluded that the positional representation of an object is highly dependent on others such that changing position of nontarget objects vastly affects the memory of the target object.

The limited nature of VWM has also been studied with respect to its temporal dynamics. How is memory consolidated? It has been known that the process that consolidates the initial perceptual representations into durable working memory representation is slow and attention demanding (Jolicœur & Dell'Acqua, 1998; Potter. 1976). The consolidation process has been extensively studied in the context of the Attentional Blink task, in which observers are asked to detect two targets presented in a RSVP stream (Rapid Serial Visual Presentation) and then report them at the end of the stream (Broadbent & Broadbent, 1987). The task is likely to involve working memory components since observers have to remember what the targets were. The crucial finding in this task is that accuracy for reporting the second target (T2) is lower, or sometimes totally forgotten, when T2 was presented after a few hundred milliseconds following the T1 presentation (Chun & Potter, Zuvic, Visser, & Di Lollo, 2000). It should be noted that studies reported that T2 was fully perceived although it was not reported at the end of stream (Jolicœur & Dell'Acqua, 2000; Shapiro, Driver, Ward, & Sorensen, 1997; Vogel, Luck, & Shapiro, 1998), suggesting failures to report T2 were not due to the fact that observers were not able to see it. These results suggest that perceptual representation requires time to be consolidated otherwise it cannot be maintained in VWM.

The neglected role of object correspondence in VWM

These numerous approaches to understanding the structure of VWM resources, the contents and organization of VWM representations, and the consolidation and decay processes involved have rarely considered a role for correspondence computations. This is surprising because most VWM tasks share a common task structure. That is, at the start of a trial, observers attempt to encode object information in a test display which is presented for a short duration (e.g. 100ms~1sec.). They maintain the encoded information for another short duration (e.g. ~900 ms), and then make different kinds of decisions about an object or objects presented in a test display, depending on the task demand (Figure 1.06). For example, in a common change detection task, observers report if objects in the study display are the same or different from those in the test display (Luck & Vogel, 1997). In a change direction judgment, observers report how the feature of a probe object has been changed (Bays&Husain, 2008). Lastly, in a delayed estimation task, observers estimate the feature value of a probed object by responding on a continuous response scale (e.g. colorwheel) (Fougnie & Alvarez, 2011; Zhang & Luck, 2008; van den Berg et al., 2012).

Crucial feature of these working memory tasks is that they require decisions about which memory objects correspond to which test objects. And it is almost always assumed that these correspondences are simply known by observers. Imagine a spatial working memory task where observers remember positions of a set of items in a study display and then judge if one of the items has changed its horizontal position in the test display (Bays & Husain, 2008). Obviously, the position comparisons require finding a corresponding memory item to the probe. But if correspondence computations are not perfect —which we should not expect them to be— correspondence failures can lead to mistakes. Thus, it is crucial to understand why and how correspondence computation becomes difficult in order to properly evaluate the limits and structure of working memory, generally.

In fact, some recent studies have attempted to account for correspondence computations in a model by taking non-target based responses into account (Bays, Catalo, & Husain, 2009; Emrich & Ferber, 2011). They found that incorporating non-target based responses in a model improves the model fits for a color delayed estimation task and changes important parameter estimates. But the model did not propose how correspondence computations are addressed, and it did not identify any systematicity in the correspondence errors. Thus it makes the point that correspondence errors likely contribute to performance without properly measuring them or characterizing them. Chapters 3 and 4 supply evidence for the role of correspondence errors in visual working memory, as well as formal modeling that implements correspondence computations based on the general correspondence principles associated with the tunnel effect and apparent motion.

Chapter 2. Correspondence Computations in Multiple Object Tracking

(The results and experiments in this chapter were previously reported in Bae & Flombaum (2012). Close encounters of the distracting kind: Identifying the cause of visual tracking errors. Attention, Perception, & Psychophysics, 74, 703-715.)

A Typical MOT trial involves a subset of identically looking objects moving around a display. Observers attempt to mentally track a subset of objects and then to identify them at the end of the motion sequence. On average, observers fail to track more than three or four objects at once, and sometimes with fast movement speed, they fail to track even as few as one object (Alvarez & Franconeri, 2007; Holcombe & Chen, 2012). This limit in tracking is typically thought to reflect the capacity of attention and/or other mental resources such as working memory.

A prevailing theory proposed that the limits in MOT are due to the limited number of slots, buffers, or FINST-like representations. These capacity limits become evident when an observer attempts to track more than the theoretical upper bound, which is typically 3 or 4 objects (Cowan, 2000; Drew & Vogel, 2008; Luck & Vogel, 1997).

However, recent research has shown that tracking performance is not fixed at a particular level, but rather varies depending on numerous display factors. Studies have found that MOT becomes harder when objects move fast (Alvarez & Franconeri, 2007; Holcombe & Chen, 2012; Liu et al., 2005; Pylyshyn & Storm, 1998), when spacing between objects is tight (Alvarez & Franconeri, 2007; Franconeri, Jonathan, & Scimeca, 2010), and when tracking duration is longer (Franconeri, Jonathan, & Scimeca, 2010). These results suggest that tracking capacity is a dynamically determined limit that depends on several display parameters.

So why and how do these display factors limit MOT performance? One possible reason is that attention has limited spatial resolution. Because attention cannot always select two objects successfully when they are close enough, there must be limits in how finely attention can discriminate between the two nearby objects. This possibility was tested Intriligator and Cavanagh (2001). Cleverly, they manipulated the visible density of the tracking display by simply varying the viewing distance. This way, the same tracking display can be seen as crowded or uncrowded when the viewing distance is longer or shorter. Results showed that tracking performance was poorer when the viewing distance was farther, and thus more crowded, suggesting that attentional resolution limits MOT performance.

In a similar vein, a recent study proposed that, among many display parameters, inter-item proximity may be the primary limiting factor in MOT (Franconeri, Jonathan, & Scimeca, 2010). According to Franconeri and his colleagues, many display factors limit MOT by introducing more frequent moments where objects are in close distance. Similarly, it has been suggested that object speed does not independently limit MOT, but it does by increasing the frequency of nearby item interactions.

Numerous related works on the effects of inter-item proximity has been interpreted in a number of ways with a particular focus on how limited mental resources may reduce the resolution of attention as tracking load increases. But there is perhaps a simpler explanation. In particular, tracking errors are by definition correspondence errors, responses to targets instead of nontargets. But perhaps they are also correspondence errors in a local sense —confusions between targets and nontargets that take place at the moments that those objects incidentally approach one another? In other words, perhaps MOT errors can reveal the limitations of correspondence computations from moment to moment?

Some research supports this point of view indirectly. Sears and Pylyshyn (2000), for example, combined MOT with probe detection in order to test which objects observers track from moment to moment. Specifically, in their task, observers tracked the pre-designated targets and, at the same time, they detected probes that appeared on either targets or nontargets. The rationale behind this task is that observers should detect probes better when the probe appears on object they are tracking. Thus, if they track a nontarget by mistake, then they should detect better when a probe appears on that nontarget. Results showed that this is the case. When they increased the number of nontargets while maintaining the same number of targets, observers more frequently detected probes that appeared on nontargets. Further, on perfect tracking trials, probes that appeared on nontargets were not detected, suggesting that on those trials nontargets were not tracked. In addition, O'Hearn, Landau, and Hoffman (2005) found that the distance between incorrectly identified objects and missed targets were shorter than successfully rejected nontargets and missed targets, suggesting targets and nontargets are more confusable when they are close in distance.

But no direct evidence has been provided to support the presence of local correspondence errors, and no work has directly investigated the moments when they should take place. Broadly, the literature on tracking has at once suggested that object spacing has a strong influence on MOT performance and that it interacts in intuitive ways with speed, tracking load, and trial duration, while at the same time focusing on resource-based and/neural explanations for tracking limits in general, and the effects of spacing in particular. Is it possible that much of the effect of spacing and related variables arises because correspondence computations are imperfect, and error prone at short distances? In the experiments reported in this chapter, I sought to investigate the specific moments during which tracking errors emerge, the moments when targets and nontargets are close enough to one another to be confusable.

The experiments broadly relied on the following logic: if failures of object correspondences predominantly constrain performance, then preventing such failure should improve performance. More specifically, I sought to improve tracking performance via an intervention that would make targets and nontargets less confusable only at those critical moments, as opposed to a global manipulation. In Experiment 1, I attempted to prevent confusions by supplying distinguishing surface properties to nontargets during the moments that I thought a confusable close encounter might take place. The logic of the experiment was that when in close proximity, featural selection could help to keep targets and nontargets separate (Egeth, Virzi, & Garbart, 1984).

Experiment 1: Preventing correspondence challenges in MOT

If attentional resolution places the most proximate constraint on object tracking abilities, then errors should take place mainly when targets and non-targets approach close enough to one another to make it possible for a non-target to be selected, and then tracked in lieu of a target. Can the hazard created by such close encounters be reduced — in other words, can the frequency of tracking errors be reduced — by providing participants with an ability to favor target selection during close spatial encounters? It would possible that supplying targets and non-targets with distinguishable surface features might facilitate target selection. To this end, I caused non-targets to change into a distinct random color during any moments that they approached within a critical distance of a target. The critical distance was 4˚, chosen based on pilot testing and an assessment of related literature (Franconeri, Alvarez, & Enns, 2007). I chose to change non-target colors because changing target colors could facilitate recovery of lost targets, as opposed to maintained tracking of a currently selected target (Makovski & Jiang, 2009a and 2009b). Overall, the logic of the experiment was straightforward. During most moments of a trial, targets and non-targets were identical, requiring sustained tracking of targets by updating knowledge of their locations over time. When a non-target approached within a close encounter distance of a target, the relevant non-target took on a random, distinguishable color that was different from all the other identically colored items in the display (Figure 1.01b). I called these ‘Color Change’ trials, and I compared performance in Color Change trials to ‘Standard’ trials wherein targets and non-targets always possessed identical colors. Better performance in Color Change compared to Standard trials would constitute evidence that close encounters cause errors by leading to tracking confusions.

Method

Participants. A group of 11 Johns Hopkins University undergraduates participated in exchange for course credit. Each had normal or corrected-to-normal visual acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Apparatus. The experiment took place in a dimly lit soundproof room. All displays were presented on a Macintosh iMAC computer at a viewing distance of approximately 60 cm such that the display subtended 39.56° X 25.35° of visual angle.

Stimuli and Procedure. At the beginning of each trial, 16 green disks (diameter 1.24˚) were presented on a black background in randomly selected positions. A white fixation cross (0.5˚ x 0.5˚) was present in the center of the display and remained present throughout each trial. Four of the green disks flashed on and off four times over the course of two seconds, identifying them as targets. All the disks then moved haphazardly through the display for six seconds. Motion speed was manipulated across trials such that 1/3 of all trials were 'Fast' (6.69˚/s), 'Medium' (5.43˚/s), and 'Slow' (4.18˚/s). Objects mostly moved in a fixed direction, but on each frame each object had a 2% chance of turning. When an object did turn, the angle was selected randomly from between 1˚ and 359˚. Additionally, if any two objects came with in 2˚ of one another, they repelled, forcing turns so that two objects could never be closer than 2,˚ measured center to center.

Half of all trials were Standard trials, wherein all the moving disks remained green. In contrast, half of all trials were Color Change Trials. In Color Change trials non-targets turned a distinct, random color whenever they approached within 4˚ of a target. These colors were assigned by randomly choosing R, G, B values between 0 and 255. When more than one non-target was within the critical distance of any targets, those non-targets did not take on the same colors. Importantly, when non-targets moved more than 4˚ away from any targets, it changed back to green. In both trial types, all items stopped moving after six seconds, at which point all the items disappeared from the screen for one second to prevent observers from using color information at the last moment of motion. All circles then reappeared, colored in green, and observers used the mouse to select all the items they thought were targets. Selected disks turned yellow to prevent participants from choosing the same item twice. After they selected a total of four items, the true targets turn red, providing feedback. Each observer completed a total of 60 trials. The experiment lasted approximately 25 minutes. Although eye movements were not monitored, observers were instructed to fixate on the white fixation cross.

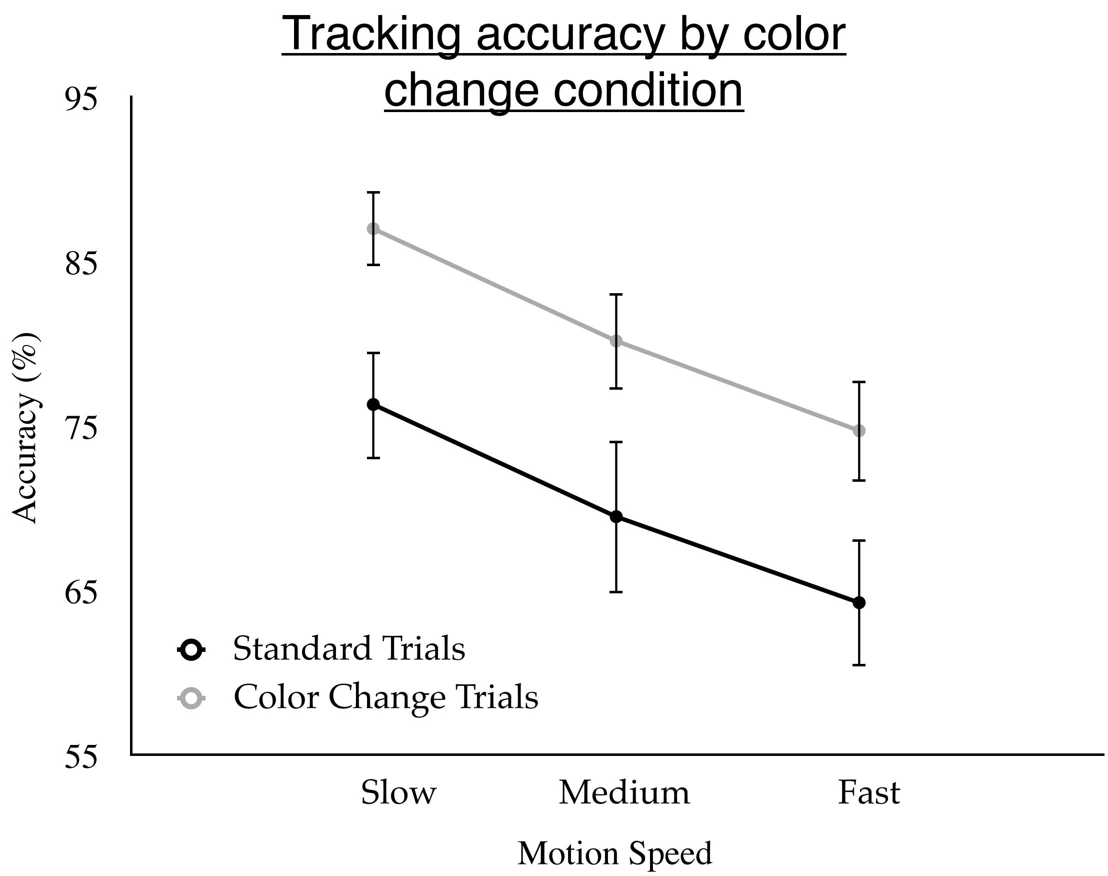

Figure 2.01. Mean tracking accuracy as a function of speed and color change condition in Experiment 1. Error bars indicate ±1 standard error of the mean.

Results and Discussion

Figure 2.01 displays mean tracking accuracy as a function of speed and color change condition. A two-way ANOVA confirmed significant main effects of both speed and color change. Performance decreased as speed increased, $F(2,20) = 14.22$, $p < .001$, and it was improved when non-targets changed color, $F(1,10) = 21.885$, $p = .001$. There was no significant interaction between color change condition and speed, with a constant advantage of about 10% for Color Change trials at all speeds $F(2,20) < 1$.

These results supply clear evidence that tracking errors are caused by close spatial encounters between targets and non-targets. In other words, under standard tracking conditions, a correspondence challenge arises when targets and non-targets come too close to one another: with limited spatial resolution, participants are unable to distinguish targets from non-targets, and they may end up selecting and tracking a non-target by mistake. Changing non-target colors reduced the impact of these interactions by supplying a basis upon which participants could preferentially maintain targets, despite limited spatial resolution.

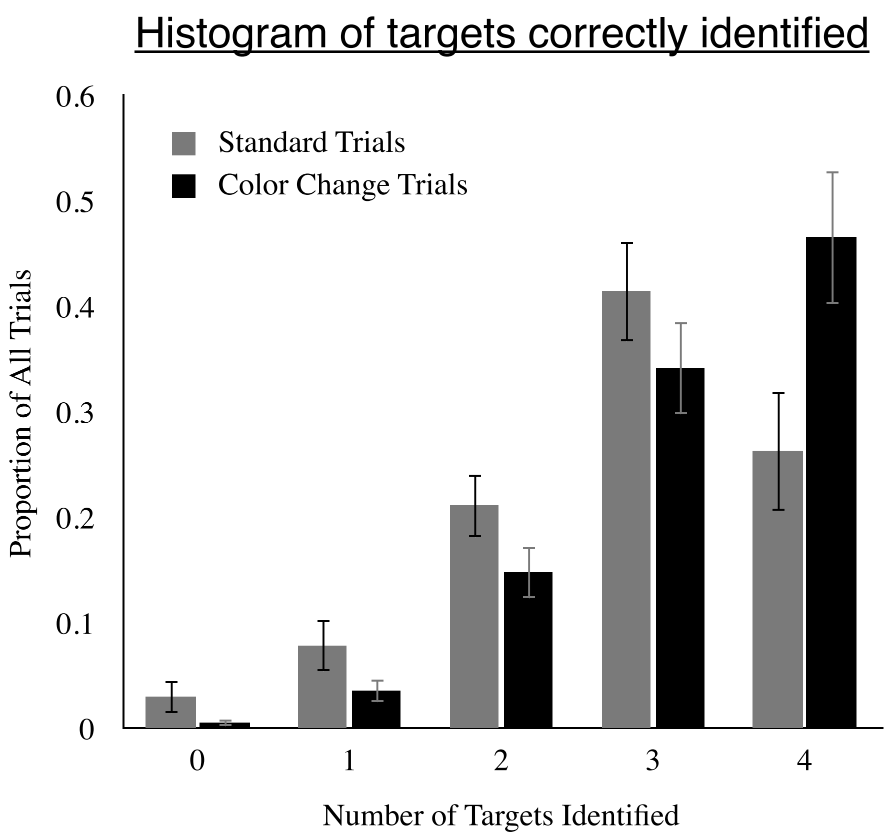

Figure 2.02. Proportion of trials in which one to four targets were correctly identified as a function of color change condition in Experiment 1. Color change trials produced almost twice as many perfect performance as standard trials.

If tracking is most proximately constrained by limited spatial resolution, and my color change manipulation alleviated the challenges that emanate from close spatial encounters, then one may wonder why performance was not perfect in Color Change trials. Figure 2.02 displays the proportion of trials in which participants successfully identified between one and four targets. This is to emphasize the size of the effect induced by my color change manipulation. In Standard trials, participants only tracked perfectly — identifying four targets — in 26% of all trials, whereas in Color Change trials this number nearly doubled to 47%. It is also likely that at least some poorly performed trials reflected general inattention, either at the start of a trial (i.e. when targets were flashing) or for extended durations during a trial. Obviously, errors induced by general inattention would not be eliminated by my color change manipulation, meaning that some erroneous trials to remain no matter how well my color change manipulation prevented item confusions. Finally, color changes do not always prevent confusions during a close encounter, just as not all close-encounters result in confusions. Instead, each of these happens probabilistically (Vul et al., 2009). This perspective is consistent with the general fact of declining performance as a function of speed, since higher speeds should produce more close encounters (Franconeri et al., 2010). Thus the observed advantage of 10% across speeds is a consequence of both more individual close encounters taking place and more individual confusions prevented as speed increases.

Experiment 2: Are color change benefits caused by target recovery?

Experiment 1 evidenced a tracking benefit when non-targets took on unique colors within a 4˚ distance from a target. I argued that this demonstrates how correspondence problems during close spatial encounters typically lead to tracking errors. But previous work might suggest an alternative interpretation, one based on the recovery of already lost targets (Makovski & Jiang, 2009a and 2009b; St. Claire, Huff, & Seiffert, 2010). Perhaps this is all that my color change manipulation achieved? In one previous study, for instance, targets and non-targets each possessed distinguishing surface features as groups (e.g. targets were all green and non-targets were all red; Makovski & Jiang, 2009b), and the benefit of unique surface features was larger when the minimum distance allowed was shorter (Makovski & Jiang, 2009a). Broadly speaking, when these group-defining colors were relatively permanent, participants' performance was significantly better. But, because benefits often did not manifest when these group-defining features changed rapidly and unpredictably, or when they failed to define groups exhaustively, the results suggested that participants could use the surface properties to realize that they were tracking a non-target and then find, or recover, the target(s) that they had lost (based on knowledge of their surface properties). It has also been demonstrated that participants can switch to tracking a new set of items during the course of a trial (Wolfe, Place, & Horowitz, 2007). In the current experiments, one might be concerned that my color change manipulation did not facilitate online tracking mechanisms by disambiguating targets and non-targets. Instead, perhaps participants just figured out that if they had lost a target, they could recover it by looking for a green item next to a uniquely colored item.

To rule out this possibility, I directly investigated how well participants could use changing non-target features to find a missing target. Experiment 1 was different from prior experiments on target recovery because, among other things, only non-targets changed colors, they changed colors at different times from one another, and each instance of a color change involved taking on a new and unique color. Therefore, to use non-target features in my experiment, participants would have had to (a) realize that they had lost a target, and then (b) use the fact that non-targets changed color close to targets in order to infer the identity of the target that they lost. In the current experiment, I investigated the possibility of achieving step (b) — that is the possibility of inferring a missing target's identity — by revealing three targets at the start of the trial, and then instructing participants to both track those three and to find and track the fourth, 'hidden' target. The dynamics of non-target color changes in this experiment were the same as in Experiment 1 such that non-targets changed colors at the appropriate times when they moved close enough to the hidden target (or to any other target). If color change benefits are due solely to target recovery, then participants should be able to use the color changes to find and track the missing target in these Hidden Target trials.

Method

Participants. A new group of 11 Johns Hopkins University undergraduates participated in exchange for course credit. Each had normal or corrected-to-normal visual acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Apparatus, stimuli, and procedure. This experiment was identical to the medium speed condition of Experiment 1, except as follows. I employed three tracking conditions. Standard and Color Change conditions were identical to those employed in Experiment 1. I also included a new, third condition, the 'Hidden Target' condition. This was identical to the Color Change condition in Experiment 1, except that only three targets flashed at the beginning of each trial so that observers did not know which of the remaining 13 items was the fourth target. (Non-targets changed color during motion when they were close enough to this 'hidden' target, as in Experiment 1). The experiment was conducted in two blocks. In the first block, Standard and Color change trials were presented randomly intermixed. The second block included only Hidden Target trials. Participants completed 40 trials plus 10 practice trials in the first block, and 20 trials plus five practice trials in the second block. The total duration of this experiment was around 30 minutes.

Data analysis. I derived a formula for computing the effective number of items tracked (ENIT) on each trial. Scholl, Pylyshyn, and Feldman (2001) previously derived a formula for trials with equal numbers of targets and non-targets (see also, Barker, Allen & McGeorge, 2010). I adjusted their formula as reflected below to account for the fact that in this experiment there were more non-targets than targets:

$$p = \frac{m}{T} + \left(1 - \frac{m}{T}\right) \cdot \frac{T-m}{N-m}$$

In this formula p is the proportion correct on each trial, m is the effective number of items tracked, T is the target load, and N is the total number of items. It was assumed that tracking performance depends on two terms, the measured proportion of correctly tracked targets among the total number of targets, and the expected proportion of correct guesses among remaining objects. Because of the unequal numbers of targets and non-targets in my experiment, I adjusted this second term. I solved for m on each trial, thus calculating ENIT on a trial-by-trial basis for each subject.

Results and Discussion

Figure 2.03 displays mean tracking accuracy and ENIT as a function of tracking condition. A one-way ANOVA confirmed a significant main effect of tracking condition, $F(2,20) = 25$, $p < .001$. Planned contrasts showed that tracking performance for the Color

Change condition was reliably better than for the Standard condition, F(1,10) = 12.83, p < .01, replicating the color change benefit observed in Experiment 1. Most importantly, performance in the Hidden Target condition was reliably worse than performance in both the Color Change condition, F(1,10) = 33.08, p < .01, and crucially, worse than in the Standard condition, F(1,10) = 19.33, p < .01.

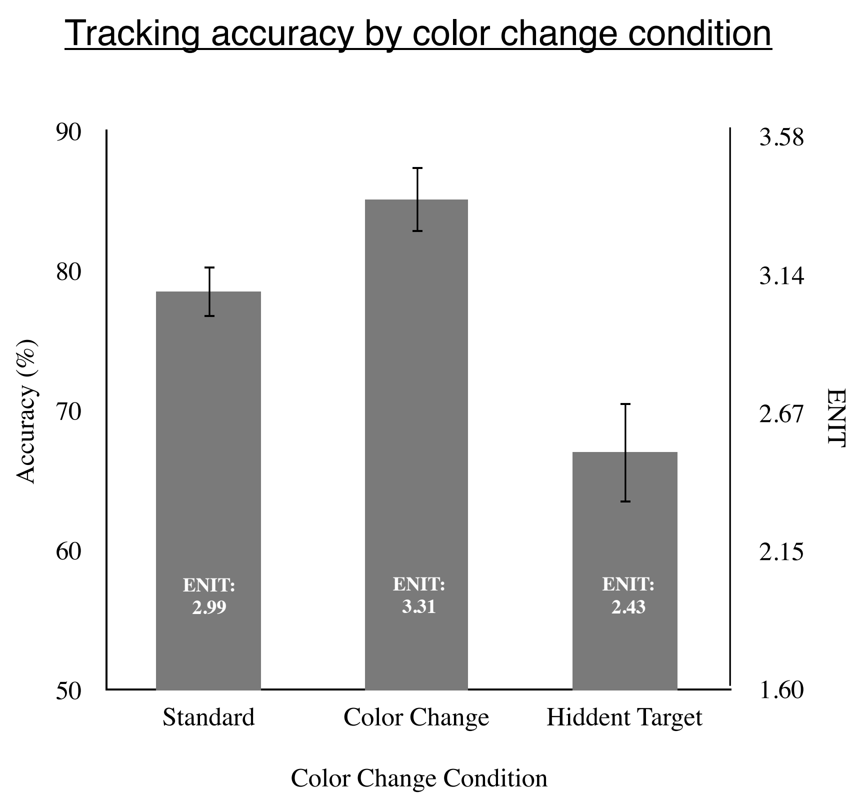

Figure 2.03. Mean tracking accuracy and Effective Number of Items Tracked (ENIT) as a function of condition in Experiment 2. Error bars indicate ±1 standard error of mean accuracy. Standard trials and Color Change trials were identical to Experiment 1, while Hidden Target trials where those in which only three targets (out of four targets) were revealed at the start of a trial. Color Changes took place in both Color Change trials and Hidden Target trials.

These results demonstrate directly the difficulty of using non-target color changes to infer the identity of a target while also tracking other objects. It appears to be the case that explicitly engaging in target recovery mechanisms deteriorates tracking performance. Participants tracked below ENIT of 3, on average, in Hidden Target trials. Given their performance in the Standard trials there is no reason to think that they could not track three items successfully under the motion dynamics and non-target loads employed. Accordingly, the failure to do so in the Hidden Target trials suggests that explicitly attempting target recovery is not possible while successfully tracking three items. But the ability to do so is at the core of a target recovery interpretation of Color Change trial results. Therefore, these results suggest that color change benefits are not likely due to target recovery.

Experiment 3: At what distance do objects become confusable?

This experiment explored the impact of my color change manipulation as a function of inter-item distance. To this end, I systematically varied the distance at which color changes took place. I reasoned that the effect of color change by distance should depend on the underlying resolution of spatial attention.

Method

Participants. A separate group of 11 Johns Hopkins University undergraduates participated in exchange for course credit. Each had normal or corrected to normal visual acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Stimuli, Apparatus & Procedure. This experiment was identical to Experiment 1, with two exceptions: (1) items always moved at a speed of 5.43˚/s (previously called, 'Medium.'), (2) Only 1/6 of all trials were 'Standard.' All others included color changes that took place when targets and non-targets were within 2.5˚, 3.0˚, 3.5˚, 4.0˚, or 4.5˚ of one another. There were an equal number of trials using each of these distances. Each participant completed a total of 60 trials.

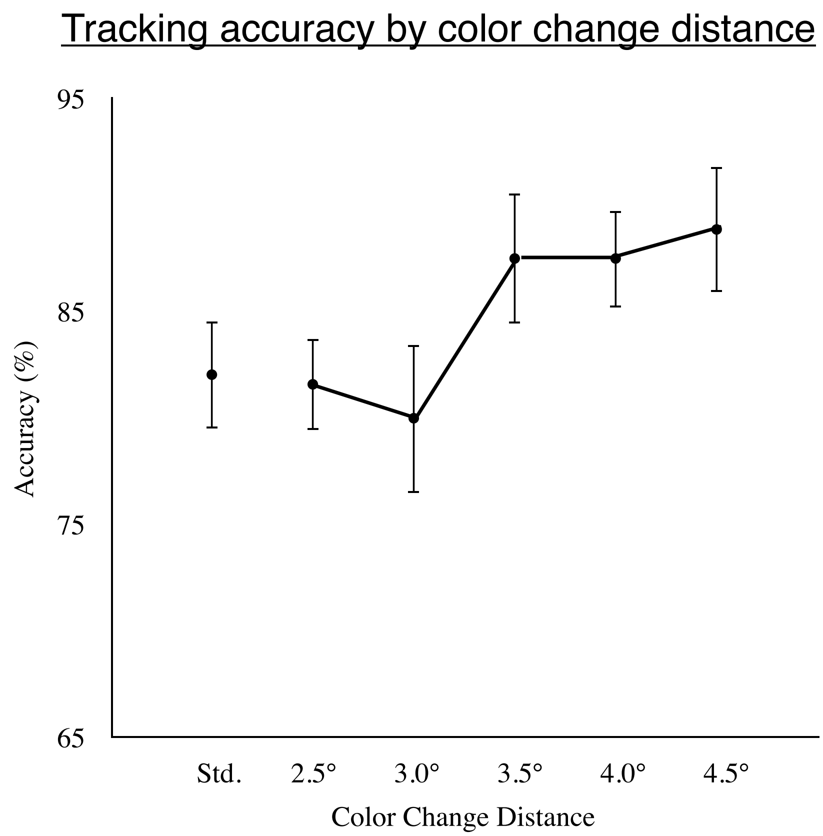

Figure 2.04. Mean tracking accuracy as a function of color change distance in Experiment 3. Error bars indicate ±1 standard error of the mean. 'Std.' refers to Standard Trials in which no color changes took place. Experiment 3 employed 'medium' speeds from Experiment 1.

Results and Discussion

Figure 2.04 displays mean tracking accuracy as a function of color change distance. A one way ANOVA revealed a significant main effect of change distance, $F(5,50) = 4.87$, $p = .001$. A linear contrast was significant, $F(1,10) = 19.18$, $p < .01$, accounting for 78% of the variance. But inspection of the graph suggested a major improvement in performance when changes took place at a distance of 3.5˚ or more. A post-hoc analysis confirmed this intuition. A Scheffé correction compared Standard, 2.5˚, and 3.0˚ conditions with 3.5˚, 4.0˚ and 4.5˚ conditions, revealing a significant difference, $F(1,10) = 23.14$, $p < .0001$. This contrast accounted for 95% of the variance, far more than the linear contrast. These results imply that the color change benefit was effective only when the distance between a target and a non-target was about 3.5˚. Surprisingly, there was a distance outside which the color change manipulation was not helpful, and when color changes were provided below a critical threshold they had no impact as well.

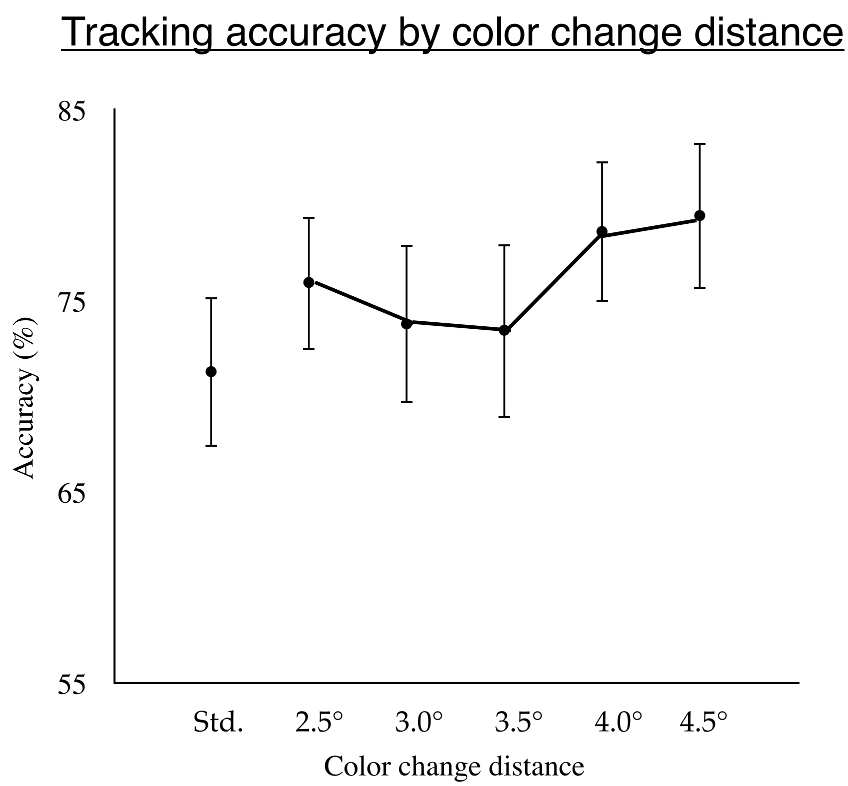

Figure 2.05. Mean tracking accuracy as a function of color change distance in Experiment 3b. Error bars indicate ±1 standard error of the mean. 'Std.' refers to Standard Trials in which no color changes took place. Experiment 3b employed 'fast' speeds from Experiment 1.

To replicate this counterintuitive effect I repeated Experiment 3 with a new group of 15 observers and object speeds that were 30% faster (Experiment 3b). Figure 2.05 displays mean tracking accuracy as a function of color change distance in the replication experiment. The same basic pattern was found —that is, a clear inflection point— only in this experiment the critical distance appeared to be 4.0˚. A linear contrast using Scheffé correction was not significant, $F(1,14) = 10.11 < F_{crit} = 11.72$, but a contrast between a group including Standard, 2.5˚, 3.0˚, and 3.5˚ trials compared with a group including 4.0˚ and 4.5˚ trials was significant, $F(1,14) = 12.73 > F_{crit} = 11.72$, explaining 77% of the variance.

Taken together, these results suggest that there is a critical distance that depends on the resolution of spatial attention at which close encounters become hazardous. Of course, I expect that attentional resolution becomes poorer at greater distances from eccentricity, and obeying Bouma's rule (Bouma, 1970; Pelli, Palomares, & Majaj, 2004; Whitney & Levi, 2011). Indeed, this is an explicit feature of a recent tracking model that depends on spatial precision (Vul et al., 2009). Thus, I emphasize here not the specific numbers discovered, 3.5-4.0˚, but the step function that seemed to describe the influence of color changes on performance. The specific numbers may reflect an average of critical distances at different eccentricities. However, the step functions reveal the dynamics of object confusions. Specifically, confusions happen at only those moments when targets and non-targets are confusable, and providing disambiguating information when they are not confusable (when they are far enough apart) does not help. Moreover, attempting to disambiguate objects only when they are very near one another — presumably, after they have already become confusable — does not help either.

Experiment 4: Are there color change benefits at large distances?

Experiment 3 tested the impact of color changes to non-targets at various distances, but in such a way that if a color change was induced at one distance, it remained present at all shorter distances. As a result, those data may not warrant the inference that no benefits can be gained when color changes are induced at greater distances. In Experiment 3, it appeared that there was no difference in performance for changes that were initiated at 3.5˚ compared to 4.5˚, suggesting that there was no added benefit to providing color information well before a close encounter. But if changes initiated at 3.5˚ brought performance up to ceiling, then I simply may not have been able to observe the additional gains accrued by initiating changes at longer distance. Accordingly, in the current experiment, I initiated changes to non-targets at a distance of 5.5˚, but those changes were undone when a non-target moved within 4.5˚ of a target.

Method

Participants. 10 Johns Hopkins University undergraduates participated for course credit. Each had normal or corrected-to-normal visual acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Apparatus, Stimuli, & Procedure. This experiment was identical to Experiment 1, except that in Color Change trials, color changes to non-targets were induced whenever a non-target was within 4˚ to 5.5˚ of a target.

Results and Discussion

Figure 2.06 displays mean tracking accuracy as a function of speed and color change condition. A two-way ANOVA revealed no main effect of color change condition on tracking accuracy, $F < 1$. There was a main effect of speed, $F(2,18) = 17.588$, $p < .001$, however, and an interaction between speed and color change condition, $F(2,18) = 3.844$, $p = .04$. The interaction was caused by better performance in Standard trials compared to Color Change trials at Medium speed, but a simple main effect using

Scheffé correction was not significant, $F(1,18) = 6.96 < F_{crit} = 22.06$. These results confirm the impression left by Experiment 3, that there are no gains to changing non-target colors when those non-targets are outside a confusable distance of a target.

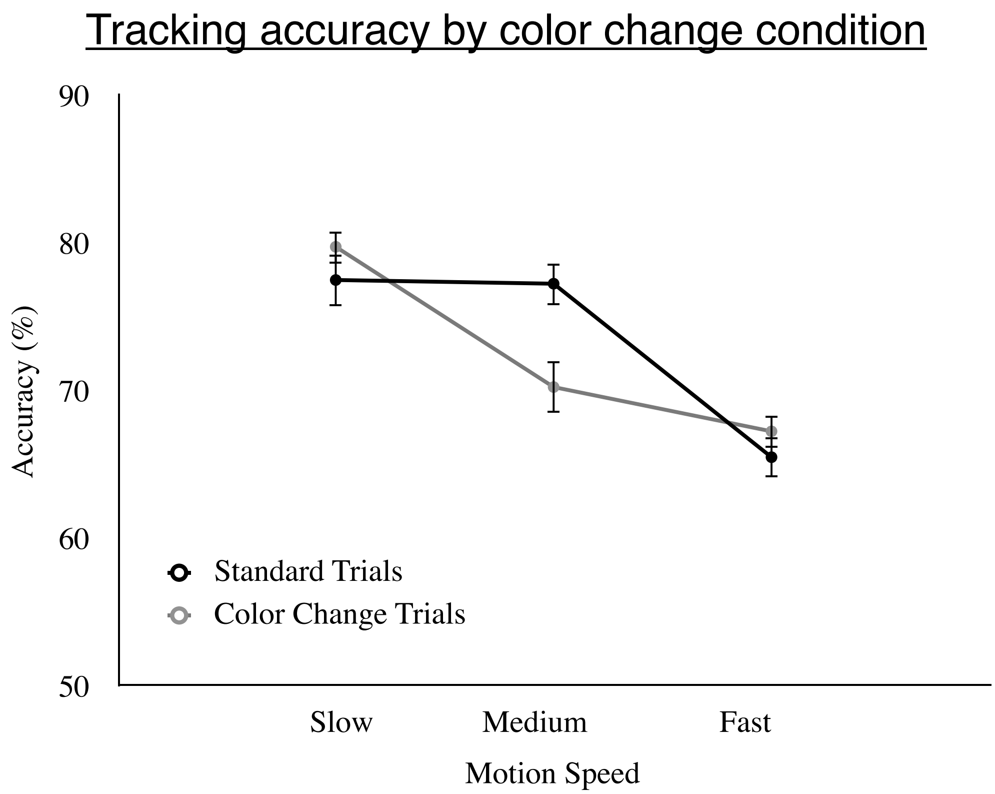

Figure 2.06. Mean tracking accuracy as a function of speed and color change condition in Experiment 4. Error bars indicate ±1 standard error of the mean. Color Changes in this experiment took place only when a non-target was within 4˚ to 5.5˚ of a target.

Experiment 5: Does performance usually depend on the number of close encounters?

The experiments conducted so far have suggested that tracking is limited by correspondence problems that emerge when objects approach within a critical distance of one another. Changing non-target features during these critical moments demonstrated that correspondence mistakes can be prevented. In two experiments, performance only improved when color changes were induced at a fixed distance from a target, and surprisingly, in Experiment 4 no benefit accrued for changes that took place at further distances.

The current experiment sought to demonstrate that close encounters at a distance of about 4˚ predict tracking performance in a standard MOT task, one without any color changes. As participants tracked, I counted the number of close encounters that took place between targets and non-targets at a distance of 4˚. I expected that, trial-by-trial, the number of close encounters that took place would predict tracking performance.

Method

Participants. A separate group of 11 Johns Hopkins University undergraduates participated in exchange for course credit. Each had normal or corrected-to-normal visual acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Apparatus, stimuli, and procedure. This experiment was identical to Experiment 3, except as follows. (1) There were never any color changes; every item in the display was indistinguishable from every other item throughout a trial. (2) In order to produce a wide range of number of close encounters, I tested participants in trials with four or five targets among ten or twelve non-targets. Half of all trials included four targets, while half included five. Counterbalanced with this manipulation, half of all trials included ten non-targets, while half included 12.

Data analysis. ENIT was computed on a trial-by-trial basis for each subject. For each trial I also counted the number of 4˚ close encounters. In order to compare these two measurements, data was collapsed across participants and binned trials into 9 bins for each target load so that each bin included at least 10% of all trials (i.e. 35 trials). This was done because there were relatively small numbers of trials with many and few close encounters. Total of nine bins were produced for each target load, each with a range of number of close encounters. The independent variable in the correlational analyses, below, and on the x-axes in Figure 2.08 was the average number of close encounters in each bin.

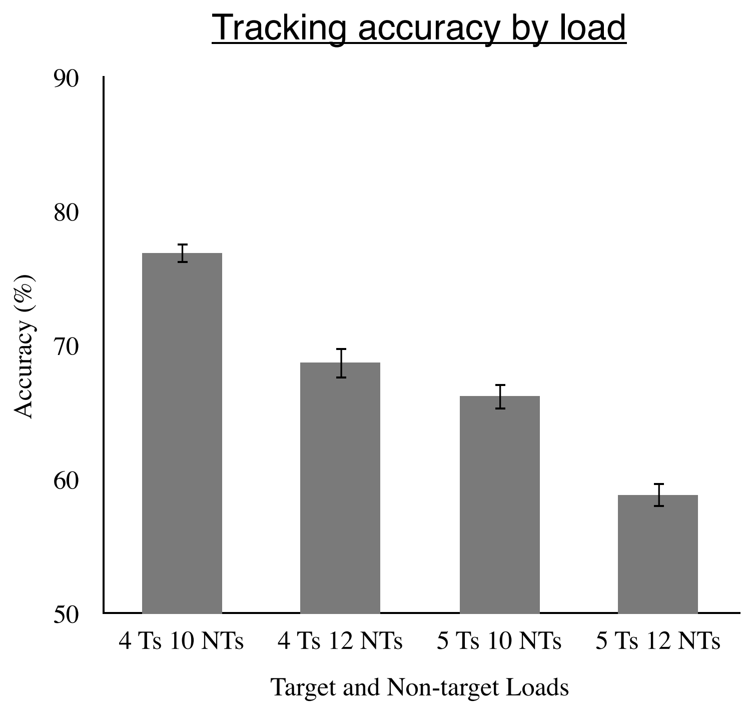

Figure 2.07. Mean tracking accuracy as a function of target (T) and non-target (NT) loads in Experiment 5. Error bars indicate ±1 standard error of the mean.

Results and Discussion

Figure 2.07 displays mean tracking accuracy as a function of number of targets and non-targets. Since the intention was to compare performance to the number of close encounters in each trial, simple accuracy seems to be not appropriate as a measure of performance. This is because accuracy reflects different levels of performance depending on target load.

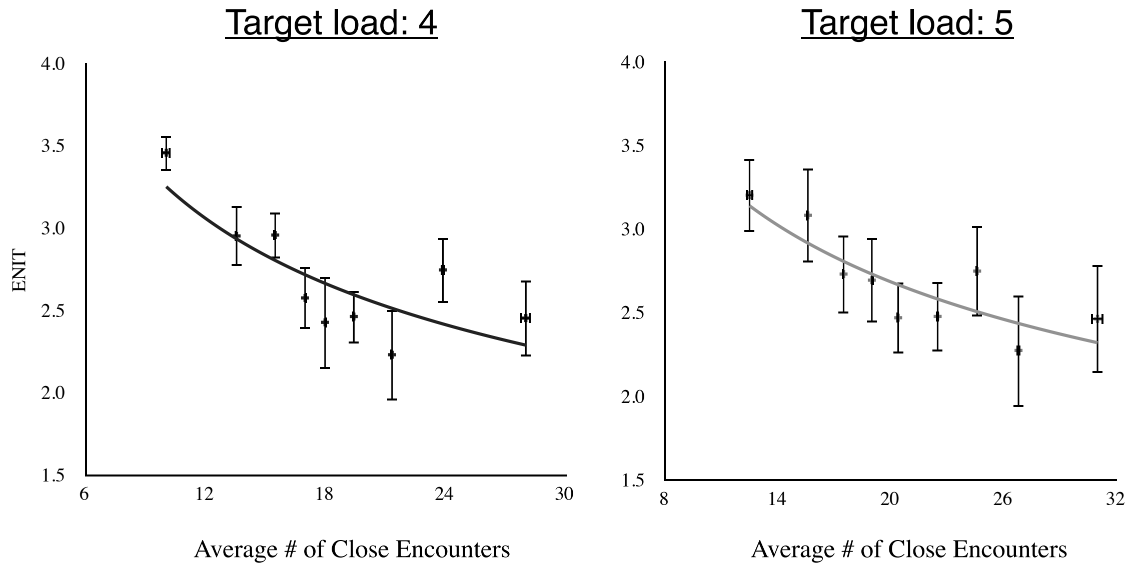

Figure 2.08. Effective number of item tracked (ENIT) as a function of the number of average close encounters (in nine bins) in Experiment 5. Close encounters were counted when a non-target approached within 4˚ of a target. Bins were created to each include at least 10% of the data, and the average number of close encounters in each bin is the independent variable (See Methods of Experiment 5). Error bars indicate ±1 standard error of the mean. Curves reflect fitted power law functions.

Figure 2.08 displays the relationship between ENIT and the average number of close encounters in each bin by tracking load. Power law functions were fit to these results because inspecting the data suggested these fits, and because once lost, a target cannot be lost again. These relationships were significant for both target load four, $F(1,7) = 11.039$, $p = .013$, $R^2 = 0.612$, and target load five, $F(1,7) = 17.075$, $p = .004$, $R^2 = 0.709$ suggesting power law declines in performance as a function of the frequency of close encounters. Linear fits were also significant, $t(7) = -2.73$, $p = .029$, $R^2 = 0.516$ for a target load of four, and for a target load of five, $t(7) = -3.56$, $p = .009$, $R^2 = 0.644$. These results furnish proof of concept that the step function inflection of about 4˚ revealed in Experiment 3 can be used to understand tracking performance in a standard version of a MOT task.

MOT Summary and Conclusion

The current study sought to supply evidence that tracking errors in MOT are caused by close spatial encounters in which imprecise spatial knowledge causes targets and non-targets to become confusable. Experiment 1 supplied this evidence by improving tracking performance through a color change manipulation that reduced the confusability of targets and non-targets whenever they were engaged in a close encounter. Experiment 2 demonstrated that participants could not use non-target changes to infer the identity of hidden target, excluding a target recovery account of the results of Experiment 1. Experiment 3 varied the distance at which non-targets changed colors, and suggested that items become confusable when they are about 4˚ from one another, at least when tracking four targets. Moreover, this experiment suggested that color change benefits do not accrue when changes are supplied at very short distances, perhaps because errors accrue quickly within the margin between the distance at which objects become confusable and smaller values. Similarly, Experiment 4 revealed no advantage for changes induced only at large distances (and not persisting at shorter distances). In other words, this experiment suggested that items that are not confusable in the first place — because they are far enough apart from one another — cannot be rendered any less confusable. Finally, Experiment 5 demonstrated that the distances at which color changes had their impacts in Experiment 1-4 could be used to predict tracking performance in trials when no color changes were supplied.

Overall, these experiments demonstrated that tracking errors can be reduced or prevented, that the distances at which targets and non-targets become confusable can be measured, and that the measured distances obtained relate to performance limits in standard MOT tasks. Future studies should explore in even greater detail why specific distances are the ones at which objects become confusable, and the relationship between these distances and the time or effort that may be needed to resolve potential confusions favorably. Additionally, future studies can employ my color change manipulation to investigate the distances at which objects become confusable as a function of, for example, speed, eccentricity, target load, and object size, among other factors.

<u>What constrains tracking performance?</u>

Since the seminal experiments of Pylyshyn & Storm (1988) perhaps the central question about object tracking has been, ‘why is it limited in the first place?’ Pylyshyn took a theoretically motivated approach, arguing early on that it is limited because I possess only a limited number of discrete pointer-like representations, FINSTs. More recently, numerous studies have addressed this question by cataloguing display features that influence tracking performance, including object speed, trial duration, target load, number of non-targets, and the spatial separation between items (e.g. Oksama & Hyona, 2004: Franconeri, Lin, Pylyshyn, Fisher, & Enns, 2008; Franconeri et al. 2010; Sears & Pylyshyn 2001; Shim, Alvarez, & Jiang, 2008). Looking through this list, it becomes apparent that these display features may share a more proximate reason for limiting tracking performance, namely, they influence the likelihood that targets and non-targets will come in close enough proximity to one another to become confusable. Several studies have supplied evidence that confusions sometimes take place, resulting in the tracking of non-targets (O'Hearn et al., 2005; Sears & Pylyshyn, 2000), and recently Franconeri and colleagues (2010) have suggested that inter-item spacing accounts for all tracking limits, even providing striking evidence that other factors can be reduced to interactions with spacing. The current results fit well with this point of view, supplying evidence that confusions during close encounters constitute the most proximate causes of tracking errors by demonstrating that such confusions can be prevented.

Previous work has left ambiguous why close spatial encounters lead to tracking errors. Franconeri and colleagues, for instance, argued that target and non-target representations include inhibitory surrounds that engage in destructive interference when items approach in close proximity, including when targets approach within close proximity of on another. Though my study was not designed to test such an account, it does not accord well with the current results. Specifically, it is not obvious why changing a non-target's color would affect the center-surround inhibitory structure of target and non-target representations and their resulting interactions. Franconeri and colleagues also argued that targets interfere with one another when they approach in close proximity. But my manipulation likely had no impact on how such encounters were negotiated, though it improved performance nonetheless. It may be that my results can be reconciled with an account of tracking limits that relies on inhibitory interactions, and it may also be that some of the errors in my experiments were caused by target-target interactions. But these possibilities do not seem necessary to understand the results presented.

Moreover, prior work has demonstrated clearly that attentional resolution is limited (Intriligator & Cavangh, 2001; Tsal & Bareket, 2005), and that the precision of spatial working memory is limited (Bays & Husain, 2008). Put simply, human spatial knowledge is, and really must be, imprecise. As a consequence, nearby items should often be confusable (Moore, Lanagan-Leitzel, & Fine, 2008), particularly the more similar they are in appearance. If confusions between targets and non-targets are to be expected because of imprecise spatial knowledge, then it seems unnecessary to hypothesize additional mechanisms to account for errors emerging during close spatial encounters. At a minimum, I would want to quantify the number of errors that can be accounted for on the basis of confusions, alone, in order to know how much variance remains to account for by other mechanisms. Doing this will be challenging given that failures due to close spatial encounters likely emerge probabilistically. Indeed, a recent model of MOT accounts for many tracking phenomena simply by formalizing the task as probabilistic inferences about target identities given uncertain spatial knowledge (Vul et al., 2009). Results of the current study are consistent with this model, identifying the moments during which inferences about target identities are most likely to fail.

The nature of multiple object tracking commodity limits

While recent theories regarding tracking limits differ considerably from earlier approaches, all theories to date seem to share an assumption that, at some point, tracking is a resource limited process. But newer theories have conceived of the underlying resource differently, as a flexible, continuous commodity rather than a fixed and discrete one. According to such flexible-resource accounts, a limited commodity gets consumed by targets, and it provides less representational resolution the more times it is divided (Horowitz & Cohen, 2010). Accordingly, when more targets are tracked, each is tracked with less spatial precision, and perhaps features such as speed, crowding, and the number of non-targets also consume limited resources, resulting in their own contribution to declining spatial precision in target representations (Alvarez & Franconeri, 2007; Bettencourt & Somers, 2009; Holcombe & Chen, 2012).

Crucially, while all prevailing theories of multiple object tracking assume commodity limits that constrain performance, fixed and flexible theories make very different predictions about when and how errors emerge. Fixed-resource theories, like Pylyshyn's FINSTs (1989, 2001), but including other views as well (e.g. Luck & Vogel, 1997; Drew & Vogel, 2008) predict that mostly random errors should arise, probably at the start of a trial, because more targets are presented to track than can be accommodated by a limited number of discrete representations. In contrast, flexible-resource theories should predict that errors arise at various points during a trial because degraded representational precision leads to confusions among tracked objects (e.g. Alvarez & Franconeri, 2007, Horowitz & Cohen, 2010; Vul et al., 2009). A vigorous debate concerning fixed and flexible resources and the specific kinds of errors they predict has appeared over recent years in the visual working memory literature (Alveraz & Cavanagh, 2004; Awh, Barton, & Vogel, 2007; Bays & Husain, 2008; Fukuda, Awh, & Vogel, 2010; Zhang & Luck, 2008).

The current results are problematic for fixed resource views. My experiments demonstrate that errors arise at various points in a trial, and that they are caused by specific kinds of interactions between objects, not simply an inability to represent some objects, causing them to be left out at the start of a trial. I acknowledge, however, that it may be that both kinds of errors emerge, and that multiple kinds of limits exist —limits on both the total number of representable items and on the precision of items. What this study demonstrates then is the possibility of directly studying the kinds of errors that produce tracking performance and their micro-genesis. The recent move towards explicit and quantitative models of how tracking takes place should be followed by a move towards observing and preventing, when possible, the specific kinds of errors that the models predict.

Implications for forthcoming experiments

With respect to the remainder of this dissertation, the experiments conducted above reveal two general points that are worth noting here. The first is that questions of how correspondence computations work, really, how *well* they work, are intimately related to questions concerning the limits of performance. To the extent that correspondence computations limit performance, how much needs to be characterized in order to properly characterize additional constraints. Second, these experiments reveal a general strategy that will become apparent in the subsequent two chapters. Namely, all the reported studies seek ways of alleviating local correspondence challenges as a method to show the effects that they otherwise have on performance.

Chapter 3. Correspondence Computations in Spatial Working Memory

(The results and experiments in this chapter have been submitted for publication)

Visual working memory (VWM) allows recent visual experiences to inform judgments about the world faced presently. It links one moment in time to the next, providing a means to integrate information across saccades and supporting everyday activities such as visual comparison, search, and change detection (e.g., Hollingworth 2009, Hyun et al. 2009; Jiang et al. 2009). Given this functional significance, it may seem surprising that the ability to use VWM is quite limited and that judgments made on its basis are frequently erroneous. But common laboratory tasks reveal significant difficulty in using VWM when more than three or four items must be held in memory at the same time (e.g., Luck & Vogel, 1997; Alvarez & Cavanagh, 2004). The fact that VWM is both practically beneficial and highly limited is a central puzzle for cognitive science.

Several decades of research have delineated the nature of working memory limits and explored their underlying cognitive and neural bases. This research has focused primarily on the *contents* of VWM: the format, number, and fidelity of the representations that this memory system can hold. For example, vigorous debate surrounds the issue of whether visual working memory stores bound objects or unbound features, and whether visual memory resources are allocated in discrete packets or continuously (e.g., Anderson, Vogel, & Awh, 2011; Alvarez & Cavanagh, 2004; Awh, Barton, & Vogel, 2007; Barton, Ester, & Awh, 2009; Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Fukuda, Awh, & Vogel, 2010; van den Berg, et al., 2012; Wilken & Ma, 2004; Zhang & Luck, 2008).

The intense focus on contents has led to a relative neglect of the procedures involved in *using* memory. Storing information is not sufficient for a functional working memory system, regardless of what is stored and how faithfully it is represented. There must in addition be computations that relate stored representations to the current view (Ullman, 1984): operations that can determine whether a previously seen object has changed in some way, whether an old object is now absent or a new object has appeared, and so on (Hollingworth, 2003; Mitroff et al., 2004; Simons, et al., 2002; Hyun et al., 2009). Very little cognitive research has attempted to catalogue such computations or articulate explicit algorithms that carry them out. In this chapter, I investigated a computational problem that is logically fundamental to many tasks involving VWM: how does an observer establish a relation of *correspondence* between items in memory and those presently observed? In this section, I provide independent evidence that correspondence challenges lead to errors in a spatial memory task. I then describe a computationally explicit method for establishing correspondence, and show that its performance matches that of human participants. Finally, I present novel task manipulations that facilitate correspondence inferences, and demonstrate that these improve working memory performance.

I chose to focus on spatial working memory for a number of reasons. The particular spatial task that I adopt has been theoretically influential (Bays & Husain, 2008). More importantly, I expected to be able to articulate concretely and formally the ways that correspondence computations might be addressed. Specifically, I expect that when spatial position is the task-relevant property correspondences are established in a way that favors perceived spatial proximity. This proximity assumption is sensible in the natural world, where objects are perceived to move continuously (in the absence of occluders), and it is a central heuristic in the perception of motion, apparent motion, and the tunnel effect, as I described in Chapter 1.

In spatial working memory, therefore, I expect correspondence errors to arise largely from spatial confusions, affording the opportunity to *predict* trials in which they would be more likely to arise. Knowing why correspondence illusions arise in turn allows me to design manipulations that reduce their impact and to provide a computational account of spatial VWM that captures key properties of human performance.

Experiment 6 addresses the main empirical goal of the current study: I replicate a common spatial working memory task and show by analysis of participant errors that some spatial configurations are more likely to give to rise to correspondence errors. Relative error rates can be understood by considering the relations of proximity among the items in the memory display and the probe at test. Experiments 7 and 8 employ a novel preview manipulation designed to facilitate the solving of correspondence problems, and thereby improve human performance. Experiment 9 is a control for these two experiments, and Experiment 10 is an additional control exploring the potential role of a non-spatial feature (color) in solving spatial correspondences. In addition to these empirical results, I show how a formal model that is uncertain about correspondences, but solves them through Bayesian inference under the proximity assumption, supplies an excellent fit to observed human data from Experiments 6 & 7.

Comparison of this model with alternatives yields two additional findings, both of which are unexpected in light of the focus of previous VWM research. First, providing the model with perfect correspondence knowledge — knowledge of which memory item should be compared to the observed probe — *decreases* its ability to predict human performance. This highlights the non-trivial nature of correspondence inferences, and suggests that all future models of human spatial VWM should incorporate a correspondence component. Second, imposing storage or resource limits on spatial memory does not significantly improve the fit beyond what is possible with correspondence errors alone, suggesting that the major limiting factor on human performance in the present task arises from memory use rather than memory contents. These surprising results underscore the need for further research on the procedures of memory use.

Experiment 6: Correspondence challenges impair spatial working memory

To investigate whether correspondence errors contribute to performance limits in spatial working memory, I tested participants in a common laboratory task (Bays & Husain, 2008). Participants memorized the locations of one to six colored squares, and after a brief delay they were asked to report the direction of displacement (left or right) of a redisplayed item (i.e., the probe; Figure 3.01.). In previous experiments, error rates increased with increasing memory load, suggesting that working memory failures are more common as participants attempt to remember more items. These results are typically interpreted to reflect well-known resource limits in visual working memory. It has been suggested that participants make more errors at larger loads either because they exhaust available ‘slots’ in which store items (Thiele, Pratte, & Rouder, 2011) or because they allocate dwindling amounts of a continuous resource to the memory representation of each item (Bays & Husain, 2008). I discuss these contrasting theories at greater length in the Summary and Conclusions section of this chapter. For now, my main interest is the extent to which correspondence problems may contribute to observed error rates.

Independent from any storage limits, are some of the errors observed in this task caused by failures to correctly solve the correspondence problem?

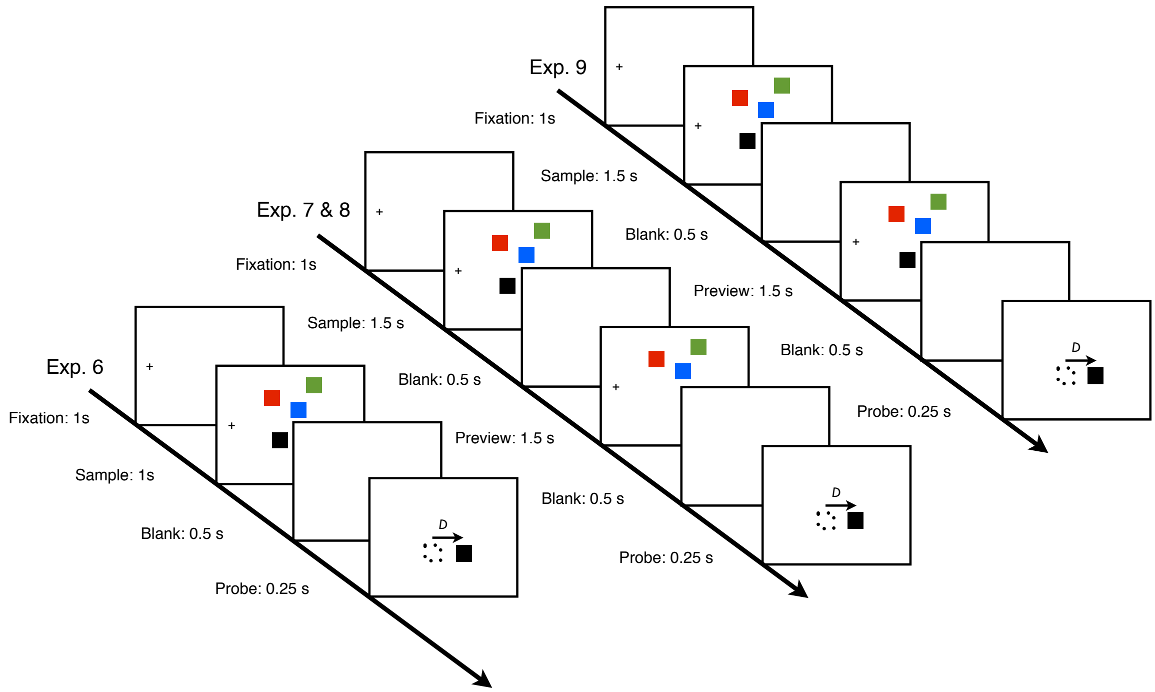

Figure 3.02. Experimental procedures for Experiments 6-9. In all experiments, participants memorized the positions of briefly presented colored squares. At test, one of the squares —the probe— reappeared, displaced to varying degrees (*D*). Participants made a response indicating whether the displacement was to the right or to the left. In Experiments 7 and 8, a preview redisplayed all the original memory items, except for the probe, just before test. In Experiment 9, all items, including the probe, were redisplayed in the preview.

Clearly, in this task, the response given to the probe depends on the memory representation to which it is compared. As discussed in the introduction the process of establishing correspondence for the purpose of spatial comparison is expected to favor memory items that are perceived to be proximal to the probe. If participants infer correspondence on this basis, then memory items that are closer to the probe than the correct target will act as lures for illusory correspondence.

Crucially, incorrect correspondence should translate into observed errors only when they lead to a task response — left or right — that is different from the one demanded by the correct correspondence. In short, wrongly inferred correspondences can result in errors only if they support a response that is inconsistent with the response that would be made based on the correct correspondence. To reveal the potential influence of correspondence errors on performance, for each trial I first identified the non-target memory item nearest to the probe. I then sorted these trials into two categories.

Consistent trials were those on which the probe was displaced in the same direction relative to the nearest non-target and the target. (In other words, these are trials in which a comparison with either of these items would demand the same final response). *Inconsistent* trials were those on which the perceived direction of a probe's displacement would differ depending on which memory item was selected for comparison. Thus, *Inconsistent* trials were those on which mistakenly comparing the probe to the nearest non-target item —a comparison based on incorrect correspondence— would result in an erroneous response. If incorrect correspondence is a source of error in this task, I predicted greater error rates in *Inconsistent* compared to *Consistent* trials.

Methods

*Participants***.** 12 Johns Hopkins University students participated for course credit or a small monetary compensation. All had normal or corrected-to-normal visual acuity. The experimental protocol was approved by the Johns Hopkins University IRB.

Apparatus. The experiments took place in a dimly lit sound-attenuated room. Stimuli were presented on a Macintosh iMAC computer with an LCD display at a viewing distance of 60 cm such that the display subtended approximately 39.56° X 25.35° of visual angle. MATLAB and Psychophysics toolbox were used to generate stimuli and collect responses (Brainard, 1997).

Procedures. At the start of each trial a black fixation-cross (0.5° x 0.5°) appeared in the display, displaced 10° horizontally from center, either to the left or the right (side was randomly counterbalanced across trials). The fixation-cross remained in the display for one second. Next, a sample display was presented for one second. The sample display consisted of between one and six colored squares (0.8° x 0.8°).

Colors were randomly selected on each trial (without repetition) from a set of highly distinguishable colors (white, black, red, green, blue, yellow, dark brown, and cyan). Memory load was determined by the number of squares presented (1-6 items), and was counterbalanced across trials such that each participant observed 100 trials at each load. Sample items were presented in an invisible square (9° x 9°) centered vertically, and displaced 10° horizontally from the fixation cross. Within the presentation square, item positions were chosen randomly, with the restriction that they could never be less than 1.5° from one another.

After one second, a blank gray screen appeared for 500 ms. A probe display followed for 250 ms. The probe was a single square, drawn in the same color as the corresponding sample item, but displaced horizontally from its original location by a deviation of 0.5º, 1.5º, 2º, 3.5º, or 5º (left or right, counter-balanced across the trials). The identity of the probe item was selected randomly at the start of each trial. The task was to report the direction of displacement (left or right) via a keypress. A blank screen remained present until a response was made. The next trial began 500 ms after a response. Each participant completed a different unique set of 600 trials.

Analysis. I identified two types of trials, *Consistent* and *Inconsistent*, among all trials with a memory load greater than one. In each trial, I computed the (Euclidean) distance from the probe to each item in the memory display. I then identified the non-target item nearest to the probe and determined the response that would be made if the probe were (mistakenly) placed in correspondence with that item. If the response matched the response dictated by the true target, then the trial was classified as *Consistent*, since on these trials comparing the probe to the nearest non-target item would have (accidentally) resulted in a correct response. In contrast, if the response demanded by the nearest non-target opposed the response demanded by the target (i.e., if one was to the left of the probe and the other was to the right) the trial was classified as *Inconsistent*.

Results and Discussion

As in previous experiments, error rates increased with increasing memory load, suggesting that working memory failures were more common as participants attempted to remember more items (Figure 3.02; $F(5,66)=14.20$, $p < .001$). Additionally, I analyzed performance as a function of the spatial disparity between the probe and the true target item. These results are displayed by memory load in Figure 3.03. The abscissa of each graph displays the size of the disparity (in degrees of visual angle) with negative numbers designating leftward displacements and positive numbers designating rightward displacements. The ordinate axis plots the probability with which participants reported a probe as having moved rightward relative to its original position. Black circles designate average human performance at each displacement and memory load. It should be clear that participants made more errors at smaller displacements, and also that they made more errors with smaller displacements at larger loads. This is a typical feature of performance on tasks like this one, a feature that is often taken to reflect continuous storage limits (Bays & Husain, 2008). I will discuss this pattern of results as potentially reflecting correspondence challenges, in light of my model, in the next section.

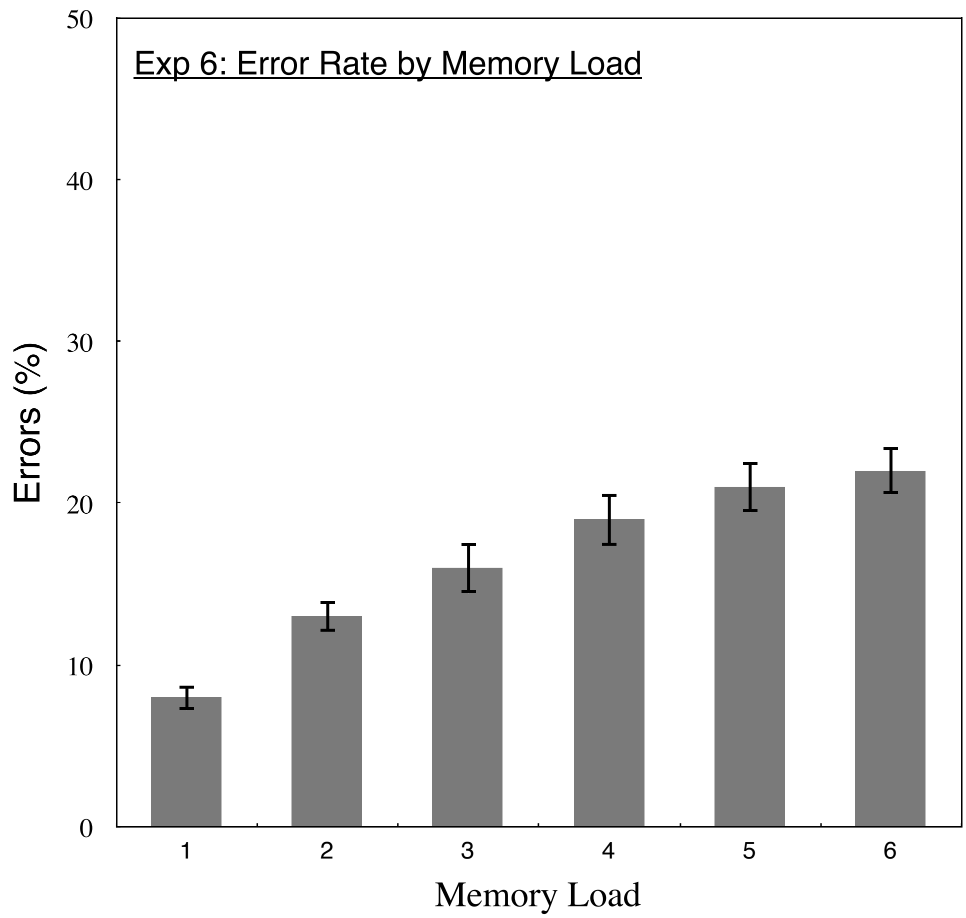

Figure 3.02. Error rate (± 1 SE) by memory load in Experiment 6 (N =12). Mean error rate increased as the number of objects displayed in the memory display.

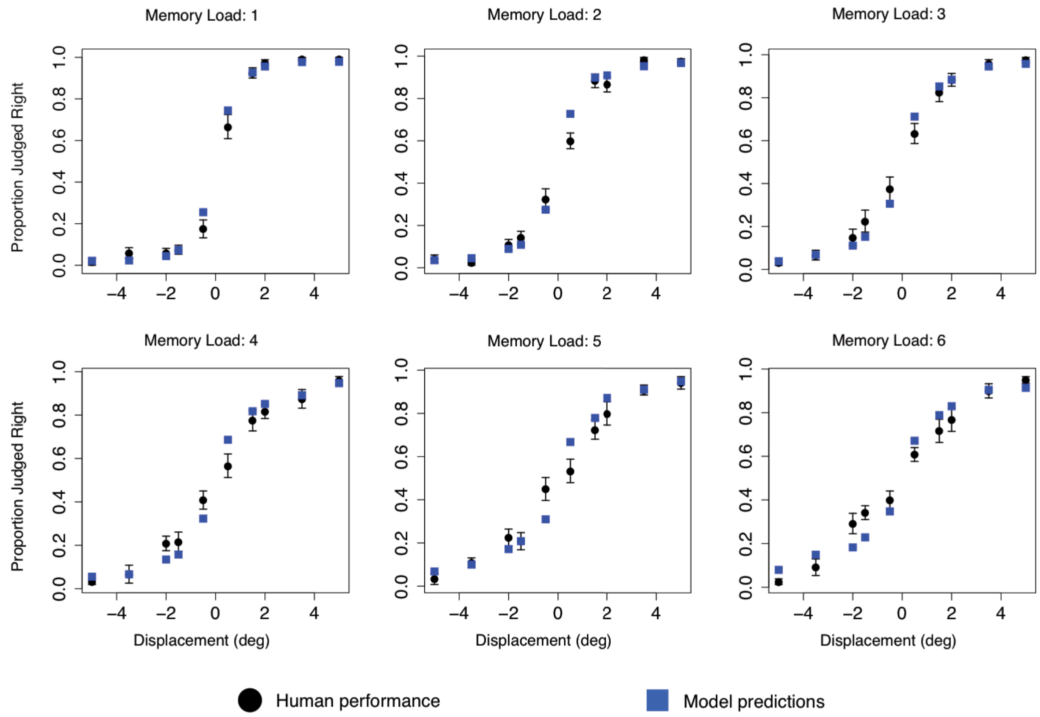

Figure 3.03. Human performance and model predictions as a function of spatial disparity and memory load in Experiment 6. Participants judged whether a probe moved left or right relative the starting position of a remembered target item. True displacements to the right are designated by positive numbers and true displacements to the left are designated by negative numbers on the x-axis of each graph. On the y-axis, each graph depicts the probability of a rightward response. Perfect performance would be reflected by a 0.0 probability of a rightward response for all negative displacements and 1.0 probability for all positive displacements. Generally, participants supplied correct responses more frequently for larger displacements than smaller ones. Error bars reflect ± 1 SE.

For current purposes, I was primarily interested in the impact of nearest non-target response consistency on performance. Accordingly, I conducted a logistic regression with disparity size, *Consistent* vs. *Inconsistent* trials, and memory load as predictors of performance. The primary result from this experiment was that trial consistency was a significant predictor of performance ($z = 4.56$, $p < .001$). (This was also true at each individual set size, largest $p < .001$). *Inconsistent* trial error rates were considerably larger than *Consistent* trial error rates (Figure 3.04), with 62% of all errors occurring in *Inconsistent* trials.

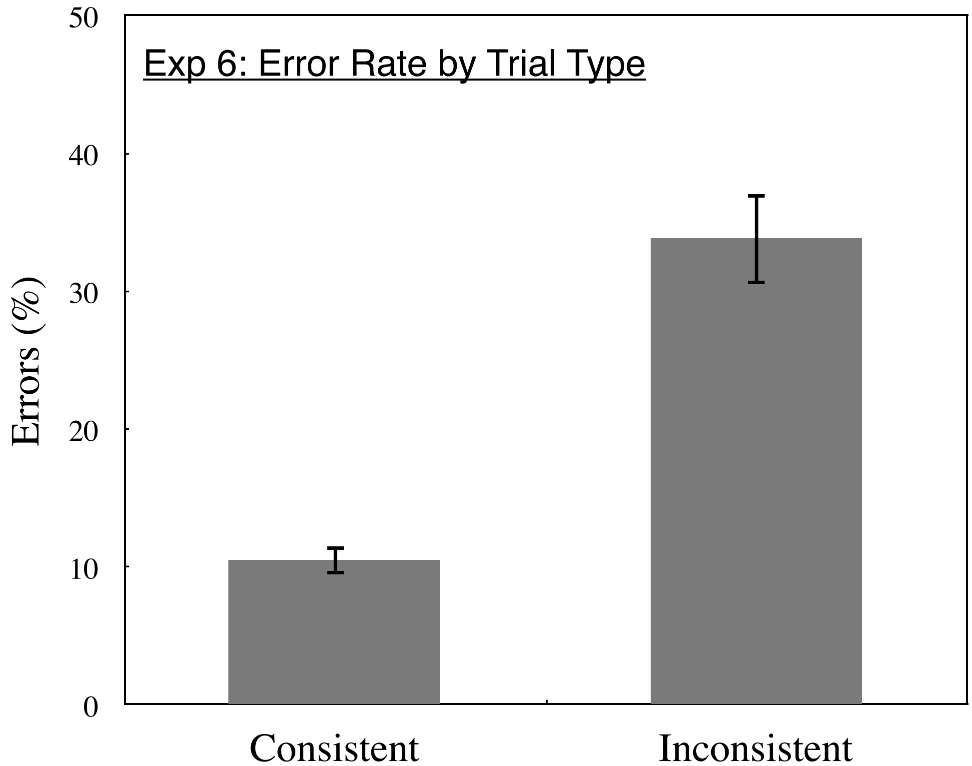

Figure 3.04. Error rates (± 1 SE) in Experiment 6 in *Consistent* compared with *Inconsistent* trials. As described in the text, all trials with memory loads of two or larger were classified as either *Consistent* or *Inconsistent* based on the directional response demanded by the nearest non-target memory item to the test probe. *Inconsistent* trials were those in which a comparison with the nearest non-target would have led to an erroneous response. Error rates in *Inconsistent* were considerably larger than in *Consistent* trials.

I also conducted analyses to consider a potential strategy that may have been available to participants in this task. Specifically, with certain item configurations and probe displacements, probes may have sometimes appeared to the left or right of all the items previously in the display (Thiele et al., 2011). Participants may have detected such instances, knowing what to report by comparing the probe to the group's envelope as opposed to any individual memory representation. Accordingly, I identified all trials in which a probe appeared to the left or right of all the items previously in the display, and I then excluded those trials from a subsequent comparison between *Consistent* and

Inconsistent trials. Some trials that I categorized as *Consistent* fell into this category, and all *Consistent* trials for a memory load of two items necessarily fall into this category (leading us to exclude this load from the subsequent analysis). Even with these trials excluded, for memory loads of three through six, trial consistency significantly predicted performance ($z=2.20$, $p = .027$). While a configural strategy may have been available in this task, it cannot entirely account for the difference in performance between *Consistent* and *Inconsistent* trials.

The large proportion of errors in *Inconsistent* trials suggests that many of participants' errors may not have been caused by storage limits, but instead by failures to compare the probe to the correct memory representation. In retrospect, that this should be the case may seem somewhat obvious. Under prevailing theories, errors arise in comparing a probe to the correct target because of uncertainty in an observer's knowledge of spatial positions (Bays & Husain, 2008). But this same spatial uncertainty would also leave observers uncertain about correspondences. Considering the task from the perspective of an observer who cannot know in advance which memory item will be relevant, and must therefore infer correspondences between the original display and the probe position, it may be inappropriate to cast instances of mistaken correspondence as failures. It is possible that participants made (near-) optimal decisions given the information provided by the displays and their inherent uncertainty about item positions.

A correspondence-uncertain Bayesian model of spatial working memory

To explore the possibility that working memory limits reflect optimal behavior given uncertain inputs, I developed a probabilistic model that performed the same task as the human participants in Experiment 6. The goal was to implement a straightforward computation for selecting among possible correspondences (when multiple memory items are present), followed by comparison of the probe item with its correspondent in memory to generate a response.

The comparison between a probe and its correspondent in memory was formalized as a signal detection problem, following previous work (e.g., Bays & Husain, 2008; Wilken & Ma, 2004). Because all sensory measurements are subject to a certain level of noise, even comparisons that are based on correct correspondence can result in errors. However, in accordance with the *Consistent/Inconsistent* analysis of the previous section, I are particularly interested in errors that arise from the logically prior process of inferring correspondence between the probe and an item in memory. My model is the first one to address this issue in the context of spatial working memory.

The primary goal of these modeling efforts was to explore the ceiling on human performance put in place by the challenge of identifying correspondences. Accordingly, I designed my initial model with a small and fixed amount of sensory uncertainty. Performance limits arising under these conditions would thus reveal computational limits imposed on the task that are independent of storage limits. For the fixed level of uncertainty, I used a value derived from performance in the one-item trials of Experiment 6, trials in which participants could not make correspondence errors, but which did reveal uncertainty about item positions (as well as variability between participants). In this way, I implemented a model without any commodity-like storage limits — neither a limit on the total number of items that could be stored, nor a limit on representational precision that varied with memory load. But the model was nevertheless prone to correspondence errors, since correspondences had to be inferred from information about items in the memory and probe displays.

Below, I first provide an overview of the model followed by a formal and detailed description. I then report results of a comparison between the model and observed human performance, and finally, I describe further model variations that were used to determine whether representational resource limits impose constraints on human performance above and beyond the computational limits set by the correspondence problem. Since larger memory loads supply a greater number of opportunities for correspondence mistakes, the model will inevitably perform more poorly at larger memory loads (see also Duncan, 1980; Eriksen & Spencer, 1969; Kantowitz, 1985; Navon, 1984). Comparison of my model with minimally different variants establishes the centrality of correspondence inferences and suggests that resources limits are not a main factor in predicting human performance in the present task.

Model overview

The probabilistic model that I developed was motivated by the fact that all sensory measurements include noise. As a result, I acquire probabilistic descriptions of the visual world (Purves, Wojtach, & Lotto, 2011). Research with both natural (Cheng et al., 2007; Kording & Wolpert, 2004) and artificial (Seger, Ulrich, & Christian, 2007) computational systems has described how properties of an object such as position, color, velocity, or orientation can be encoded in terms of a distribution of values that are compatible with the retinal image and prior beliefs. In using visual working memory, probability distributions lead to a well-known signal detection problem: any comparison between an observation and a memory (to determine whether they are the same or different) will end erroneously with some non-zero likelihood dependent on the parameters of the relevant distributions (Bays & Husain, 2008; Wilken & Ma, 2004). Thus comparing objects amounts to probabilistic inference.

A fact that is often overlooked, however, is that probabilistic representations lead to correspondence problems as well. If every feature of the visual world is represented probabilistically, an observer can have no way of knowing with certainty which observed items (if any) correspond with each of the items currently stored in memory. Algorithms must be in place to make these unavoidable judgments. Thus the process of executing the task in Experiment 6 involves at least two decision-making steps that should be hamstrung by probabilistic knowledge of item positions. The first involves making a decision about which item in memory corresponds with the observed probe, and the second involves making a directional judgment about the probe's displacement relative to its (inferred) correspondent in memory.

My probabilistic model carried out these two steps and it represented the positions of memory items and probe items probabilistically, as follows: (1) when a sample display appeared, the model inferred a probability distribution over the position of each object in the display, and it then stored those distributions during the delay period. The variances of these distributions were derived from the one-item trials in Experiment 6. (2) At test, the model inferred a probability distribution over the position of the observed probe. The variance in this distribution was also derived from the uncertainty observed in the one-item trials of Experiment 6. (3) Next, the model computed the probability of a correspondence between the probe and each of the remembered memory items on the basis of their vertical spatial proximity. (4) Finally, the model generated a binary task response (left or right) conditional on the inferred position distributions of the probe and the most likely corresponding memory representation. Calculations of all model components were derived by simple applications of Bayes' Theorem. Monte Carlo simulations were employed to generate the probability of a rightward response in each experiment trial (from Experiment 6; see e.g. Girschick et al. 2011). By representing spatial position probabilistically, the model could select different correspondences as the most probable on different simulations of a single trial. In this way, the model was probabilistically prone to correspondence errors.

Because I was interested in identifying computational limits, independent of any storage limits*, I fit model parameters to participant data from the one-item trials in Experiment 6, leaving them unchanged when I generated responses to the multi-item trials*. (Thus modeling those trials without free parameters). Specifically, I fit the model to each participant's one-item data with just two parameters: the standard deviation of the Gaussian positional noise distribution representing an observed object, and the probability of a random lapse (alpha), a common component of psychophysical models of discrimination thought to reflect (unbiased) guessing or inattention (Zhang & Luck, 2008). Flexible resource theories generally predict increased imprecision with increasing memory load (e.g. Bays & Husain, 2008), and fixed resource theories generally predict increased guessing with increasing memory loads (e.g. Zhang & Luck, 2008). By setting model parameters to those fit to one-item trials, I was able to explore ideal performance without any storage limits.

Model details

In the simulation of a given trial, the model is first provided with a noisy sample of the position of each item in the memory display. To provide a simple source of measurement uncertainty, these samples were drawn from isotropic Gaussian distributions centered on the object's true position:

$$z_i \sim N(\tilde{z}_{i,} \sigma^2 I) \quad [1]$$

Note that I adopt the convention of using $\tilde{z}_i$ to stand for the experimenter-determined position of the *i*th item, while z_i denotes a noisy measurement of that position. By a straightforward application of Bayes' Theorem (Bishop, 2006), and under the common assumption that the model has knowledge of its own measurement variability (σ in the above equation), the model infers a posterior distribution for each object's position as follows:

$$p(\hat{z}|z_i) \propto N(z_i; \hat{z}, \sigma^2 I) \quad [2]$$

Here $\hat{z}_i$ is a random variable over possible positions.

Following the processing of the memory display, the model receives a noisy sample of the probe's position (z_*) and infers a posterior distribution over possible probe positions ($\hat{z}_*$) exactly as in equation (2). This model of how the memory and probe displays are processed could certainly be elaborated — for example, I have not addressed how the model determines the number of items in the memory display (Ma & Flombaum, 2013; Smith et al. 2005) or, relatedly, how it combines a stream of perceptual measurements into the individual samples assumed above. But it makes the formulation of the comparison process straightforward and in line with previous models of VWM, allowing us to focus on the role of correspondence inferences.

Having inferred a probability distribution over the position of the probe item, the model next identified the probability of a correspondence between the probe and each memory item. In trials with a single memory item, correspondence is deterministic and the result of the comparison process fully determines the response. When there is more than one memory item, the model must determine which of the items to place in correspondence with the probe for the purpose of making the comparison. According to Bayes' Theorem, the posterior probability that a given memory item *k* (from among *n*) corresponds to the probe, conditioned on all of the measurement samples, is given by:

$$p(k|z_1, \dots, z_n, z_*) = \frac{p(z_*|k, z_1,\dots,z_n)p(k|z_1,\dots,z_n)}{\sum_{l=1}^{n} p(z_*|l, z_1,\dots,z_n)p(l|z_1,\dots,z_n)} \quad [3]$$

Without any a priori reason to favor one memory item over another, the model assumes that the prior probability of correspondence $p(k|z_1, \dots, z_n, z_*)$ is identical for all items and independent of their positions. The remaining factor $p(z_*|k,\ z_1, \dots, z_n)$ is constrained by the task-specific knowledge that the probe is displaced horizontally, but not vertically, relative to its position in the memory display. Intuitively, the model should favor correspondence with memory items that have vertical coordinates similar to that of the probe, and it should ignore the horizontal coordinates (which are subject to change). More technically, the model assumes that the *x*-coordinates of the memory items and the probe are generated independently; terms related to these correspondences drop out of the equation, together with the uniform prior probabilities. What remains after is a dependency between the vertical positions: if the probe is identical to the *k*th memory item, then $\tilde{z}_{ky}$ must be equal to $\tilde{z}_{*y}$.

Because an observer cannot know the true values of the *y*-coordinates — but must infer them from noisy measurements of position — the probability of correspondence with the *k*th item is determined by integrating over all possible values conditional on the samples:

$$p\left(k \middle| z_1, \ldots, z_n, z_{probe}\right) = \frac{\int d\hat{z}_{ky} p\left(\hat{z}_{*y} \middle| \hat{z}_{ky}\right) p\left(\hat{z}_{ky} | z_{ky}\right)}{\sum_{l=1}^{n} \int d\hat{z}_{ly} p\left(z_{*y} \middle| \hat{z}_{ly}\right) p\left(\hat{z}_{ly} | z_{ly}\right)}$$

$$= \frac{\int d\hat{z}_{ky}\, N\left(z_{*y}; \hat{z}_{ky}, \sigma^2\right) N\left(\hat{z}_{ky}; z_{ky}, \sigma^2\right)}{\sum_{l=1}^{n} \int d\hat{z}_{ly}\, N\left(z_{*y}; \hat{z}_{ly}, \sigma^2\right) N\left(\hat{z}_{ly}; z_{ly}, \sigma^2\right)}$$

$$= \frac{N\left(z_{*y}; z_{ky}, 2\sigma^2\right)}{\sum_{l=1}^{n} N\left(z_{*y}; z_{ly}, 2\sigma^2\right)} \quad [4]$$

where the last step follows from the Gaussian convolution identity.

By equation (4), the probability of correspondence decreases monotonically with the square of the difference between the sampled *y*-coordinates of the probe and the *k*th item. This accords with my intuitive understanding of how an ideal observer would assess the probability of correspondence in the present task, and can be seen as a refinement of the definition of *Consistent/Inconsistent* trials in the analysis of Experiment 6. Crucially, it is the distance between the samples (not the experimenter-determined positions) that is relevant for this computation. Therefore, the same measurement noise that creates errors in comparison can also lead to mistaken correspondence[1].

It remains to formalize the generation of a response — left or right — given an inferred correspondence between the probe and an item in memory. I assume that the model acts as an ideal observer, responding 'right' if and only if the conditional probability of a rightward displacement exceeds 0.5. To generate this conditional probability, and analogously to equation (4), the model integrates over all *x*-coordinates of the probe and its correspondent conditional on the relevant samples. Only contributions from pairs of probe/item values that support a rightward response contribute to this probability, which is formally given by:

$$
\begin{aligned}
p(Right|z_k, z_*) &= \int d\,\hat{z}_{kx} \int d\hat{z}_{*x}\, p(\hat{z}_{kx}|z_{kx})p(\hat{z}_{*x}|z_{*x})\, I\,[\hat{z}_{kx}<\hat{z}_{*x}] \\
&= \int_0^\infty d\Delta_x N(\,\Delta_x|z_{*x} - \, z_{kx}, 2\sigma^2) \\
&= 1 - \,\Phi(0;\; z_{*x} - z_{kx}, 2\sigma^2) \;\; [5]
\end{aligned}
$$

[1]There are at least two ways in which the observer could use the distribution over possible correspondences to inform the binary decision about displacement direction. The observer could commit to the single memory item that has the maximum *a posteriori* probability of being the target (MAP estimation). Or the observer could average over all memory items, each weighted by its posterior probability (mixture estimation). The modeling results reported in the text adopt MAP estimation, which I have found to slightly outperform mixture estimation in accounting for human performance. Importantly, even under MAP estimation the probe-item correspondence — and hence the binary decision — could vary over repeated simulations of the same memory display, since the vertical distances that determine the posterior correspondence probabilities depended on random quantities. Future work could explore the differences between MAP and mixture estimation methods. I also investigated models that employed two-dimensional computations of probe to memory item distances, as opposed to just vertical distances. The fits of these models were extremely close to one another, though the model employing only vertical distances outperformed slightly, as would be expected from the perspective of a Bayesian ideal observer possessing knowledge that an object could displace horizontally, but not vertically, between sample and test.

In this equation, $I\,[\hat{z}_{kx}<\hat{z}_{*x}]$ is an indicator function for rightward probe displacement and $\Phi(\cdot)$ is the Gaussian cumulative distribution function. It follows, trivially, that $p(\text{Right}|z_k, z_*) > .5$ if and only if $z_{kx} < z_{*x}$. As in the case of establishing correspondences, the response depended on the noisy samples of positions, and could therefore differ over repeated simulations with the same display, capturing the basic property of a signal detection perspective, that responses depend probabilistically on the properties of the underlying distributions compared.

Monte Carlo simulations The equations derived up to this point determine a binary response given particular noisy measurements $z_{1,\ldots\ldots,}z_n, z_*$. To derive predictions about binary response probabilities given the experimenter determined positions, $\hat{z}_{1,\ldots\ldots,}\hat{z}_n, \hat{z}_*$, in each trial, I did the following: I repeatedly generated samples according to the noise distribution in [1]. I formed posterior distributions on positions according to [2]. I then established correspondences by identifying the memory representation with the maximum *a posteriori* probability (MAP), and finally, I accumulated binary responses according to [5] (Girshick et al., 2011). When this process was simulated *N* times for a given trial, the average of the *N* binary responses provided an estimate of the predicted response probability for that trial; note that this probability is in general not identical to $p(Right)$, given above, which are conditioned on a particular set of noisy samples. For the simulations reported in the text, I set *N* = 10,000.

Parameter fitting As I noted previously, the model included two parameters: the imprecision of the Gaussian distributions describing object positions (σ) and a lapse parameter (α) reflecting unbiased guessing. I fit these two parameters individually to the data of each participant from Experiment 6 so that I could make predictions about performance for that group of participants. However, I fit these parameters only to the data obtained from one-item trials in Experiment 6, using those same values for evaluating the model at all memory loads. Predicted response probabilities provided a means for parameter fitting by maximum-likelihood (ML) to the observed responses in the one-item trials of Experiment 6.

Model results and comparison with human performance

My simulation produced results remarkably similar to those of human participants. For each trial, the model assigned a probability of a rightward response, as described above. I then assigned each trial to one of 100 equal-width bins with probabilities of rightward responses ranging from 0.00 to 1.00. Figure 3.05 displays the relationship between the probabilities assigned to each trial-bin by the model and the measured probabilities obtained from human participants in those trials. The model's fit was confirmed statistically by a strong linear relationship between the predicted probability of a response for each bin of trials and the proportion of trials in which the response was observed ($r = .96$).

Indeed, the model evidenced typical features of human behavior in this task. In Figure 3.03, model predictions are plotted as blue squares and results from human participants are plotted as black discs. By generating a model-based prediction for each trial and each participant in Experiment 6, I could compare model predictions and group data as a function of memory load. Like the human observers, the model made increasing numbers of errors as set size increased. Moreover, the model became less sensitive as memory load increased; the curve characterizing performance became less sigmoidal and more linear with increasing memory loads. This kinds of change in human performance

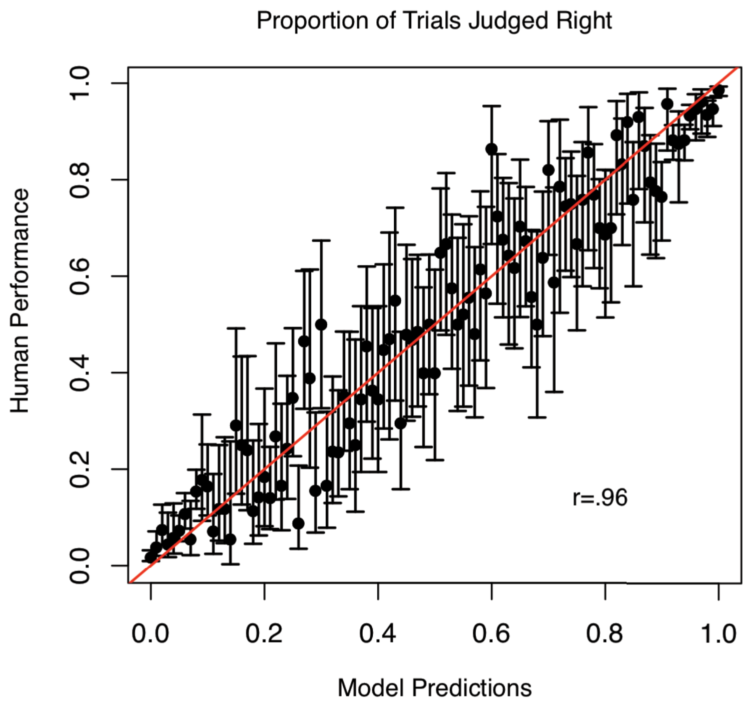

Figure 3.05. Relationship between model predictions and observed human responses for Experiment 6. My correspondence uncertain model supplied a probability that the group of subjects tested would make a rightward response for each trial in Experiment 6. These predicted probabilities are plotted on the x-axis. These trials were then placed into 100 bins based on the probability assigned by the model. The x-axis plots the observed probability of a responding rightward for each bin of trials. Error bars reflect 95% confidence intervals for these observations. There was a strong linear relationship between model predictions and observed responses.

has previously been taken to reflect a decrease in representational precision as memory load increases —that is, the pattern has been viewed as evidence that human observers become more uncertain about the position of each individual object the more they need to remember (Bays & Husain, 2008). But the model did not become more uncertain about each individual object's position when memory demands increased because I set uncertainty about position at the value derived from one-object trials. Patterns like the one observed here have also been taken to reflect discrete storage limits, limits on the number of objects an individual can store in memory simultaneously (e.g. Zhang & Luck, 2008). Such limits should lead to increased guessing at larger memory loads. But again, I fixed the guessing rate at a level obtained in one-item trials. Thus with load-independent uncertainty, and in the absence of any storage limits, the model produced key qualitative and quantitative features of spatial working memory performance. Solving the correspondence problem naturally becomes more error prone as memory load increases.

Model comparisons

With estimates of uncertainty and guessing derived entirely from single-item trials, my model achieved remarkable agreement with human performance across a range of loads. Among the main implications of this result is that limits on the contents of visual working memory cannot be measured accurately without first accounting for variance explained by the challenges associated with correspondence uncertainty. Yet all prevailing theories posit extreme storage limits of one form or another, and they have measured these limits through models that either do not allow for correspondence errors at all (Bays & Husain, 2008; Wilken & Ma, 2004; van den Berg et al., 2012; Zhang & Luck, 2008), or that characterize 'misbindings' as random events in which all non-targets are equally likely as mistaken sources of responses (Bays, Catalao, & Husain, 2009; Bays, Wu, & Husain, 2011).

To explore if and how storage limits impose constraints on human performance above and beyond the computational challenges associated with correspondences, I developed variations of my model in which representational precision and guessing rates could vary with set size. I fit three additional models to compare with the primary model described above, resulting in four models over all. The primary model employed precision and guessing parameters derived and unchanged from trials with a memory load of one. Since it included no storage limits I call it the '*Unlimited Storage'* model. Additionally, and following prior approaches (Bays & Husain, 2008; Zhang & Luck, 2008), I refit my model independently for each memory load and each participant, thus allowing the precision and guessing parameters to vary across memory loads. Changes in the value of the precision parameter would imply changes in representational precision, and increases in guessing rates would imply discrete capacity limits. Since declines in representational precision and discrete capacity limits both reflect a kind of commodity limit —a limit wherein one object consumes resources that are no longer available for another object— I called this a *'Commodity Limited'* model. This kind of Commodity Limited model has recently been the focus of attention in a rapidly growing literature (Bays, Catalao, & Husain, 2009; Bays et al., 2011; Bays & Husain, 2008; Bays, Wu, & Husain, 2011; Burnett et al., in press; Fougnie & Alvarez, 2011; Gorgoraptis et al., 2011; Van den Berg et al., 2012; Zhang & Luck, 2008, 2009).

Finally, I considered versions of each of these models in which perfect correspondences were assumed. In other words, I considered models in which errors always arose from noisy comparisons between an observed probe and the correct memory representation or as a result of guessing (Bays & Husain, 2008; van den Berg et al., 2012; Wilken & Ma, 2004; Zhang & Luck, 2008). The purpose of this was to demonstrate that including a need for solving correspondences improves a model's fit. Note that I implemented this need without adding free parameters to the respective models, and that any theory that acknowledges noisy representation of object features should require a mechanism by which correspondences are inferred.

Because the Commodity limited models included considerably more free parameters than the Unlimited Storage models, I compared them on the basis of the Bayesian Information Criterion (BIC), a measure that is based on the maximum likelihood estimate (MLE), while taking into account parametric complexity (Myung & Pitt, 1997; Wagenmakers, 2007; van den Berg et al., 2012). Table 3.1 reports the BIC value obtained for each of the four tested models. Note that better models have lower BIC values. Under the column titled 'Correspondences,' a designation of 'Inferred' refers to models that were uncertain about correspondences, dealing with them via probabilistic inference (as in my model), and 'Known' designates models in which perfect correspondences were always assumed.

Comparison between models for Experiment 6

Model	# of Parameters	Correspondences	BIC
Unlimited Storage	24	Inferred	5628.265
Unlimited Storage	24	Known	7267.230
Commodity Limited	144	Inferred	6103.848
Commodity Limited	144	Known	6613.691

Table 3.1. Bayesian Information Criterion (BIC) values for the four models tested against the data of Experiment 6 (N=12). Each model was tested under conditions where correspondences had to be determined on the basis of spatial proximity (Correspondence = Inferred) and under conditions where perfect correspondence between the probe and the relevant target was assumed (Correspondence = Known). The Unlimited Storage models included two free parameters —precision and guessing rate at a load of one— for each of 12 participants, and as a result, these models included 24 free parameters each. The Commodity Limited models included the same two parameters per participant, but the values of the parameters varied for each tested memory load (one to six). Accordingly, those models included a total of 144 free parameters each. Lower BIC values suggest better models.

Evidencing the importance of correspondence problems in limiting human performance, I found that both my original model, and the Commodity Limited model supplied a better fit to the human data when correspondence was established by proximity, compared to when perfect correspondences were assumed. This is of particular importance because one might have been concerned that, in fact, human participants could solve correspondences perfectly in this task, perhaps on the basis of color – the corresponding memory item was always in the same color as a probe. But prior experiments have demonstrated that participants do not use color in spatial working memory experiments like this one (Logie et al., 2011), a result that is supported by my model comparisons. I return to this issue in Experiment 10 and in the Summary and Conclusions. Overall, that model fits improve when correspondences had to be inferred demonstrated that correspondence errors contribute to human spatial working memory errors, in agreement with the *Consistent* vs. *Inconsistent* analysis conducted for Experiment 6.

Surprisingly, the model comparisons that I performed even suggested that correspondence challenges and attendant errors impose the only limit on human performance in this spatial working memory task. The simplest model —one that did not include any storage limits— supplied a better fit to the data than the model with declining precision and increasing guess rates as a function of load. However, all prevailing theories hypothesize that commodity-based limits impose the primary constraint on visual working memory, differing mainly on the nature of the underlying commodity. Is it possible that hypothesized storage limits may not describe a genuine architectural feature of working memory? This may be too strong a conclusion to draw from these experiments, taken alone. But I note that simple models that would support such a conclusion are rarely if ever tested against the more complex models common in the literature. I return to this issue in the Summary and Conclusion section. At a minimum, my model comparisons suggest that, to a considerable extent, spatial working memory errors derive from representational uncertainty —an ineluctable fact of sensory processing— and the concomitant problem of establishing correct correspondences across perceptual episodes.

Experiments 7-9: Facilitating correspondences through preview

The *Inconsistent/Consistent* analysis and modeling results from Experiment 6 suggest that failures of correspondence underlie many spatial working memory errors. It follows that facilitating correct correspondence could result in high levels of performance even for large memory loads. I minimally modified the previous paradigm to test this hypothesis (Figure 3.01, Experiments 7-8). The procedure was nearly identical to Experiment 6. Just as in that experiment, a probe eventually reappeared and participants reported its direction of displacement. Before the presentation of the probe, however, participants were shown a preview display comprising all of the non-probed memory items in their original positions. Because the probed item was not redisplayed, and because participants did not know in advance which item would be probed and which items previewed, the preview did not provide additional direct evidence about the position of the probed item. However, it could provide information about the identity of that item: if participants successfully established a correspondence among the preview items and the memory items, the probe could be identified as the 'odd man out' that did not appear in the preview. Thus I expected that the preview display in Experiments 7 and 8 would bring performance in high load trials close to performance in lower load trials. Experiment 9 was a control wherein I 'previewed' the entire sample to demonstrate that the preview facilitates performance by exclusion, rather than by reactivating a general representation of the scene. I predicted that performance in Experiment 9 —the complete preview experiment—would decline, as it usually does, as a function of memory load.[2]

Methods

Participants. Separate groups of 12, 6, and 6, participants took part in Experiments 6, 7, and 8, respectively. All had normal or corrected-to-normal visual acuity. The experimental protocol was approved by the Johns Hopkins University IRB.

Apparatus. The apparatus was identical to that used in Experiment 6.

Procedures. The overall procedure for these experiments was identical to Experiment 6 except as follows. Memory loads of two, three, four, and six were tested in Experiment 7, and loads of four, five, six, and eight were tested in Experiment 8. Displacements employed were of 0.5º, 2º, or 5º. The sample display duration was increased to 1.5 seconds. Most importantly, 500 ms after a sample display disappeared, a preview display was presented for 1.5 seconds. The preview display consisted of each of the previous sample items presented in their original locations, excluding the to-be-probed item, which was absent entirely. The preview items then disappeared from the display, and following another 500 ms blank, a probe appeared for 0.25 seconds. The

[2] The present notion of 'preview' is different from that used in previous studies (e.g., Kahneman, Treisman & Burkell, 1983; Watson & Humphreys, 1997; Al-Aidrous et al., 2011), which presented a subset of the non-target items *prior* to the presentation of the memory display rather than after it. Both types of preview facilitate VWM performance, and I anticipate that my model could account for the earlier results in a way similar to that discussed below: essentially, by using the preview to (probabilistically) exclude memory items from being placed in correspondence with the probe. For discussion of alternative accounts of the earlier results, see Donk & Theeuwes (2001, 2003).

procedure for Experiment 8 was identical to Experiment 7, except for the fact that sample sizes of 4,6,7, and 8 were tested. For both experiments, participants completed 60 trials for each memory load, which amounted to a total of 240 trials. Experiment 9 was identical to Experiment 7, except that all sample items, including the eventual probe, were present during the preview.

Analysis. In Experiments 7 and 8, I estimated the precision of participants' knowledge about item positions as a function of memory load. Following prior reports (Bays & Husain, 2008), a probit regression model was used to estimate the parameters of the cumulative Gaussian distributions that best fit participants' responses as a function of test item displacement (for each memory load). Precision was estimated as the reciprocal of the estimated standard deviation (1/degrees) of the fitted Gaussian distribution for each memory load.

Results and Discussion

One way ANOVAs revealed no effect of memory load on error rate in either Experiment 7 or 8 (Experiment 7: $F(3,33) < 1$; Experiment 8: $F(3,15) = 1.267$, $p = .321$). To further explore these data, I estimated the precision of participants' knowledge about item positions by using the fact that probes were displaced to varying degrees. This was done by fitting cumulative Gaussian functions to participants' responses and then estimating the standard deviations of these functions for each memory load. Precision was estimated as the reciprocals of the estimated standard deviations (1/degrees), and estimated parameters were regarded as influencing precision if they had a significant effect on the slope term of the fitted regression model. Figure 3.06 displays the fitted response functions for each memory load of Experiment 7 (black lines) and Experiment 8 (red lines). Figure 3.07 displays the estimated precision parameters for each memory load of each of the two experiments.

Increasing memory load had no significant impact on precision in either experiment (Experiment 7: $X^2(3) = 4.2$, $p = .24$; Experiment 8: $X^2(3) = 4.3$, $p = .23$). Comparisons between loads of two and six (Experiment 7: $z = -.061$ $p > .5$) and between loads of four and eight (Experiment 8: $z = -.144$ $p > .5$) were not significant, despite a doubling or tripling of the number of memory items. I also note that precision was unchanging throughout my tested range, for loads both above and below hypothesized fixed capacity limits of about four (Anderson et al., 2011; Luck & Vogel, 1997; Zhang & Luck, 2008). When the potential for correspondence errors was alleviated via my preview displays, performance did not decline with increasing memory load.

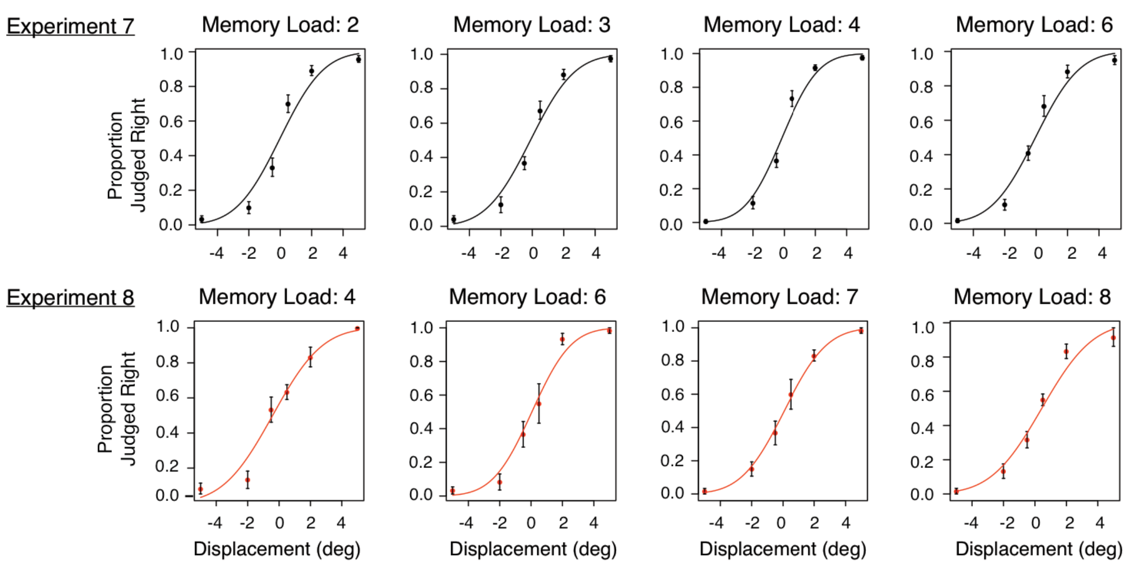

Figure 3.06. Proportion of trials judged rightward as a function of displacement magnitude and memory load in the preview experiments, Experiment 7 (top row) and Experiment 8 (bottom row), along with fitted cumulative Gaussian functions. Error bars reflect ± 1 SE.

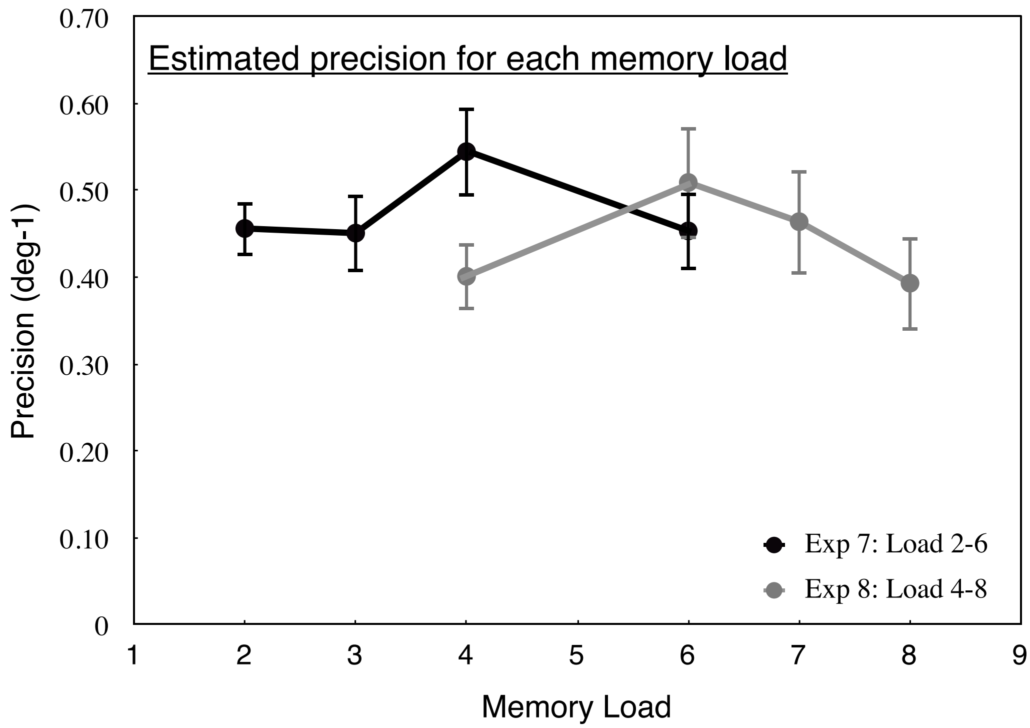

Figure 3.07**.** Estimated precision (± 1 SE) by memory load in the preview experiments (Experiment 7 in black, Experiment 8 in red). Representational precision, estimated as the reciprocal of the SD of best-fit cumulative Gaussian functions, did not decline as memory load increased. Experiment 7 (N=12) included memory loads of two, three, four and six; Experiment 8 (N=6) included loads of four, six, seven, and eight.

One might wonder whether the preview display could provide indirect information about the position of a probe item, perhaps through cues about configuration or other encodings of spatial relations among memory items (Jiang, Olson, & Chung, 2000: Thiele et al., 2011). Below I report modeling results that weigh against such a hypothesis, by showing that the model supplies an excellent fit to the data without recourse to such cues. Moreover, Experiment 9 was conducted to address exactly such a possibility. In Experiment 9, all items were previewed, including the probe. This complete preview should have provided exhaustive information about any stored configurations, but no additional information to aid in correspondence matching. As in Experiment 6, and unlike in Experiment 7, error rates increased significantly.

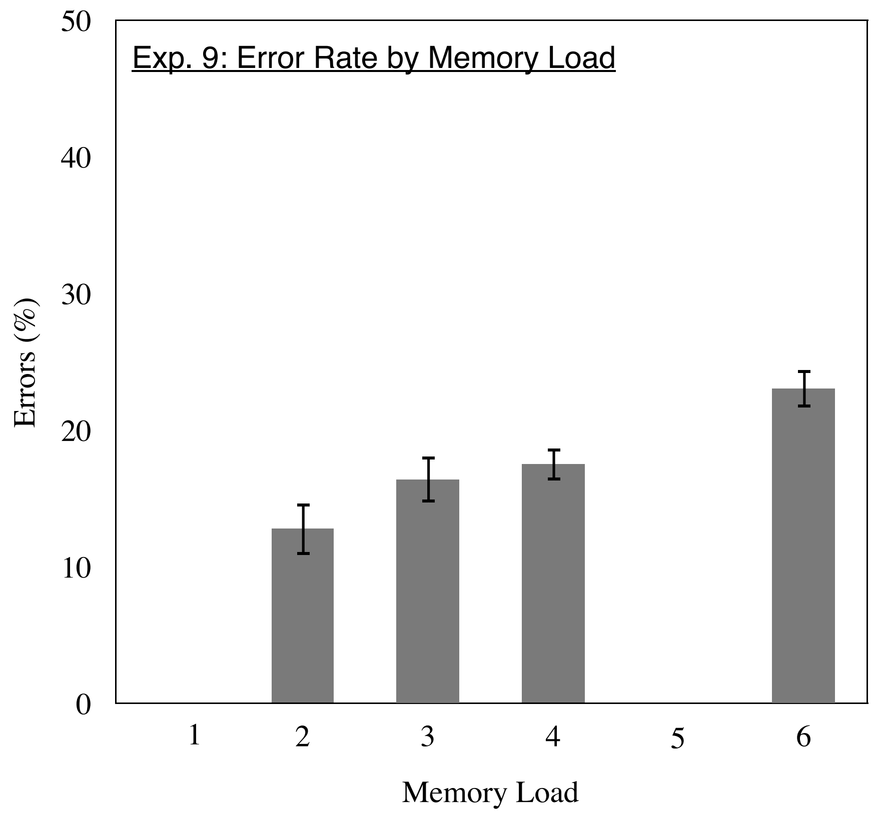

Figure 3.08. Error rates (± 1 SE) by memory load in Experiment 9 (N =6). Mean error rate increased as memory load (the number of sample items) increased.

Modeling Experiment 7

I minimally modified my Unlimited Storage model from Experiment 6 to model the results of Experiment 7. Methodologically, these two experiments were identical, except for the fact that in Experiment 7, a preview display appeared just before test and it included all the sample items save for the eventual probe. Therefore, my model computed correspondences between sample items and preview items during the preview display, leaving a single sample representation for subsequent comparison with the probe at the time of test. In other words, when the probe appeared at test, a correspondence ‘decision’ had already been made, and all that was required was a comparison between the remaining sample item representation and the representation of the probe.

To compute the correspondences between sample items and preview items, the model calculated the distance between the position of each stored sample item (of which there were as many as the memory load) and each of the observed preview item positions (of which there were as many as the memory load, minus one), and it then selected the most probable joint set of correspondences. Crucially, this procedure meant that even with the preview displays, the model (and human participants) would sometimes make correspondence errors, selecting a wrong sample item for subsequent comparison to the probe.

Since Experiment 7 did not include trials with a memory load of one (because there would be nothing to preview in those trials), I could not estimate the participant-specific value of a lapse parameter and a spatial precision parameter from those trials. In place of these estimates, I substituted the median values across all participants from load one trials of Experiment 6.

The mathematical details for this model were identical to the previous model, except with respect to how correspondence was established. Because the preview items (unlike the probe) were known to share the same horizontal and vertical coordinates as the original memory items, an ideal observer would use both distance dimensions in establishing correspondence. And because there could be multiple preview items (one fewer than the number of memory items), correspondences were determined by jointly minimizing the distance among corresponding items, technically, minimizing the following matching function (where

μ is an injective mapping from preview items to memory items):

$$C(\mu) = \sum_{i=2}^{n} N\left(z_i'; z_{\mu(i)}, 2\sigma^2 I\right) \quad [6]$$

Here z_i' was the sampled position of i in the preview display, and $\mu(i) \epsilon \{1, \dots, n\}$. Each term in this sum has the same form as the numerator of [4], except that Gaussian density is two-dimensional rather than one (since this model used both vertical and horizontal coordinates to infer correspondences). The derivation of this matching function for Bayes' Theorem was a simple extension of the derivation of [4] with the additional constraint that there were multiple, mutually exclusive elements of the correspondence. Given the least-cost mapping of preview items to memory items, the target was identified as the memory item that was not in the range of μ, and then compared to the probe as in the model for Experiment 6.

When modified so that correspondence was solved during the preview display — leaving a single memory representation for comparison to the probe — the Unlimited Storage model provided an excellent fit to the results of Experiment 7 (r = .91). As I did for Experiment 6, I also considered a version of the Unlimited Storage model wherein perfect correspondences between the probe and the tested memory item were assumed. (This model did not perform any additional steps during the preview display, since it was already assumed that the right correspondence would be known at test). The model that was uncertain about correspondences supplied a better fit to the data (Table 3.2).

Similarly, I considered a Commodity Limited model, both a version with perfect correspondences assumed and a version that solved correspondences using the preview display. (For this model, cost function [6] was also employed, but parameter values were fit for each participant at each memory load). Again, the correspondence uncertain model supplied a better fit between these two Commodity Limited models, and as before, the correspondence uncertain Unlimited Storage model supplied the best fit among all the models as measured by BIC. Taken together the modeling results for Experiment 7 and the empirical results of Experiments 7-9 demonstrate how correspondence challenges arise naturally from sensory uncertainty about spatial position, and how these challenges impose computational limits on human performance independent from any storage limits.

Comparison between models for Experiment 7

Model	# of Parameters	Correspondences	BIC
Unlimited Storage	0	Inferred	2772.615
Unlimited Storage	0	Known	2788.703
Commodity Limited	96	Inferred	2783.493
Commodity Limited	96	Known	2798.735

Table 3.2. Bayesian Information Criterion (BIC) values for the four models tested against the data of Experiment 7 (N=12). Each model was tested under conditions where correspondences had to be determined on the basis of spatial proximity (Correspondence = Inferred) and under conditions where perfect correspondence between the probe and the relevant target was assumed (Correspondence = Known). The Unlimited Storage models included 0 free parameters because the precision and guessing rate parameters adopted were the median values obtained from participants in the load one trials of Experiment 6. The Commodity Limited models included two parameters per participant —guessing rate and precision— and the values of these parameters varied for each tested memory load (2,3,4 & 6). As a result, those models included a total of 96 free parameters each. Lower BIC values suggest preferred models.

Experiment 10: Do participants use color to solve the correspondence problem?

The results of the preceding four experiments, and the models presented, suggested that human observers face a correspondence problem when using spatial working memory, and that they contend with this challenge on the basis of inferences over spatial positions. But one nagging issue is that in these experiments the colors of the objects could have provided a reliable means for solving correspondences. The probe always matched the color of the corresponding memory item. But participants did not appear to systematically use color. If they had, it is not clear why *Inconsistent* trials would have produced so severe an impact on performance (in Experiment 6), or why my correspondence uncertain models would have provided systematically better fits than the models that assumed perfect correspondences. These results were more generally consistent with the fact that human observers often appear to prioritize spatial features relative to surface properties when making judgments about object correspondences (Flombaum et al., 2009), and also with the well-known 'binding problem' between color and position (Treisman & Schmidt, 1982), which I discuss more extensively in the Summary and Conclusions section.

Nonetheless, Experiment 10 was designed to directly test whether participants used color to inform their correspondence inferences. The experiment included two groups of participants in a between-subjects design. Half the participants did the same task as the one used in Experiment 6 (with memory loads of 1, 2, 4, and 6 tested). For the other half of participants, memory displays were identical to those used in Experiment 6, but each probe appeared in an entirely new color, one that was not at all present in the relevant memory display. My intent was to compare performance between participants in the two conditions, with the expectation that participants who observed color-changing probes would perform more poorly if they generally use color as a means of solving the correspondence problem. In contrast, if they do not or cannot generally use color then I expected comparable performance in the two groups. I ran the experiment using a between-subjects design so that I could compare performance in participants who saw color-changing probes to the performance of participants who should have legitimately expected color to be a stable and reliable cue for correspondences.

Methods

Participants. 20 participants were included in this experiment

Apparatus. The apparatus was identical to that used in Experiment 6.

Procedures. The overall procedure for this experiment was identical to Experiment 6, except as follows. Memory loads of one, two, four, and six were tested at displacements of 0.5º, 2º, and 5º. There were two conditions, run between subjects. The *Same-Color* condition was identical to Experiment 6, and it included 10 participants. The *Different-Color* condition also included 10 participants, and it deviated from Experiment 6 in only one way: the probe appeared in a color not present at all in the memory display of a given trial. In both conditions, as in Experiment 6, participants judged whether a probe moved left or right relative to the position of the corresponding memory item.

Results and Discussion

A three-way ANOVA revealed significant main effects of memory load ($F(3,54) = 23.158$, $p < .001$) and spatial disparity ($F(2,36) = 262$, $p <.0001$) on accuracy, as well as an interaction between them ($F(6,108) = 2.265$, $p = .042$). However, a main effect of *Same-Color/Different-Color* condition was not significant ($F(1,18) = 1.772$, $p = .2$).

Similarly, there was no significant two-way interaction between memory load and condition ($F<1$) or between displacement and condition ($F(2,36) = 1.984$, $p = .152$). The three-way interaction between memory load, spatial disparity, and condition, also was not significant ($F<1$). Out of concern that these results were driven primarily by the inclusion of a memory load of one (which was included to reinforce, for participants in the *Same-Color* condition, that color was a reliable cue), I conducted planned pair-wise comparisons between performance in the *Same-Color* and *Different-Color* conditions at each memory load greater than one. None of these comparisons were significant (smallest $p = .17$). Thus performance was comparable when color could have been used to solve correspondence problems and when it could not have been, suggesting that, overall, color information was not used by participants as a means to addressing the correspondence problem.

SWM Summary and Conclusions

The vast majority of research on visual working memory has made inferences about the contents of human memory from task performance. I sought to explore a neglected aspect of even the simplest working memory task, the need for a computational procedure that identifies which items in memory correspond with which new observations. I focused on spatial working memory as a case study in which I could directly observe the impact of correspondence problem failures, and where I could formalize a potential algorithm to address the correspondence challenge.

Experiment 6 evidenced the impact of correspondence problems on human memory performance via an analysis of *Consistent* compared with *Inconsistent* trials, trials in which correspondence failures were less likely to lead to incorrect responses compared to trials where they were more likely to do so. Experiments 7 and 8 further demonstrated the importance of solving correspondence problems by utilizing a novel preview manipulation that allowed participants to more effectively solve the correspondence problem prior to making a final, task-relevant judgment. Surprisingly, in these experiments, I found that the preview manipulation not only improved performance at larger memory loads, it also directly equalized performance for larger loads with performance at smaller loads. Experiments 9 and 10 were controls, respectively demonstrating that my preview did not just potentiate all representations, and that participants did not use color as a cue to correspondences.

I also developed a Bayesian model that performs the tasks in both Experiment 6 and 7 and that addressed the correspondence problem through a basic proximity-based computation over uncertain positional inputs. The model provided an excellent fit to human data, and it did so most effectively at all memory loads when utilizing only the representational imprecision associated with remembering a single object. Thus my modeling implies that spatial working memory is limited mainly by the uncertainty associated with perceiving a single object's position, perhaps without any additional storage constraints of either a slot-like (Anderson et al., 2011; Awh et al., 2007; Barton et al., 2009; Cowan, 2000; Fukuda et al., 2010; Luck & Vogel, 1997; Rouder et al., 2008; Zhang & Luck, 2008) or flexible nature (Bays et al., 2009; Bays & Husain, 2008; van den Berg et al., 2012; Wilken & Ma, 2004).

Why is there correspondence uncertainty?

A motivating assumption behind the experiments reported is that performing any working memory task requires computations that address the correspondence problem.

This assumption follows from first principles. Just as the presence of objects and their features must be inferred from impoverished retinal inputs (Marr, 1982), the relationships between inputs separated in time must also be inferred. From a probabilistic point-of-view, these inferences suffer from the same challenges that plague any comparative or decision-making process. Several recent and influential studies have pointed out that probabilistic representations lead to a signal-detection problem when comparing two objects to determine whether they have the same features (Bays & Husain, 2008; Wilken & Ma, 2004). Even when objects are genuinely identical, they should sometimes be judged as different; and even when they are really different, they should sometimes be judged as the same. Similarly, the probabilistic nature of an observer's knowledge entails that observers cannot know with certainty which objects to compare with which, and as I have demonstrated here, that decisions about which to compare will end erroneously some of the time. Indeed, the correspondence problem and the more commonly acknowledged signal-detection problem are both challenges to identifying objects as 'the same.' The signal-detection problem deals with sameness along feature dimensions —do two objects share the same features? While the correspondence problem deals with a broader sense of sameness, sameness in identity over time— which object observed now is the same object observed previously (Flombaum et al., 2009; Yi, et al., 2008)?

While the correspondence problem has not played a prominent role in research on visual working memory, it has been influential in other domains of visual cognition and perception. It serves as an explanation for many apparent motion phenomena, wherein correspondences are inferred between disparate visual events producing the impression of motion where it never really took place (Anstis, 1980); it also furnishes an explanation for the tunnel effect, wherein inferred correspondences between object encounters maintain a sense of persisting objecthood over time, despite episodes of complete occlusion (Burke, 1959; Flombaum & Scholl, 2006); and it is the motivation behind the influential Object File theory of Kahneman, Triesman, and Gibbs (1992), which proposes mechanisms by which correspondences are inferred over time. It also appears perennially as an unavoidable challenge to machine vision systems (Belhumeur & Mumford, 1992; Kahn, Balch, & Dellart, 2005; Ogale & Aloimonos, 2005; Smith, Gatica-Perez, & Odobez, 2005; Szeliski, 2010; Zhao & Nevatia, 2004).

One feature of these examples that appears to be shared with the experiments reported here is the apparent discounting of surface properties, such as color, in favor of spatial properties, such as proximity, when solving correspondences. This phenomenon is sometimes called 'spatiotemporal priority' in related literatures (for a review, see Flombaum et al., 2009). Experiment 10 demonstrated that participants did not use color to solve correspondences by showing that performance did not degrade when the task was designed so that color could not be used. The *Inconsistent* vs. *Consistent* analysis of Experiment 6, the preview results of Experiments 6 and 7, and the modeling results, throughout, all reinforced this point, since none of them should have turned out as they did if participants used color.

In part, participants may not have used color in this task because color memory was not directly probed, and therefore, it perhaps seemed irrelevant. More generally observers may not rely on color because color representations are also probabilistic (Wilken & Ma, 2004; Zhang & Luck, 2008), and they may serve just as poorly, or perhaps worse, as cues to correspondences. Moreover, color information may be weakly bound to individuals and locations (Triesman & Schmidt, 1982), rendering it relatively ineffective as a means for correspondence. Prior works has demonstrated that it can be very difficult to detect when colors swap positions in a visual working memory task, even when the specific colors present are remembered successfully (Wheeler & Treisman, 2002). With respect to the primary purpose of this research project, the fact that participants did not use color further suggests that correspondences may be addressed *in the same ways* by the visual system under contexts that experimenters intuitively treat differently.

Implications for commodity limited models of visual working memory

The purpose of this study was to explore the extent to which correspondence challenges constrain the ability to use spatial working memory. The results surprisingly suggested that correspondence problems not only constrain spatial working memory, but also that they may be the most significant factor constraining it. Two pieces of evidence led to this conclusion. The first came from my modeling. I built a model of human performance in Experiment 6 that did not include storage limits. I did this in order to determine how well an unencumbered observer could perform my task, setting a sort of ceiling on human performance before any storage limits become relevant. But this Unlimited Storage model supplied a better fit to human responses than the Commodity Limited model I eventually tested. Model comparisons suggested no benefit for allowing precision and guessing rates to change as a function of memory load, changes typically interpreted as reflecting storage limits. A second piece of evidence buoyed this conclusion. Specifically, in Experiments 7 and 8, my preview displays not only improved performance at larger loads; they made performance indistinguishable at all tested loads.

When I fit psychophysical functions to the responses in these experiments, they were indistinguishable at all memory loads. I tested wide ranges of memory loads (one to eight), and wide ranges of displacement. If storage limits substantially constrain spatial working memory, then they should have had an impact in my experiments.

That they did not conflicts with prevailing theories of spatial working memory (Dent & Smyth, 2006; Latcher & Hayhoe, 1995), and visual working memory more generally (for review, see Brady, Konkle, & Alvarez, 2011). This is because prevailing theories suggest that the potential contents of visual working memory are limited by a commodity — something that becomes depleted as increasing numbers of objects are encoded. A long-standing and vigorous debate has focused on what the commodity might be, how it is allocated, and whether it becomes exhausted completely or only dwindles to the point of producing very poor representations. Broadly, two kinds of theories dominate this debate. Fixed resource theories have argued that the relevant commodity allocates discretely, and supports only about three representations simultaneously (Cowan, 2000; Luck & Vogel, 1997; Rouder et al., 2008; Zhang & Luck, 2008). In contrast, Flexible resource theories suggest that a continuously allocated commodity determines the precision of object representations, with the capacity to represent more than three objects, but somewhat poorly (Bays et al., 2009; Bays & Husain, 2008; van den Berg et al., 2012; Wilken & Ma, 2004). While hybrid theories have also been proposed (Alvarez & Cavanagh, 2004; Anderson et al., 2011), all prevailing models include some kind of commodity-based constraints on the contents of working memory. I found no need to include any such constraints to explain human performance in my experiments. Instead, the challenge of identifying correspondences appeared sufficient to account for human errors.

This conclusion may seem rather surprising in light of decades of visual working memory research and the dominance of commodity-limited theories beginning with Sperling (1960). But none of this work has attributed any task difficulty to correspondence computations. I expect that building models to include these necessary computational steps will prove productive for elucidating the contents of working memory storage in other contexts. Indeed, Chapter 4 presents some new research that makes it clear that correspondence computations play a role in a variety of working memory contexts.

Chapter 4. Correspondence Computations in Visual Working Memory

(The results and experiments in this chapter were previously reported in Bae & Flombaum (2013). Two items remembered as precisely as one: How integral features can improve visual working memory. Psychological Science, 24, 2038-2047.)

Although there are severe debates on whether VWM is limited by the discrete upper bound defined by the number of objects or not, there appears to be consensus around the idea that within hypothesized bounds, VWM is continuously limited (Brady, Konkle, & Alveraz, 2011; Fukuda, Awh, & Vogel, 2010). Indeed, the severity of VWM limits are perhaps best reflected by the fact that even remembering just two things is harder than remembering one. Across a wide array of stimuli and experiments, the precision of each of two items simultaneously held in memory appears lower than the precision of a single item (Anderson, Vogel, & Awh, 2011; Alveraz & Cavanagh, 2004; Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Wilken & Ma, 2004; Zhang & Luck, 2008).

In this section, I challenge the consensus opinion that declines in precision for two items result from competition for a limited memory commodity. Instead I observe that a specific computational challenge plagues memory for two items but not for one, namely, a correspondence challenge. In order to use a memory to make a judgment about a presently viewed object (e.g. Luck & Vogel, 1997; Bays & Husain, 2008), an observer must determine which stored memories correspond to the observed object. Similarly, in order to use a memory to report the features of a cued object (e.g. Zhang & Luck, 2008; van den Berg et al., 2012), an observer must determine which stored memory a cue corresponds with. But if only one object is stored, then no such computation is necessary. I suggest that, in fact, the precision of two objects in memory is no worse than the precision of one. Instead, in experiments with two objects, participants make correspondence errors that have been interpreted as reflecting declining precision.

If correspondence errors account for apparent costs associated with remembering two items, then costs should be eliminated when correspondence errors are prevented (Bae et al., under review; Bae & Flombaum, 2012). To test this prediction, I sought a method that would support observers when making correspondence decisions. Logically, observers should use 'task irrelevant' features of objects to make correspondence decisions, since these are the features that usually do not change between a memory display and a test display. Correspondence decisions should be anchored to stable features of the world. But in most experiments, task irrelevant features are either identical among all memory items (e.g. identical shapes when color is the target) —providing no basis for correspondence decisions— or they are different along separable dimensions (e.g. shape and color)— providing a problematic basis because of known feature binding challenges (Treisman & Schmidt, 1982). Thus our goal was to design memory displays in which sample items differed along a stable feature dimension that could provide a sound basis for correspondence judgments.

To achieve this end, I found inspiration in prior research focusing on a distinction between integral and separable feature combinations (Garner, 1974; Garner & Felfoldy, 1970). Integral features are those that can be manipulated independently, but are not perceived independently (e.g. size and shape). Separable features (e.g. size and color) are represented independently, and binding between them is not always successful (Treisman & Schmidt, 1982). I hypothesized that distinguishing items via separable features (as is often done) leaves the correspondence problem in place. For example, if asked to remember the sizes of two triangles that also differ in color (see Experiment 14), the presence of a color at test will not necessarily help with the correspondence problem; the problem of determining which remembered size corresponds to e.g., ‘the blue one,’ is no less difficult than determining which remembered size corresponds to ‘the one over there.’ In contrast, a uniquely identifying integral feature should help. In a size memory experiment, when probed with a circle, for example, one would naturally query only the sizes of remembered circles, ignoring remembered triangles. Including only a single circle in a memory set of two could, in turn, largely prevent correspondence errors. By this logic, I hypothesized that integral feature distinctions would erase costs associated with remembering two compared to one.

I tested this hypothesis in memory experiments for three different features: the luminance of two objects when they possessed the same hue (e.g. both were a version of red) or different hues (e.g. red and green); the sizes of two objects when they were either the same shape or different; and in an auditory experiment, the amplitudes of two tones when they were either the same frequency or different frequencies. I predicted equal performance for two items compared to one when the two items in memory comprised different integral feature values. Additionally, to validate the specific integral features employed, I conducted a conceptual replication of classic sorting experiments by Garner and Felfoldy (1970).

Experiments 11-12: Sorting integral features

To independently identify integral feature dimensions in vision we conceptually replicated a now classic experiment by Garner and Felfoldy (1970). In Experiment 11, three colored squares with different luminance values were presented simultaneously and participants were asked to find the brightest and second brightest ones (Figure 4.01). Crucially, the three colored squares could be from the same hue family or from different hue families. I predicted that performance with items sharing a hue would be better than with items from different hue families because hue and luminance are integral features. This was the logic underlying Garner's experiments. Experiment 12 applied the same logic to size and shape. Participants were instructed to identify the smallest and then the second smallest item, by area, in a set of three. The items were either all the same shape, or they were different. I expected worse performance in the sets comprising different shapes.

Method

Participants. Two groups of 10 Johns Hopkins undergraduates participated in exchange for course credit. One group participated in the luminance sorting experiment, and the other participated in the size sorting experiment. All had normal or corrected-to-normal visual acuity. The protocols for these and all other reported experiments were approved by the Johns Hopkins University IRB.

Apparatus & Stimuli. Participants sat 60 cm from an iMAC computer such that the display subtended approximately 39.56° by 25.35°. Stimuli used in the experiments are shown in Figure 1.01. The lights were turned off during testing.

Procedure. On each luminance sorting trial, 3 colored squares were picked from a set of 51 possibilities (Figure 4.01A). On half of trials, all the colors presented were from the same hue family. In the other half, each item was a different hue. In all trials, the three squares varied in luminance via the L coordinate in L*a*b space. They were positioned horizontally in a random order. The task was to pick the brightest and then second brightest by pressing numerical keys corresponding to their positions on the screen.

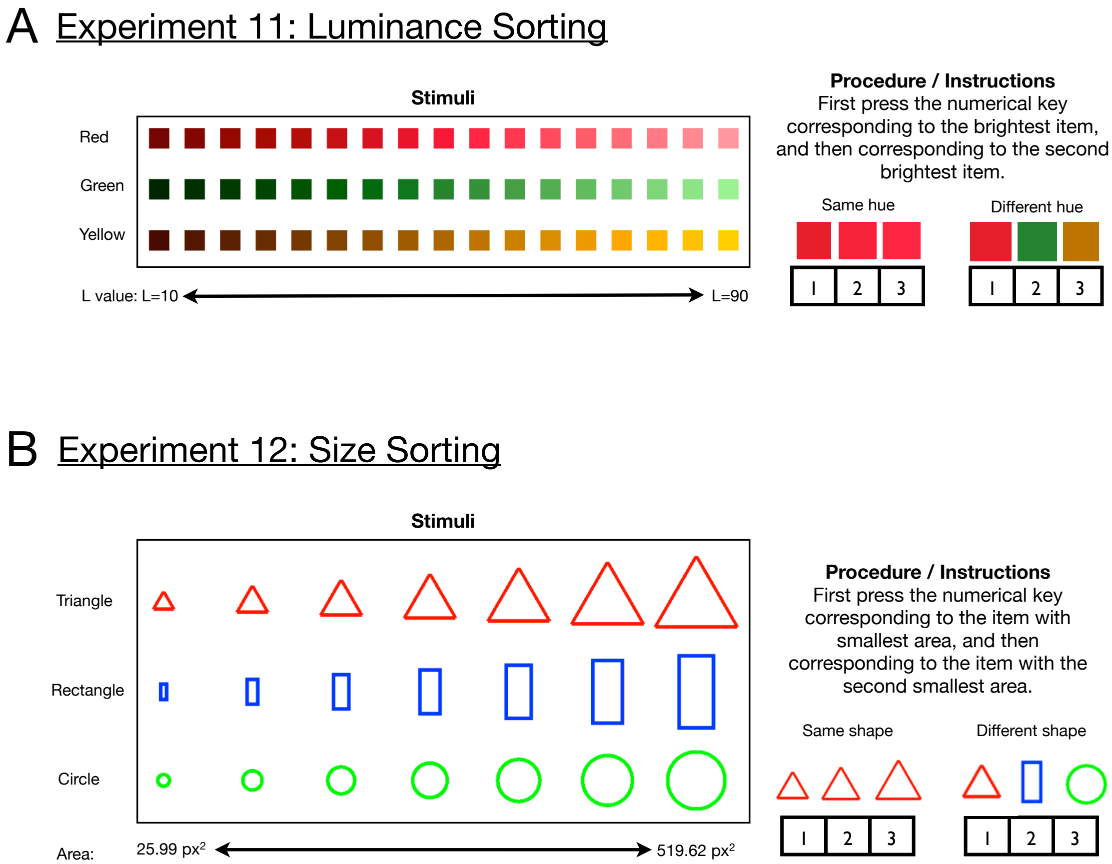

Figure 4.01. Experimental design for the luminance (a) and size (b) sorting experiments. a. A set of 51 colored chips were generated with 17 possible luminance values in CIE L*a*b color space (L values between 10 and 90, in steps of 5). Hue families were defined by a and b of L*a*b coordinates (Red: a = 80, b = 38; Green: a = -40, b = 38; Yellow: a = 20, b = 98). Within a column, stimuli shared the same L (luminance) value. On each trial, from this set of 51 possible colors, 3 were selected and presented in a random order horizontally at the center of a black screen. In half of trials, colors were selected from the same hue family (e.g. reds in the case shown), and the other half involved an item from each of the 3 hue families. Regardless of the composition of the stimulus set, the task was to sort the presented items from brightest to darkest by pressing keys corresponding to the numbers under each item. (Numbers shown in the figure were not presented in the display; but participants learned during practice that 1 through 3 designated left to right). b. Identical procedures were employed in the size sorting experiment, except here, three shapes were used. The task was to sort stimuli from the one occupying the smallest area to the one occupying the largest area. Objects in each column occupied equal area.

An identical procedure was employed in the size sorting experiment (Figure 4.01B). Three objects with either the same shape or different shapes were presented. The task was to pick the smallest and then the second smallest item by area. Each experiment included 120 trials and 10 practice trials.

Results and Discussion

Proportions of accurately sorted trials are shown in Figure 4.02. Sorting performance was worse when the items were from different hue families, $t(9) = 8.90$, $p < .05$, $d = 3.04$, or differently shaped, $t(9) = 12.48$, $p < .05$ $d = 3.73$.

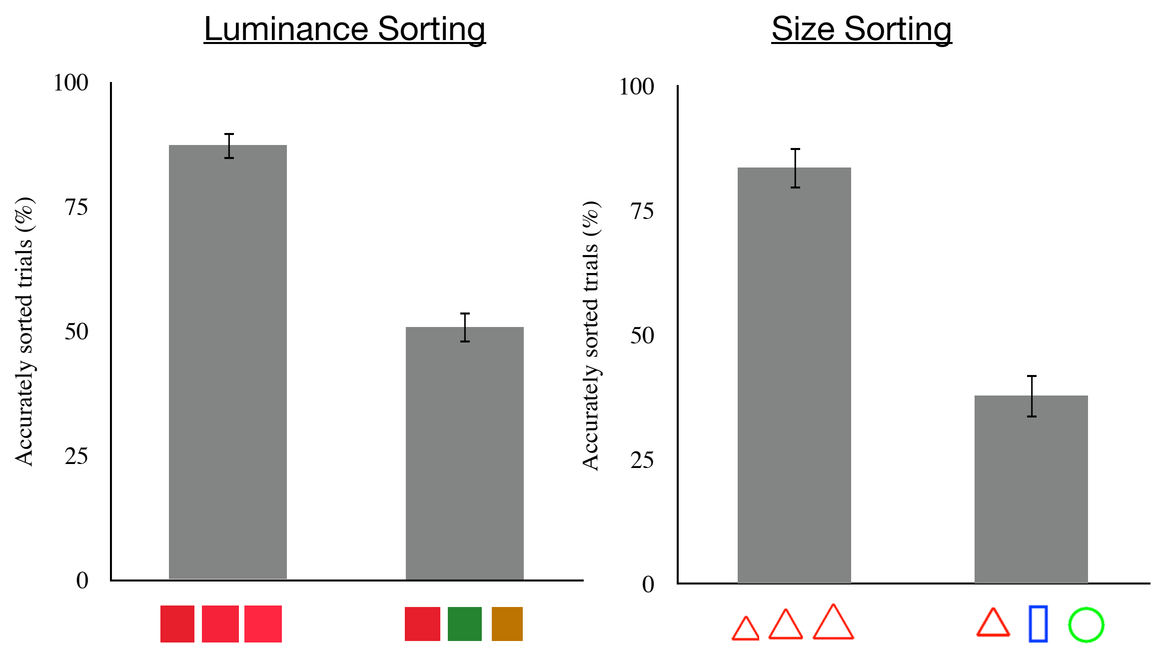

Figure 4.02. Accuracy of Luminance sorting and Size sorting experiments. Accuracy was measured as the percentage of trials correctly sorted. Error bars reflect ±1 S.E. of the mean.

These results evidence what is perhaps a cardinal rule of perception: that it is often more than the sum of its parts. Features that are independently controllable may not combine independently. For current purposes, the consequent point is that it can be challenging to compare objects along one dimension when they are different along another, seemingly irrelevant, but integral dimension. We exploit this fact in the forthcoming memory experiments.

Experiments 13-15: Preventing memory errors with integral features

Armed with specific knowledge about dimensions that are integral, and with the knowledge that differences in one integral dimension can render challenging comparisons along another, we sought to prevent correspondence problems in a VWM experiment. The logic was as follows. We employed a basic feature-judgment task (e.g. Bays & Husain, 2008). An observer is instructed to remember, for example, the luminance of either one item or two. At test, a single probe item appears, and the observer must compare its luminance to that of the corresponding memory item. In two item trials, if the items share the same hue family, the main point is that a correspondence error can take place. The observer may compare the probe to the wrong memory item, potentially producing an erroneous response (and worse performance in two item trials compared to one). In contrast, what might happen if the two objects differ along an integral dimension? Though observers may generally make correspondence errors, we know from Experiments 11 and 12 that mistaken correspondences should, in this situation, lead to difficult and unintuitive comparisons: e.g. which is brighter, the green one I see or the red one I remember? The unintuitive nature of the comparison could potentially lead an observer to realize (not necessarily explicitly), that she has made a correspondence mistake, and the observer may then address the task comparison to the correct memory item instead.

Thus I predicted that two items would be remembered as precisely as one when the two items differed along an 'irrelevant,' but integral feature dimension. When they shared the same feature value along that dimension, however, we predicted typical costs for two compared to one. Based on the results of Experiments 11 and 12, I tested these predictions in the cases of memory for luminance (Experiment 13), with objects that differed (or not) in hue, and in memory for size (Experiment 14), with objects that differed (or not) in shape. In addition, I extended our manipulation to auditory working memory (Experiment 15) using a known set of integral auditory features, amplitude and frequency (Garner, 1976; Wood, 1975).

<u>Method</u>

Participants. A new group Johns Hopkins undergraduates participated, 12 in Experiment 13, 8 in Experiment1 4, and 12 in Experiment 15. All participants had normal or corrected-to-normal vision and audition.

Apparatus, stimuli, & procedure. In Experiment 13 and 14, apparatus and stimuli were identical to those used in the sorting experiment. In the luminance experiment (Experiment 13, Figure 4.03A), one or two squares were presented in random positions around a central fixation. Each square was assigned one from 17 possible luminance values. Additionally, each square was assigned a hue (i.e. it was drawn from one of the 3 rows in Figure 4.01A). There were two kinds of two item trials. In the *Same Integral Feature (*SIF*)* trials*,* both squares possessed the same hue. In *Different Integral Feature (*DIF*)* trials*,* they each possessed a different hue. Participants were asked to memorize the luminance of each item. After a brief mask display, a probe item appeared in the same position with the same hue as one of the corresponding memory items. Participants judged whether the probe was darker or brighter than its correspondent. A probe differed from its corresponding memory item by ± 5, 15, or 25 L value steps. There were 432 trials plus 10 practice trials.

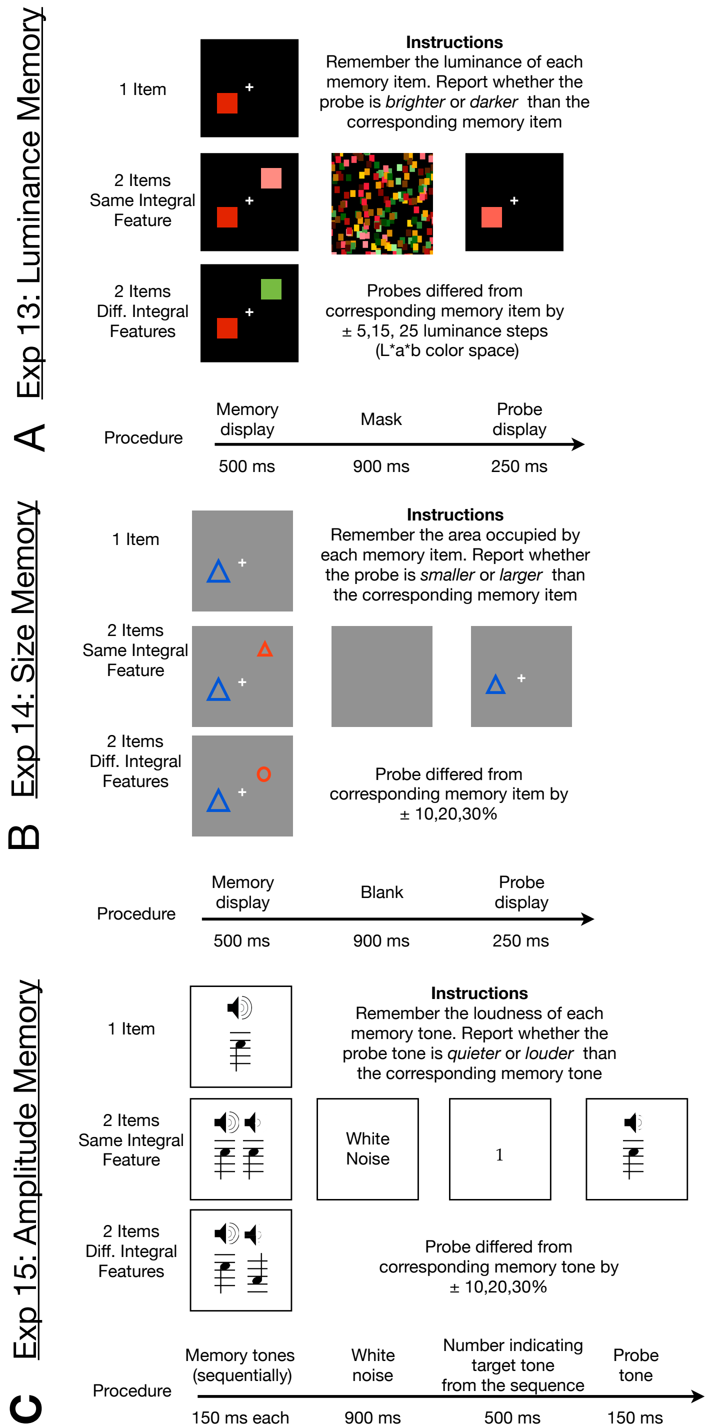

Figure 4.03. Experimental procedures. a. Luminance judgment task. One or two items from stimuli set used in the luminance sorting experiment were used as sample items. b. Size judgment task. One or two items from the stimuli set in Experiment 12 (size sorting) were used. c. Amplitude judgment task. Amplitudes were normalized to a range of 0 to 1. One or two tones with amplitude ranging between 0.4 and 0.7 were presented sequentially followed by white noise The probe tone was 10, 20, or 30 % louder or softer than its correspondent.

The size experiment (Experiment 14, Figure 4.03B) was nearly identical. Memory objects were circles and triangles. SIF trials included two objects with the same shape. DIF trials included two different shapes. The objects were always different colors (i.e. red and blue). One item trials included either a circle or a triangle. The task was to remember the area occupied by each item. At test, a probe item appeared at the same position and with the same shape and color as the corresponding memory item. The size of a probe was ± 10, 20, or 30% relative to its correspondent. Participants reported whether the probe was larger or smaller. There were 216 trials plus 10 practice trials.

In Experiment 15 (Figure 4.03C), auditory stimuli were delivered in stereo through headphones. On each trial, one or two pure tones of either 220 Hz or 400 Hz were played. In two item trials, the tones were separated by a 500 ms blank screen and a 150 ms burst of white noise. The amplitudes of the tones were chosen randomly from the range of 57.1 dB to 64.6 dB, with a minimum difference of 1.25 dB. Participants were instructed to remember the ‘loudness’ of each tone. After a second burst of white noise (in two item trials), a 1 or 2 was displayed indicating which tone from the sequence to compare with the upcoming probe. The probe tone always had the same frequency as its correspondent, but varied in amplitude by ± 10, 20, or 30%. Participants judged whether a probe’s amplitude was softer or louder than its correspondent. In SIF trials, the two memory tones had the same frequency and in DIF trials, they had different frequencies. There were 216 trials plus 10 practice trials.

Analysis. In order to compare the quality of memory representations, I fit psychometric functions with a probit regression model in each memory load and condition (Figure 4.03, the first column). The reciprocal of the standard deviation of these functions (the second column, Figure 4.03) reflects the precision of memory representations (Bays & Husain, 2008; Palmer, 1990).

Results and Discussion

Figure 4.04 graphs the response functions obtained from each experimental condition, and estimates of representational precision derived from these functions. There were significant effects of condition in each experiment: luminance, $x^2(2) = 15.8$, $p < .001$, $d = 3.97$; size, $x^2(2) = 26.4$, $p < .001$, $d = 5.13$; amplitude, $x^2(2) = 8.0$, $p = .018$, $d = 2.83$. Planned comparisons explored differences between one item trials and the two kinds of two item trials. A typical cost arose for remembering two items compared to one when the two items shared an integral feature value (SIF). Statistically, these cost are seen in the significant difference between the estimated standard deviations of the response functions for the relevant experimental conditions: luminance, $z = -3.892$; size $z = -4.357$; amplitude $z = -2.478$, ($p < .05$ for all). Note that in the size experiment costs were present despite the fact that the two shapes always differed in color, a separable feature dimension from shape.

But costs were eliminated for two items compared to one when the two items included different integral features (DIF): luminance $z = -1.62$; size $z = 0.32$; $z = -0.165$; ($p > .1$ for all). Representational precision was also significantly better in DIF trials compared to SIF trials: luminance, $x^2(1) = 5.5$, $p = .019$, $d = 2.34$; size $x^2(1) = 16.6$, $p < .001$, $d = 4.07$; amplitude $x^2(1) = 5.4$, $p = .02$, $d = 2.32$.

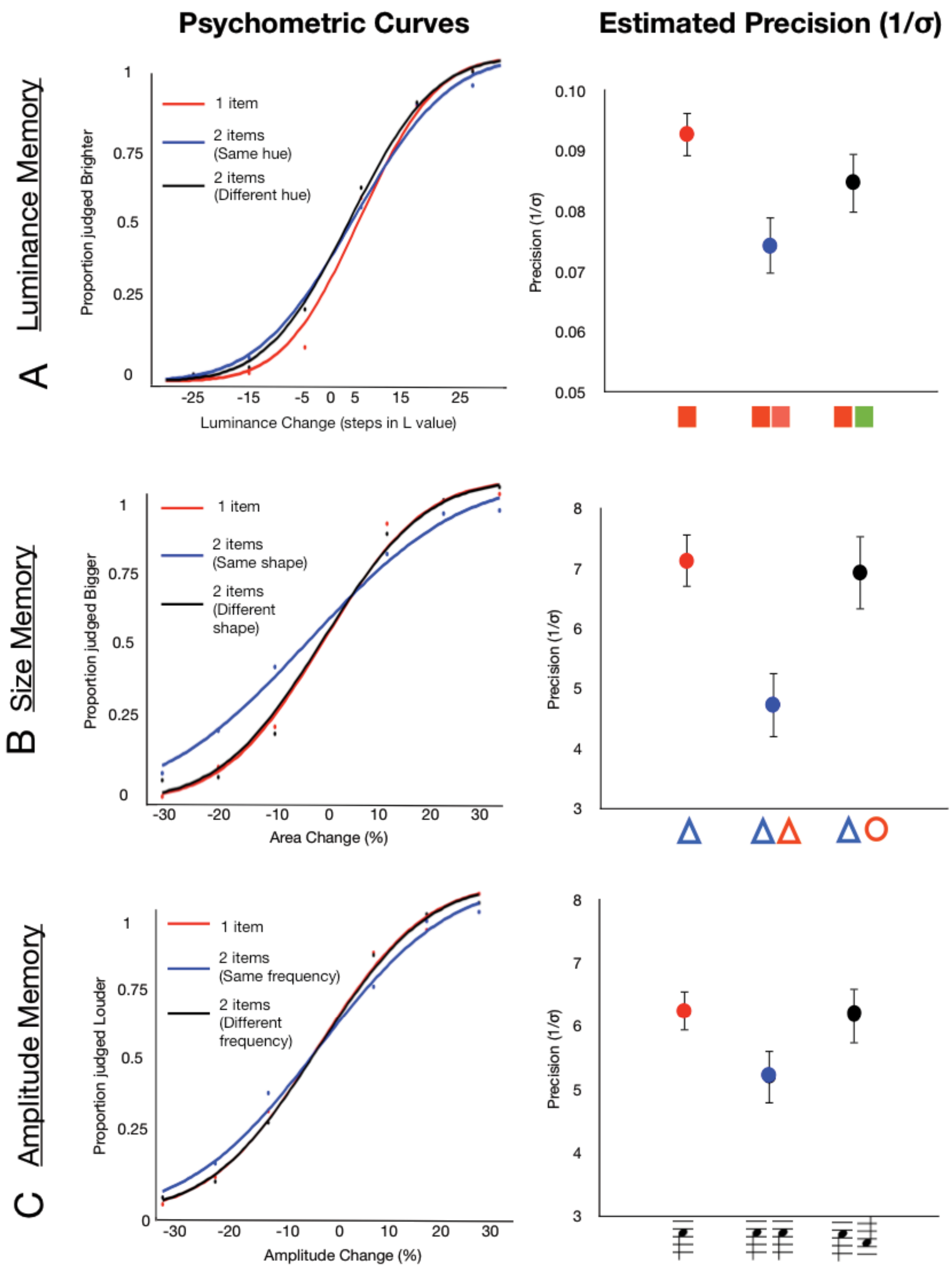

Figure 4.04. Estimated response functions (left column), and estimated precisions (right column) for each condition of each memory experiment a. Luminance judgment task. Response functions were estimated in terms of the probability of a brighter response as a function of the magnitude of luminance change. Reciprocals of standard deviations of estimated functions is plotted the right side. b. Size judgment task. Estimated response functions and reciprocals of their standard deviations shown in the middle row c. Amplitude judgment task. Estimated response functions and reciprocals of standard deviations shown in the bottom row. Error bars reflect ±1 S.E.

I turned back to the raw data associated with SIF trials in order to identify positive evidence of the presence of correspondence errors. We sorted trials into two categories based on the response to the probe demanded by the target and nontarget. *Consistent* trials were those in which both items demanded the same kind of response — the probe differed from each memory item in the same way. *Inconsistent* trials were those in which the nontarget demanded a different response than the target. If correspondence errors account for some mistakes in SIF trials, then Inconsistent trials should evidence a larger error rate than Consistent trials. We found a significant difference in this direction in the SIF trials of Experiments 13 and 15, and marginal difference in Experiment 14. (Inconsistent vs. Consistent error rate, respectively: Experiment 13, 44% vs. 9%, $t(11) = 14.199$, $p < .01$; Experiment 14, 26% vs. 16%, $t(7) = 2.143$, $p = .06$; Experiment 15, 37% vs. 15%, $t(11) = 4.04$, $p < .01$). Comparing error rate between SIF and DIF for Inconsistent trials showed higher error rate in SIF (Figures S6; SIF vs. DIF error rate, respectively: Experiment 13, 44% vs. 37%, $t(11) = 2.95$, $p = .013$; Experiment 14, 26% vs. 7%, $t(7) = 3.94$, $p < .01$; Experiment 15, 37% vs. 32%, $t(11) = 0.64$, $p = 0.64$).

These results are consistent with the prediction that preventing correspondence errors eliminates VWM costs for two objects compared to one. In the typical case —that is, in SIF trials— estimated response function standard deviations are thought to only reflect representational precision. But the contrasting results from DIF trials suggest that they may also reflect correspondence errors, errors further evidenced by our analysis of Consistent compared to Inconsistent trials.

Visual WM Summary and Conclusion

I investigated the quality of VWM for two items compared to one in situations where the two memory items differed or not along a task irrelevant, but integral feature dimension. To identify integral dimensions, Experiments 11 and 12 adapted a paradigm by Garner & Felfoldy (1970), demonstrating that differences in one integral feature dimension can produce a challenge for comparing items along another integral dimension. Exploiting this fact, I employed a VWM paradigm that relies on comparisons between a memory and a probe item at test. In Experiments 13-15, the probe and the test item always shared the same integral feature value. The critical manipulation involved the integral feature value of an untested memory item in two item trials. When the two items shared the same integral feature value, we observed an apparent and typical decline in memory quality compared to trials with only a single memory item. But when the two memory items possessed different integral values, we did not observe declines in performance.

This is, to my knowledge, the only of many recent studies to eliminate entirely a memory cost for two items compared to one (Bays & Husain, 2008; van den Berg, 2012; Zhang & Luck, 2008; but see Bae et al., under review). The presence of such a cost is a critical prediction of both flexible-resource (Bays & Husain, 2008; van den Berg et al., 2012) and hybrid-resource theories (Alvarez & Cavanagh, 2004; Anderson, Vogel, & Awh, 2011**)**, wherein the quality of two items in memory should always be roughly half that of a single item. (Of current theories, only a traditional fixed capacity-model would not make this prediction; e.g. Cowan, 2001; Luck & Vogel, 1997). It is important that although the memory items differed along an integral dimension, it was unknown to participants which would be probed, *and the same feature of each item had to be* *remembered*. Thus the effects cannot be explained in terms of different pools of resources dedicated to the storage of different features.

Instead, an account of the current results must address the role of integral features in eliminating memory costs. I have supplied one such account, leveraging the fact that two item trials usually present observers with a correspondence problem, though one item trials do not. I hypothesized that differences along an integral feature dimension would prevent attendant correspondence errors by supplying observers with a reliable and salient anchor for correspondence decisions. In general, any theory which acknowledges noisy or probabilistic representations of object features must also acknowledge that correspondence computations are necessary for retrieving a memory, though the mechanisms underlying these computations are rarely discussed (but see Bae et al., under review; Levillain & Flombaum, 2012).

There is a second potential account of our results, also appealing to correspondence errors, but during perception as opposed to test. Specifically, researchers often conceive of perception and encoding as the noisy sampling of features from images (Girshick, Landy, Simoncelli, 2011; Vul, Hanus, & Kanwisher, 2009; Vul & Rich, 2010). On this end, so to speak, there also exists a correspondence challenge when more than one item is present. After drawing a sample with some feature content, an observer needs to assign a correspondence between the feature and one of the objects believed to be in the image. Because samples are noisy —because an observer should possess uncertainty about where exactly in time and space a sample came from— there is a risk of correspondence errors. Indeed, this exact explanation has recently been offered to account for well-known feature binding challenges (Treisman & Schmidt, 1982; Vul &

Rich, 2010). Of course, there is no feature binding challenge in one item displays. Thus typical costs associated with remembering two items compared to one may reflect encoding correspondence errors, instances during which stray samples influence the inferred properties of an object. And in turn, integral feature differences may prevent such sampling-related errors by making it clearer when pairs of samples arose from independent sources.

It may seem incompatible with previous evidence that apparent changes in precision between 1 and 2 items is actually driven by correspondence errors, particularly since several models have incorporated a 'misbinding' parameter meant to capture these instances (Bays, Catalao, & Husain, 2009; but see Anderson et al., 2011). But all modeling to date has assumed random causes of correspondence errors, an equal likelihood in all trials regardless of particular stimulus properties and arrangements. If correspondence errors have systematic causes —if they are more likely in some trials than others— then current models may not estimate their prevalence correctly. For example, correspondence errors may be more likely as a function of spatial proximity (Emrich & Ferber, 2012; Vul & Rich, 2010). Moreover, current results suggest that surface similarity may play a role. If items are more likely to be confused when they share an integral feature, perhaps the continuous extent of any similarity modulates the likelihood of correspondence error. Future research should explore this possibility.

Neurally, explaining the reported results in terms of how integral features may prevent correspondence errors requires mechanisms that represent integral feature combinations. Recent fMRI research suggests a potential mechanism, that neurons can be tuned to conjunctions of features. Integral features then, are those for which neurons reflect joint preferences, while separable features are those for which pairs of neurons represent conjunctions (Drucker, Kerr, & Aguire, 2009). Joint tuning could supply a basis for making correspondence decisions that utilize integral feature differences.

Finally, I emphasize that what is perhaps most surprising about the reported results is that memory performance *improved* when memory displays were made *more complicated*. In Experiment 14, memory performance was better when displays included a triangle and a circle, for example, as opposed to two triangles. Yet, one could more efficiently summarize the contents of two triangle displays. Redundancy in these scenes should have drawn down less memory resources than the varied displays. I therefore accounted for the performance observed not in terms of memory resources and storage, but in terms of correspondence computations that must be involved in encoding and retrieving contents to and from memory. In general, research has focused almost exclusively on the nature of VWM resources and storage, to the exclusion of the computations that must be involved in acquiring and using information. As I have shown here, accounting for such computations —and the errors that they may induce— can lead to the realization that storage limitations are less severe than they may initially seem. VWM resources appear at least ample enough to afford as precise representations of two items as of just one. But computations deciding between multiple options naturally become more error prone as the number of options increases (see also, Duncan, 1980; Navon, 1984).

Chapter 5. Summary and Conclusion

The research presented focused on Multiple Object Tracking, Visual Working Memory and Spatial Working Memory. In each case, the goal was to identify the role of correspondence computations in performance. I showed in particular that performance can be better when challenges associated with object correspondence are prevented. In the MOT experiments, correspondence challenges were prevented by distinctive surface information that was provided to nontargets when they were near targets. In spatial WM, correspondence challenges were prevented by a preview display in which all the memory items reappeared except for the to-be tested item. In the more general working memory experiments, correspondence challenges were resolved by employing integral features. When these challenges were prevented, better performance was obtained. Surprisingly, in spatial working memory with a preview display, performance with higher memory loads (up to set size 7 or 8) was indistinguishable from lower memory load. These results suggest that the correspondence computations are not only a relevant limit in some working memory contexts, but perhaps the primary or even only limit.

A second kind of evidence that I presented in several different places had to do with *Consistent/Inconsistent* classification of trials. The goal of this analysis was investigating the impact of correspondence errors on a trial-by-trial basis. In the memory experiments reported in Chapters 3 and 4, I sorted trials based on whether a correspondence error would lead to a mistake — Trials in which they would have much higher error rates. Although I could not analyze consistency, per se, in the MOT experiments presented in Chapter 2, I performed a similar trial-by-trial analysis. In particular, I analyzed the error rates in a suite of trials dependent on how frequently targets and nontargets approached one another. The more frequently they did, the larger the error rates turned out to be. These analyses are important because many theories of the kinds of resources that limit visual cognition treats all trials the same if they contain the same tracking or memory load. But from the perspective of correspondences, such trials are not all the same.

The last kind of evidence I presented came from my modeling work in SWM. My model started with the basic assumption that internal representations always possess a certain amount of noise, such that one cannot know for sure what has been seen and remembered. But unlike other models, mine actually implemented the correspondence algorithms that become necessary given this uncertainty. The model did an excellent job of accounting for human performance, with virtually no free parameters.

Taken together, these results suggest that uncertainty about object correspondence imposes the primary constraint on visual cognition, broadly. This is very different from prevailing views. Beginning with Sperling (1960), researchers have been interested in how much visual information the visual system can process at once. The work presented here suggests that this is actually the wrong kind of question. Limits may arise because of how much information *can be used effectively at once*.

Spatiotemporal vs. Surface Properties

One issue that came up throughout my dissertation has to do with the role of surface properties in correspondence problems. But the point was not always the same. Chapter 3, on spatial working memory, suggested that surface properties were not used by observers to make correspondence decisions, a result consistent with work on apparent motion and phenomena such as the tunnel effect. But in Chapter 2, surface properties were used to help participants discriminate targets and distractors, and in Chapter 4, integral surface property relationships were used to scaffold correspondences. There are important reasons, though, why features played a role in those places. In Experiment 2, color was used to aid discrimination when two objects were near one another, that is, to help resolve an *ambiguous* spatial situation. In many places where features do not play a role, the challenge is not comparing two side by side objects, but relating previous events with prior ones. Similarly, the point of Chapter 4 was that strong integral feature relationships are necessary to supply a strong featural basis for correspondences. Thus while this research reinforced the important role of spatiotemporal relations in correspondence computations, it also demonstrated how the issue is more subtle than it may appear. Future research should investigate the role of surface properties within the context of specific correspondence algorithms that can take them as inputs.

Conclusion

Identifying object correspondences is essential for successful visual cognition. However, its role has been studied only limited area of visual cognition. Here, I sought to find evidence for the role of object correspondence in two areas in visual cognition. Throughout a series of MOT and VWM experiments, I found converging evidence for the role of correspondence computation. Surprisingly, experimental results suggest that failures of correspondence computations seem to be the major limiting factor of visual cognition—measured working memory precision was unchanged across memory loads when the failures of correspondence computations were prevented. In addition, implementing a correspondence algorithm in a computational model captures human performance in typical working memory tasks. These results suggest that what is limited may not be the representational precision but the visual system's ability to solve the complex correspondence problems. More generally, these results emphasize the importance of scrutinizing cognition at the algorithmic level in order to understand what makes it difficult.

Reference

Anderson, D. E., Vogel, E. K., & Awh, E. (2011). Precision in visual working memory reaches a plateau when individual item-limits are exceeded. *Journal of Neuroscience, 31*, 1128-1138.

Alvarez, G. & 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.

Alvarez, G. & Franconeri, S. (2007). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*, 1 - 10.

Alvarez, G.A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*, 392-398.

Ariely, D. (2001), Seeing sets: Representation by statistical properties. *Psychological Science, 12*, 157- 162.

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.

Bae, G.Y. & Flombaum, J. (2011). Amodal causal capture in the tunnel effect. *Perception, 40*, 74-90.

Bae, G.Y. & Flombaum, J. (2012). Close encounters of the distracting kind: Identifying the cause of visual tracking errors. *Attention, Perception, & Psychophysics*, *74*, 703-715.

Bae, G.Y., Wilson, C., & Flombaum, J. (under review). A nearest neighbor model of object correspondence in retrieval from spatial working memory.

Barker, K., Allen, RD. & McGeorge, P. (2010). Multiple Object Tracking: Enhanced visuospatial representations as a result of experience. *Experimental Psychology*, *57*, 208 - 214.

Barton, B., Ester, E. F., & Awh, E. (2009). Discrete resource allocation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1359-1367.

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

Bays, P.M., Catalao, R.F.G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*, 1-11.

Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of features in visual working memory. *Neuropsychologia, 49*, 1622-1631.

Bettencourt, K., & Somers, D. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision, 9*,1-11.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177-178

Brady, T. F. and Alvarez, G.A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychological Science*, *22*, 384-392.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and towards structured representations. *Journal of Vision*, *11*, 1-34.

Burnett S, Zokaei N, van der Staaij I, Bays PM & Husain M. (2012). Development of visual working memory precision in childhood. *Developmental Science, 15*, 528-539.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*, 393-404.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 109-127.

Cornelissen, F.W., & Greenlee, M.W. (2000). Visual memory for random block patterns defined by luminance and color contrast.*Vision Research*, *40*, 287–299.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.

Dawson, M. (1991). The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem. *Psychological Review*, *98*, 569-603.

Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *Journal of Neuroscience*, *28*, 4183–4191.

Drucker, D.M., Kerr, W.T., & Aguirre, G.K. (2009). Distinguishing conjoint and independent neural tuning for stimulus features with fMRI adaptation. *Journal of Neurophysiology*, *101*, 3310-3324.

Duncan, J. (1980). The demonstration of capacity limitation. *Cognitive Psychology, 12*, 75-96.

Emrich, S.M. & Ferber, S. (2012). Competition during encoding increases binding errors in visual working memory. *Journal of Vision*. *12*, 1-6.

Flombaum, J. I., Scholl, B. J., & Santos, L. R. (2009). Spatiotemporal priority as a fundamental principle of object persistence. In B. Hood & L. Santos (Eds.), *The* *origins of object knowledge* (pp. 135-164). Oxford University Press.

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

Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can you select? *Journal of Experimental Psychology: Human Perception & Performance*, *33*, 1003-1012.

Franconeri, S., Jonathan, S., & Scimeca, J. (2010). Tracking multiple objects is limited only by object spacing, not speed, time, or capacity. *Psychological Science*, *21*, 920-925.

Franconeri, S., Lin, J., Pylyshyn, Z., Fisher, B., & Enns, J. (2008). Evidence against a speed limit in multiple object tracking. *Psychonomic Bulletin & Review, 15*, 802-808.

Fukuda, K., Awh, E., & Vogel, E.K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology*, *20*,177-182.

Garner, W.R. (1974). *The processing of information and structure*, Wiley, New York.

Garner, W.R. (1976). Interaction of stimulus dimensions in concepts and choice processes. *Cognitive Psychology*, *8*, 98-123.

Garner, W.R, & Felfoldy, G.L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*, 225-241.

Girshick, A.R., Landy, M.S., Simoncelli, E.P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience, 14*, 926-932.

Gorgoraptis, N., Catalao, R.F.G., Bays, P.M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *Journal of Neuroscience, 31*, 8502-8511.

Gmeindl, L., Nelson, J. K., Wiggin, T., & Reuter-Lorenz, P. A. (2011). Configural representations in spatial working memory: modulation by perceptual segregation and voluntary attention. *Attention, Perception, & Psychophysics, 73*, 2130-2142.

Holcombe A.O., & Chen, W.Y. (2012). Exhausting attentional tracking resources witha single fast-moving object. *Cognition*, *123*, 218-228.

Horowitz, T. S. & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception, & Psychophysics, 72*, 1765-1775.

Jiang, Y., & Kumar, A. (2004). Visual short-term memory for two sequential arrays: One integrated representation or two separate representations. *Psychonomic Bulletin & Review, 11*, 495-500.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 683-702.

Julesz, B. (1971). *Foundations of Cyclopean Perception*. Chicago: The University of Chicago Press.

Jolicœur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology, 36*, 138-202.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*, 171 - 216.

Lee, B., & Harris, J. (1996). Contrast transfer characteristics of visual short-term memory. *Vision Research, 36*, 2159–2166.

Levillain, F., & Flombaum, J. I. (2012). Correspondence problems cause repositioning costs in visual working memory. *Visual Cognition*, *20*, 669-695.

Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., & Enns, J. T. (2005). Multiple-object tracking is based on scene, not retinal, coordinates. *Journal of Experimental Psychology: Human Perception & Performance*, *31*, 235-247.

Logie, R. H., Brockmole, J. R., & Jaswal, S. (2011). Feature binding in visual short-term memory is unaffected by task-irrelevant changes of location, shape, and color. *Memory and Cognition, 39*, 24-36.

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

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences in cognitive ability. *Trends in Cognitive Sciences, 17,* 391-400.

Ma, Z. & Flombaum, J. I. (2013). Off to a bad start: uncertainty about the number of targets at the onset of multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 1421-1432.

Makovski, T., & Jiang, Y. (2009a). The role of visual working memory in attentive tracking of unique objects. *Journal of Experimental Psychology: Human Perception & Performance, 35*, 1687-1697.

Makovski, T., & Jiang, Y. (2009b). Feature binding in attentive tracking of distinct objects. *Visual Cognition, 17*, 180-194.

Marr, D. (1982). *Vision*. New York: Freeman.

Moore, C. M., Lanagan-Leitzel, L. K., & Fine, E. M. (2008). Distinguishing between the precision of attentional localization and attentional resolution. *Perception & Psychophysics, 70*, 573-582.

Navon, D. (1984). Resources —A theoretical soup stone? *Psychological Review, 91*, 216-234.

O'Hearn, K., Landau, B., & Hoffman, J. (2005). Multiple object tracking in people with Williams Syndrome and in normally developing children. *Psychological Science*, *16*, 905 - 912.

Oksama, L., & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Visual Cognition, 11*, 631-671.

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 332-350.

Paivio, A., & Bleasdale, F. (1974). Visual short-term memory: A methodological caveat. *Canadian Journal of Psychology*, *28*, 24–31.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*, 739–744

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Figure 1.05Distinguishing feature integration from detection. *Journal of Vision, 4*, 1136-1169,

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning & Memory*, *2*, 509-522.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial index model. *Cognition, 32*, 65-97.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*, 127-158.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*, 179 - 197.

Rouder, J.N., Morey, R.D., Cowan, N., Zwilling, C.E., Morey, C.C., & Pratte, M.S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences, 104,* 5975-5979.

Scholl, B. J., & Flombaum, J. I. (2010). *Object persistence*. In B. Goldstein (Ed.), Encyclopedia of Perception, Volume 2 (pp. 653-657). Thousand Oaks, CA: Sage Publications.

Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology*, *54*, 1 - 14.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*, 159-177.

Shapiro, K., Driver, J., Ward, R., & Sorensen, R. E. (1997). Priming from the attentional blink: A failure to extract visual tokens but not visual types. *Psychological Science*, *8*, 95-100.

Shim, W.M., Alveraz, G.A., & Jiang, Y. V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychonomic Bulletin & Review*, *15*, 390-397.

Smith, K. Ba, S., Gatica-Perez, D., & Odobez, J. M. (2005). *Using particle filters to track varying numbers of objects*. Paper presented at Computer Vision and Pattern Recognition (CVPR), Sand Diego, CA.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monograph*, *74*,1-29.

St. Clair, R. L., Huff, M. & Seiffert, A. E. (2010). Conflicting motion information impairs multiple object tracking. *Journal of Vision, 10*, 18.

Thiele, J. E., Pratte, M, S., & Rouder, J. N. (2011). On perfect working memory performance with large numbers of items. *Psychonomic Bulletin & Review*, *18*, 958-963.

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

Treisman, A. & Schmidt H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*, 107-141.

Tsal, Y. & Bareket, T. (2005). Localization judgments under various levels of attention. *Psychonomic Bulletin & Review, 12*, 559-566.

Ullman, S. (1984). Visual routines. *Cognition*, *18*, 97-159.

van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012) Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*, 8780-8785.

Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review*, *9*, 739-743.

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1656-1674.

Vogel, E. K. & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 784-751.

Vul, E., Frank, M. C., Alvarez, G. A., & Tenenbaum, J. B. (2009). Explaining human multiple object tracking as resource constrained approximate inference in a dynamic probabilistic model. *Advances in Neural Information Processing Systems, 22*, 1955-1963.

Vul, E. & Rich, A.N. (2010). Independent sampling of feature enables conscious perception of bound objects. *Psychological Sciences*, *21*, 1168-1175.

Wheeler, M. E. & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48-66.

Whitney, D. & Levi, D. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*, 160-168.

Wilken, P. & Ma, W.J. (2004). A detection theory account of change detection. *Journal of Vision, 4,* 1120-1135.

Wood, C.C. (1975). Auditory and phonetic levels of processing in speech perception: neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Perception and Performance*, *104*, 3-20.

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review, 10*, 80–87.

Yi, D-J., Turk-Browne, N. B., Flombaum, J. I., Kim, M-S., Scholl, B. J., & Chun, M. M. (2008). Spatiotemporal object continuity in human ventral visual cortex. *Proceedings of the National Academy of Sciences*, *105*, 8840–8845.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233-236.

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

Zuvic, S. M., Visser, T. A. W., & Di Lollo, V. (2000). Direct estimates of processing delays in the attentional blink. *Psychological Research*, *63*, 192-198.

VITA

Gi-Yeul Bae was born on November 23, 1980 in Daejeon, South Korea to Joongsik Bae and Soonim Park. He has a younger brother, Se-Yeul, with whom he spent his childhood in Gumi, South Korea. In 1999, Gi-Yeul moved to Seoul for college. Initially, he was interested in Economics, so he majored in Agricultural Economics at Korea University but later he became more interested in Psychology. This change was mainly triggered by his general curiosity on how human think and behave, and a popular psychology book written by a Korean Psychologists, Dr. Min Gyu Lee, specifically made him pursue human cognition as his major. After the first two years of undergraduates, he went to Korea Navy for 2.5 years for military duty. He came back to the college in 2003 and started to work as a RA in psychology laboratory run by Dr. Ki Chun Nam. After receiving his B.A. in 2006, he decided to go to Graduate school to do his own research. Under the supervision of Dr. Yang Seok Cho, he received M.A. from Korea University in 2008. While he was in the graduate school, he published his first scientific article. To continue to pursue his research interest, Mr. Bae enrolled as a graduate student at Johns Hopkins University under the supervision of Dr. Jonathan Flombaum in the fall of 2009, where his research interests have focused on visual cognition. He earned his second M.A. in 2010, and he anticipates completing his Ph.D. in 2014, at which point he will become a postdoctoral associated at University of California-Davis in the laboratory of Dr. Steven Luck.