ELECTROPHYSIOLOGICAL MEASURES OF ATTENTIONAL TRACKING AND WORKING MEMORY

by

TRAFTON DREW

A DISSERTATION

Presented to the Department of Psychology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2009

University of Oregon Graduate School

Confirmation of Approval and Acceptance of Dissertation prepared by:

Trafton Drew

Title:

"Electrophysiological measures of attentional tracking and working memory"

This dissertation has been accepted and approved in partial fulfillment of the requirements for the degree in the Department of Psychology by:

Edward Vogel, Chairperson, Psychology
Edward Awh, Member, Psychology
Ulrich Mayr, Member, Psychology
Paul van Donkelaar, Outside Member, Human Physiology

and Richard Linton, Vice President for Research and Graduate Studies/Dean of the Graduate School for the University of Oregon.

June 13, 2009

Original approval signatures are on file with the Graduate School and the University of Oregon Libraries.

An Abstract of the Dissertation of

Trafton Drew for the degree of Doctor of Philosophy

in the Department of Psychology to be taken June 2009

Title: ELECTROPHYSIOLOGICAL MEASURES OF ATTENTIONAL TRACKING AND WORKING MEMORY

Approved: ______________________________
Dr. Edward K. Vogel

In the multiple object tracking (MOT) task, observers are presented with multiple identical objects, some of which are temporarily identified as targets. After a selection period, all objects move randomly and independently for several seconds. At the end the motion period, all objects stop and observers must identify the target objects again. This task has been used to study a variety of important cognitive questions from object-based attention to cognitive development, divided attention and the development of expertise. Yet, surprisingly little is known about the neural mechanisms that underlie the ability to track multiple targets independently. Although a number of researchers have used fMRI (functional magnetic imaging) to examine what areas are active during MOT, the current set of studies is the first to employ ERPs (event-related potentials) to examine the neural mechanisms of MOT. With excellent temporal resolution, the ERP methodology allows researchers to delineate the time course of different phases of a single task with millisecond precision, something not possible with fMRI. In Chapter II, we manipulated the number of targets and difficulty of tracking and observed a lateralized contralateral negativity that was sensitive to the number of targets but not difficulty of tracking. Chapter III examined the effect of irrelevant white probes flashed briefly throughout the trial while observers tracked. We observed modulations of early visual components that indicated that during tracking, spatial attention focused on targets but did not differentiate between distractors and empty space. Finally, in Chapter IV, we examined the relationship between visual working memory (VWM) and MOT by manipulating the presence or absence of task relevant motion. We found that the waveforms evoked by an MOT task in the absence of task-relevant motion were nearly identical to waveforms evoked by the VWM task, suggesting that VWM is an important part of the typical MOT task.

This thesis includes previously published and unpublished material.

CURRICULUM VITAE

NAME OF AUTHOR: Trafton Drew

PLACE OF BIRTH: Durham, North Carolina

DATE OF BIRTH: February 20, 1980

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
University of North Carolina-Chapel Hill, Chapel Hill, NC
Georgia Institute of Technology, Atlanta, GA

DEGREES AWARDED:

Doctor of Philosophy, Cognitive Neuroscience, 2009, University of Oregon
Master of Arts in Psychology, 2006, University of Oregon
Bachelor of Arts in Psychology, 2002, University of North Carolina-Chapel Hill

AREAS OF SPECIAL INTEREST:

Neural Basis of Visual Cognition

PROFESSIONAL EXPERIENCE:

Graduate Research Fellow, Department of Psychology, University of Oregon, Eugene, OR 2004-2009
Graduate Teaching Fellow, Department of Psychology, University of Oregon, Eugene, OR 2004-2005, 2008
Post-Graduate Researcher, Department of Psychology, University of California Davis, CA 2003-2004
Research Assistant, Department of Psychology, University of Wales-Bangor, Bangor, Wales 2002-2003

PUBLICATIONS:

Drew, T. & Vogel, E. K. (in press). The capacity of working memory. In *New Encyclopedia of Neuroscience*, L Squire, T Albright, F Bloom, F Gage & N Spitzer (Eds.)

Drew*, T., McCollough*, A. W., Horowitz, T. S., & Vogel, E. K. (2009). Attentional enhancement during multiple object tracking, *Psychonomic Bulletin & Review*, *16*, 411-417.

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

Drew, T. & Vogel, E. K. (2008). Recently Attended Masks are Less Effective, *Perception & Psychophysics*, *70*(1), 96-103.

Drew, T, McCollough, A., & Vogel, E. K. (2006). Event-related potential measures of working-memory, *Clinical EEG and Neuroscience*, *37*, 286-291.

Drew, T. & Shapiro, K. (2006). Representational masking in the attentional blink, *Visual Cognition*, *13*, 513-528.

ACKNOWLEDGMENTS

I could not have completed this dissertation without the help and support of many people. I would like to thank all of the scientists I have worked with in my short, but well traveled career, from Tom Wallsten and Peter Gordon at UNC-Chapel Hill, to Kimron Shapiro at UW-Bangor, G. Ron Mangun and Barry Giesbrecht at UC-Davis and finally my current advisor, Ed Vogel. Ed has been a great mentor and I hope that we will continue to work together long after I am gone. I would also like to thank my dissertation committee: Ed Awh, Ulrich Mayr and Paul van Donkelaar for providing me with helpful interpretations of data and ideas about follow-up projects. Thanks also to fellow Vogel lab mates Andrew McCollough, Keisuke Fukuda, Veronica Perez, Jason Fair and Nathan Ashby who helped me in all phases of this dissertation from helping design experiments, to running subjects and helping clean caps. Finally, thanks to my family for supporting me through my many travels and infrequent visits home.

TABLE OF CONTENTS

Chapter Page

Chapter Page

II. NEURAL MEASURES OF INDIVIDUAL DIFFERENCES IN SELECTING AND TRACKING MULTIPLE MOVING OBJECTS

Chapter Page

LIST OF FIGURES

Figure Page

CHAPTER I

MULTIPLE OBJECT TRACKING AND ATTENTION

INTRODUCTION

Our limited ability to divide attention is one of the central limitations with cognition and this ability is thought to underlie performance on a diverse array of tasks from driving on a crowded highway to the ability (or lack thereof) to maintain a line of thought while composing an email while being interrupted by an impending appointment alarm. While there is a rich history studying the spatial division of attention using transient cueing tasks, more recently researchers have begun to study the sustained division of attention. While the two approaches are inextricably linked, the sustained approach appears to have a more ecological validity as it relates more closely to real life situations where divided attention appears necessary to complete a task. Recently, there has been an explosion of studies that study sustained divided attention using the multiple object tracking (MOT) task.

The goal of this dissertation is to explore the relationship between MOT, attention and working memory. Portions of this work have been previously published or are to be published with additional authors. Chapter II was published with Edward K. Vogel in the Journal of Neuroscience. Chapter III was published with Andrew McCollough, Todd S.

Horowitz and Edward K. Vogel in Psychonomic Bulletin and Review. Chapter IV is will be submitted with co-authors Todd S. Horowitz, Jeremy Wolfe and Edward K. Vogel.

In this task, targets are identified amongst a set of distractors at the beginning of each trial. In the second phase of the trial all objects are visually identical, forcing the subject to recall which objects were identified as targets in the first phase, and all of the objects move randomly and independently for some period of time. At the end of the trial participants are asked to either identify all of the targets that they tracked or respond to a probed item that was either a target of non-target distractor. It is my hope to use the vast visual attention literature to elucidate the role of attention during MOT. This is a useful endeavor because the MOT paradigm may be thought of as an amalgamation of number of well-studied tasks from the visual attention literature. Rather than studying each of these tasks in isolation and attempting to generalize the findings to attention in the real world, the MOT paradigm allows researchers to study a multi-faceted task that necessitates a number of different types of attention for completion of the task. It is my hope that by studying this situation, I may be able to better understand how attention is implemented in real-world situations.

What is the Role of Attention in MOT?

The standard MOT paradigm may be decomposed into a number of smaller components tasks, some of which have been studied extensively in the visual attention literature. The hope is that by examining these literatures in the context of MOT I may be able to better understand the role of attention during the task and that basic findings in MOT may in turn elucidate the role of attention in the common real world activity of mentally tracking information. At the beginning of each trial, the subject must enumerate each of the targets while ignoring the distractor items. I will terms this the selection phase and note that it is strikingly similar to the standard visual search task. Once the objects start moving, subjects must continually update the relationship between each target and it's location. To successfully complete this task, subjects must divide attention over distinct foci, rapidly switch between the locations, use some sort of higher level grouping heuristic or some combination of all three of these strategies. There is a sizable literature both for and against the 'spotlight' theory of attention and this literature will be used to better understand the tracking phase of the standard MOT task. One striking contrast between the existing MOT literature and the visual attention literature is the use of neuro-imaging: many of the most profound, important findings in the visual attention literature are thanks to neuroimaging while there is currently a relative paucity of MOT experiments that have employed neuroimaging.

Selection

The simple task of finding a target amongst distractors has been intensively studied as means of exploring the underlying mechanisms of visual perception and visual selective attention. One could argue that people perform hundred of visual searches every day: from attempting to find a corkscrew in a drawer full of similar looking tools to searching an intersection for a sign to the interstate. Each trial in a multiple object tracking experiment starts with a very simple visual search: find the blinking targets. In almost all existing MOT papers, the targets blink on and off for 2 seconds prior to motion onset. In the terminology of visual search, this qualifies as pop-out search as the targets differ from the non-targets on a single highly discriminable feature: blinking. It would therefore be predicted that the time to select the targets would not increase as the number of distractors increased. Abrupt onsets are known to elicit 'attentional capture' meaning that attention tends to quickly orient to the blinking stimulus. In most demonstrations of this effect, the item that onsets is a distractor and the attentional capture of this object increases the amount of time that it takes to find the target (Yantis and Hillstrom 1994). However, there appears to be a limit on how many items may capture attention: Yantis and Johnson (Yantis and Johnson 1990) found the effect of abrupt onset asymptotes at 4 items.

This finding is important in the context of MOT for a number of reasons. First, the fact that there is an apparent capacity for the number of items that may simultaneously attract attention suggests that selection is not an entirely pre-attentive process, even when the targets automatically grab attention. Regardless of how phenomenologically simple selection feels, the act of selection seems to require a form of **capacity-limited** attention. Second, the apparent capacity for attentional capture is roughly 4 items: strikingly similar to the capacity estimates for the number of items may simultaneously tracked, the number of items that may be subitized in parallel and the capacity of working memory (Luck and Vogel 1997; Pylyshyn and Storm 1988; Trick and Pylyshyn 1994). Miller wrote one of the most influential psychological papers ever written about continually running into the 'magic number' 7 +/-2 and it seems that in the visual domain the magic number is 4 +/-1 (Cowan 2001; Miller 1956). This may be evidence in favor of some sort of fundamental cognitive bottleneck that limits the attentional processing capabilities across a wide array of tasks. Cowan (2001) believes that working memory representation is the focus of attention and that it is the focus of attention that constrains the capacity of each of these seemingly disparate tasks. One group has attempted to explain this important limit in terms of a biologically constrained model (Raffone and Wolters 2001), but there is currently little neurological data that either confirms or refutes biological plausibility of the model. Although Yantis and Johnson's evidence makes it unlikely that the targets in an MOT experiment are automatically and simultaneously selected, by flashing the targets on and off for 2s at the beginning of each trial the hope is that there is ample time to select the targets in an serial, effortful manner if necessary. Almost all of the existing visual search literature deals with search times for a single target, but if one assumes that there is a linear relationship between number of targets and the time it takes to find all of them, 2s should be enough time to locate each of the targets. In fact, unpublished research in our lab (see Figure 1-1) has shown that given the ease of the search, 2s may be an unnecessarily long period of time for selection: when subjects were randomly given 500ms or 2s to select targets performance in the two conditions was statistically equivalent. Moreover, there was a very strong correlation between performance on the two versions of the task. This data raises the possibility that the primary limitation for selection of multiple targets may not by the amount of time, but rather by the number of objects that must be selected.

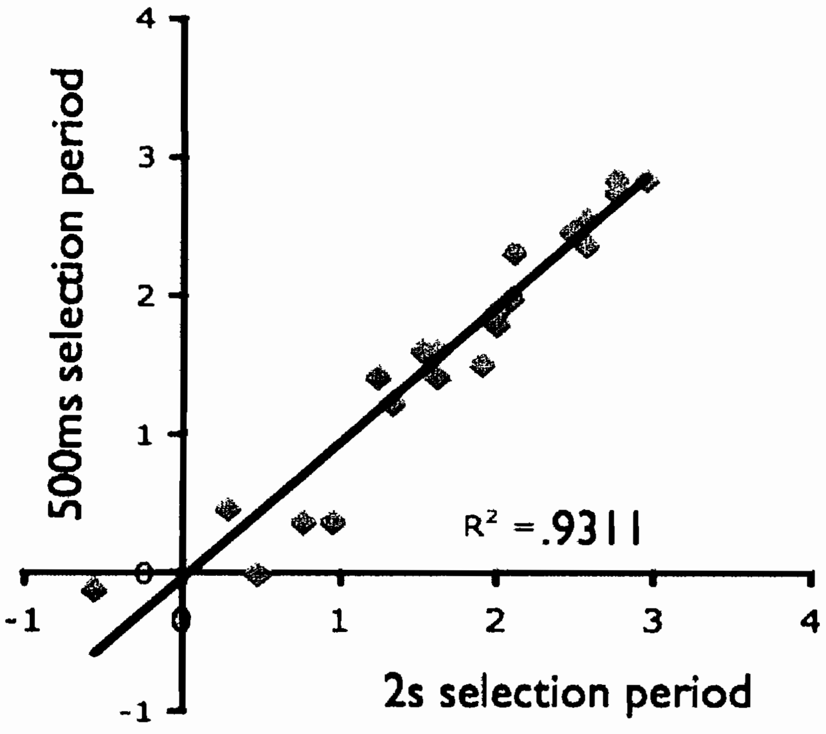

Figure 1-1: Correlation Between Tracking Performance and Selection Period.

When the selection criterion is a non-singleton feature or when the selection period is very brief (200ms), performance on the tracking task degrades (Pylyshyn and Annan 2006). This is perhaps an obvious point, but it underscores the importance of the selection phase: the MOT paradigm is designed such that if you cannot select a target initially you will not be able to track it because targets and distractors become identical once motion begins. In this sense, the selection phase serves as the oft-overlooked initial bottleneck of multiple-object tracking. Recent evidence has suggested that the number of locations that a person can simultaneously attend is dependent upon spatial precision that is necessary to select each location (Franconeri et al 2007). That is, when a very tight focus was needed to successfully select a location, participants were able to select fewer locations than when fewer items on the screen necessitated looser focus to select a location. The authors see this as evidence in favor of a flexible resource that may adjust the amount of resources devoted to each location as a function of the fidelity of representation necessary to perform the task. This results in apparent capacity varying as a function of the resolution necessary. Another interpretation for this data is that it is much easier to group objects in sparse display than in a dense display. Although it is difficult to rule out this alternative account, this result reports evidence that participants are capable of functionally selecting more items (5.6) in less time (500ms) than participants are asked to select in typical MOT experiments. However, a direct comparison between the selection in this experiment and a standard MOT experiment may not be meaningful as the purpose selection is different in the two tasks. In Franconeri's experiments, the subject must hold a group of locations in memory and then search the locations for a target while in MOT experiments the locations of the targets must be held and then iteratively updated once the tracking begins. The fact that subjects can select and hold locations in memory does not necessarily mean that they can simultaneously select rapidly updatable objects while keeping them separate from distractors. Furthermore, the static displays used by this group seem to invite grouping. In fact Yantis (1992) has shown that manipulating the initial ease of grouping targets results in an apparent increase in tracking capacity.

Still, as crowding apparently makes selection more difficult this result may help explain why tracking becomes more difficult as the visual angle of the tracking area decreases. Intrilligator and Cavanagh (Intriligator and Cavanagh 2001) found that the ability to track multiple items was decreased despite the fact that control experiments indicated that the size manipulations were not severe enough to effect perception of the objects. They conclude that the performance must be due to a limit on a higher attentional mechanism. Franconeri et al's findings indicate that density of display may adversely affect the attentional resolution of selection as well.

It seems clear that attention is an important part of the selection phase of MOT. Although the standard cueing procedure should make identification of isolated targets relatively simple, converging evidence seems to indicate that representing multiple target locations may be subject to capacity limitations irrespective of the amount time given. Little is know of the neural mechanisms of selecting multiple objects, but a great deal is known about the early attentional mechanisms that underlie simple selection. By examining this literature, we may be able to better understand how multiple objects are selected so that they may be tracked.

Neural Mechanisms of Selection

Cognitive neuroscience has revolutionized how researchers think about attention. While the focus of attention research prior to these techniques was to understand the operating principles of attention, the ability to observe brain activity while subjects perform tasks has allowed researchers to explore the neural mechanisms that underlie performance. The hope is that these methodologies will allow us to understand the processing that leads to behavior so that we may better understand the behavior. One question that has been very difficult to address through behavioral data alone is whether targets in the beginning of trial are selected serially or simultaneously. Some of the most compelling evidence in favor of parallel processing of multiple locations is thanks to neuroimaging techniques that allow researchers to index attention using task irrelevant probes or features.

While most theories agree that the primary function of visual attention is to select some stimuli while ignoring others, different theories have postulated different reasons why processing is slow when there are multiple stimuli to process. According to *feature integration theory* the visual system automatically (meaning in parallel and without any capacity limitations) decomposes the scene into maps of simple features such as orientation and color (Treisman and Gormican 1988; Treisman and Sato 1990; Treisman and Gelade 1980). Separate features cannot be coded to an object without focusing attention on the object. Without attention, there may be a coarse representation of the presence or absence of a feature but the feature will not be bound to a specific object. Attention is thus thought to play the role binding features, objects and locations together. In an attempt to apply *feature integration theory* to anatomical and physiological data, Luck and colleagues (Luck et al 1997b) invoked the *ambiguity resolution theory*. One of the central ideas of this theory is that because the size of receptive fields gets so much larger as information travels up the visual stream, the coding of information must be distributed over many neurons rather rely upon single cell that codes for the presence or absence of a specific object. The distributed network implies that perception of an object should be relatively indifferent to slight changes in viewpoint, illumination or location on retina. This method of representing information becomes much more complicated when there is more than one object in a single receptive field. If there is a red circle and a green square in a single RF, it may be ambiguous which item is red if the distributed network simply codes for the presence or absence of features. This ambiguity might lead to a binding error in Triesman's terms (Treisman and Schmidt 1982). *Ambiguity resolution theory's* (ART) main contribution is that it posits that the primary role of selective attention is to resolve ambiguous coding situations. This predicts that as the proximity of a competing stimulus decreases, the need for attention increases. It also predicts that binding errors will only occur when there are multiple items in a single receptive field, thereby necessitating attention to disambiguate. Both of these predictions have been supported empirically (Cohen and Ivry, 1991; Sohn et al 1996; Treisman and Gelade 1980).

While recording from single neurons in area V4, macaque monkeys exhibited an attentional modulation only when the item monkey was looking for was in same the RF as another stimulus. In V1, where RFs are too small to contain multiple stimuli, no attentional modulation was found. Furthermore, attention effects were larger when the target and distractor were presented simultaneously than when presented sequentially. It appears that both temporal and cortical proximity of representation play a role in determining the necessity of attention. In areas with larger receptive fields, neurons tend to be driven by specific stimuli, meaning that if a stimulus is presented in the neuron's RF, the firing rate will increase above the baseline firing rate (Moran and Desimone 1985). Chelazzi *et al.* (Chelazzi and Desimone 1994; Chelazzi et al 1993) used this finding to demonstrate strong evidence that selection of a simple target in a visual search task begins approximately 175ms after stimulus onset. As expected, when a stimulus was presented in neuron's RF, the neuron's firing rate quickly increased above baseline. If the stimulus was not effective for the neuron in question, beginning around 175 ms the firing rate quickly slowed to the baseline rate while the same stimulus elicited a prolonged elevated firing rate if the stimulus was effective for that neuron. This could theoretically resolve ambiguity by allowing only one stimulus to be active (as indicated by an increased firing rate) approximately 250 after stimulus onset. In general, these attention affects observed by Chelazzi are larger with more difficult (via complex stimuli or asking the monkey to localize rather than just detect) tasks. This seems to suggest that it may serve as a measure of focal attention. In keeping with ART, Luck and colleagues later showed that the attention effects were larger when there was more than one item in the neuron's receptive field (Luck et al 1997a).

There are a number of striking parallels between this attentional of selection and an electrophysiological component known as the N2pc (Luck et al 1997b). Perhaps most striking, Woodman and colleagues recently recorded ERPs from macaque monkeys and replicated many of classic demonstrations that had previously been confined to human subjects (Woodman et al 2007). The N2pc is an ERP waveform that is specific type of the N2 component (meaning it is generally second negative going component evoked in response to a stimulus) that has a posterior contralateral focus. Typically, the N2pc is observed as a negative deflection at sites contralateral to the target in a search display that tends to occur between 175-275 post stimulus. The N2pc is an index of covert visuospatial attention and is generally thought to reflect the process of attentional filtering via distractor suppression (Woodman and Luck 2003), although some argue that the component is simply sensitive to target selection rather than distractor suppression (Eimer 1996). Luck and Hillyard (Luck and Hillyard 1994a, b) were the first to use the N2pc to better understand the role of attention during visual search. Eimer (1996) later showed that a robust N2pc may be found when the target is one of just two items and the distractor item is on the opposite hemifield from the target and suggested that this indicates that the N2pc must represent attentional selection rather than distractor suppression. On the other hand, Luck and Hillyard (Luck and Hillyard 1994a) found that the N2pc was absent in absence of distractors, but the same target elicited an N2pc in the presence of distractors. Furthermore Luck and colleagues have demonstrated that the N2pc is larger for identification tasks than for detection tasks, is increased as the similarity between target and distractor increases and is larger for tasks that require localization of the target (Hopf et al 2002; Luck et al 1997b). Finally, a significant N2pc is elicited by non-targets that are very similar to targets, but not when decisions must be made on the basis of global context (Luck and Hillyard 1994a). A recent MEG study suggested that distractors in the field opposite to the target are suppressed first, leading to a large effect on the ipsilateral side and this activity is followed by contralateral activity that seems to reflect suppression of the distractors in the same field as the target (Hopf et al, 2002). The N2pc has also been used to demonstrate strong evidence in favor of serial deployment of attention during a difficult visual search (Woodman and Luck 1999; Woodman and Luck 2003). In these studies, the experimenters used the contralateral nature of the N2pc to their advantage by deliberately placing probable targets on specific visual hemifields. They found that when the most probable potential target was on the right side and the second most probable target was on the left side, there was a negative deflection contralateral to the most likely target followed by a ipsilateral deflection that was thought to be due to orienting to the second most probable target. This does not prove that search is always done in serial, but does provide strong evidence that difficult searches can be done in serial. In the context of MOT, this is interesting because a similar methodology could be used to determine whether targets are selected serially or in parallel.

There are number of important limitations to applying this technique to understanding the selection period of an MOT paradigm. The first is that although Woodman & Luck used multiple *potential* targets to index the implementation of a covert attentional search, unlike MOT tasks there was only one target. Although the N2pc data clearly shows that the potential targets were attended, there is no reason for a subject to hold any information about a non-target in mind once it has been identified as a non-target. It may be that the N2pc is therefore indexing *inspection* of potential targets rather than selection. Furthermore, it is not clear how attention moves during a simple (or pop-out) search and it would very difficult to adapt Woodman and Luck's procedure to address this question. This is because the paradigm depends upon the presence of distractors that are similar enough to targets that they require attention to definitively reject. By definition, the presence of these similar distractors eliminates the possibility of a pop-out search. One way around this problem would be to have multiple targets. The literature on visual search with more than one target is strikingly small given the intimidatingly large visual search literature, but at least one study have shown that searching for multiple targets may be categorically different than searching for a single target (Gibson et al 2000). They showed that people are incapable of completing a search when the task is to determine whether one or two targets are present and all items change locations every 107ms. On the other hand, Horowitz and Wolfe (Horowitz and Wolfe 1998) have shown that the same type of location change did not affect the search rate in a more standard search task with one target. It seems that when there is the possibility of more than one target, each target must be tagged and the tag must be held until a complete search is completed for the presence or absence of an additional target. When there is not enough time to complete this secondary search, performance drops to chance. It seems that search for a single target does not necessitate memory, yet the addition of just a single additional item makes the task unassailable without some form of memory. This may be taken as evidence that memory is necessary to perform the selection phase of an MOT task, but the targets in Gibson et al's task were very difficult to discriminate (mirror images of a horizontal '5') and it is unclear whether the task could be done without memory given simple targets such as blinking versus non-blinking items.

Although a great deal is understood about the neural mechanisms of selection, it is currently unclear from the existing literature whether items in MOT are selected in a serial or parallel manner. Although the targets tend to be easily distinguishable from distractors, it is not certain that even simple targets are selected in parallel when there are multiple targets. One way to approach this question would be use the N2pc as an index of selection. If all targets were lateralized, parallel selection would predict a single transient deflection that would increase in magnitude as the number of targets increased. In a centralized, unbalanced display (such as 2 targets on the left and 1 on the right), serial selection would predict two N2pc deflections of opposite polarity (similar to what was found by Woodman & Luck, 1999) while parallel selection would predict a single deflection contralateral to the side with more targets.

While the literature has predominantly used the N2pc as an index of selection, it may also be used to assess the capacity limitations of a selection process. To whit, our lab has recently found that in a lateralized MOT paradigm, the amplitude of the N2pc increases with then number of targets, but does not increase when subjects are that was a supraliminal number of targets. Amplitude increased from 1 to 2 to 3 items but did not increase from 3 to 5. A similar result was found when subjects were asked to simply enumerate the number of targets: once again N2pc amplitude increased up to set size 3, before reaching asymptote. On the other hand, when subjects complete a lateralized memory task, the N2pc is unaffected by the number of items in the display (Vogel and Machizawa 2004). One suggestion is that this apparent dissociation indicates that different processes are taking place during the initial phase of these three tasks. Whereas it is necessary to individuate to track or enumerate items, it may not be necessary to do so when selecting all items on the screen and consolidating the representation of as many as possible into visual working memory. Perhaps there is a capacity limitation on the number of items that may be individuated, but not the amount of information that may be initially parsed for late individuation. This implies information processing may be fundamentally changed through the presence or absence of distractors: The capacity limited process of individuation may take place offline (during the maintenance period in this example) in the absence of distractors, but must take place during initial selection in the presence of distractors. Preliminary data from our lab back up this claim by showing that when subjects are asked to enumerate items in the absence of distractors, there is no evidence of an N2pc set size effect. One piece of evidence against this interpretation is that N2pc amplitude was not affected in a VWM filtering task that included two targets and two distractors (Vogel et al 2005). However, closer inspection of this data suggests that N2pc amplitude may be higher for 4 items than 2 targets and two distractors (see

Figure 1-2). More importantly, in order to rule out the effect of distractors on the presence or absence of an N2pc set size effect it is necessary to have more than one condition with distractors.

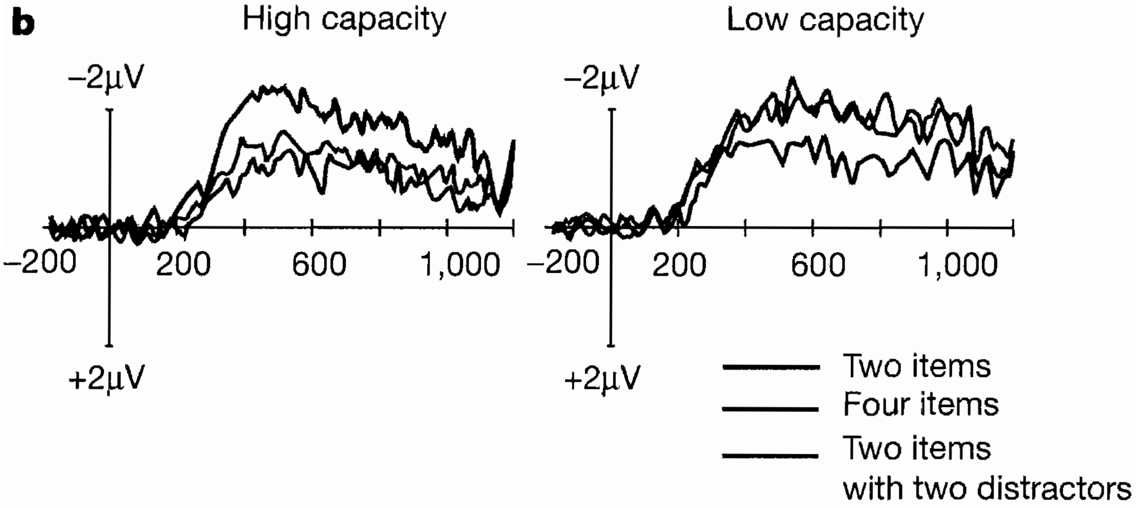

Figure 1-2: CDA Filtering Waveforms from Vogel et al., 2005

Divided Attention

The most well-known metaphor for attention is as a spotlight that enlightens a single area while moving about in an analog fashion. This metaphor was proposed by William James and the basic idea has been echoed by attention researchers ever since (James 1890). Eriksen and colleagues proposed a variation on the spotlight metaphor by suggesting that attention is more like a zoom lens with a variable size who fidelity increases as the size of the area attended decreases (Eriksen and St. James 1986; Eriksen and Yeh 1985). More recently researchers have suggested that these models may be overly simplistic: under certain circumstances attention seems to be constrained to a single foci, while in others it seems capable of splitting to multiple foci depending on task demands (Cave and Bichot 1999; Handy and Mangun 2000). On the surface, multiple object tracking seems to be a vivid demonstration that attention may be divided over multiple locations. However, it is still not clear whether:

A. Attention is simultaneously split between each of the objects being tracked

B. Attention rapidly switches between attended objects

C. Grouping of tracked objects allows attention to be implemented in a single malleable spotlight that encompasses each of the tracked objects.

This uncertainty mirrors the debate of over whether the attentional spotlight may be split when spatially cued to more than one location. Shaw and Shaw (Shaw and Shaw 1977) showed that target identification was enhanced when it appeared in one of two highly probable locations, but other researchers drew this conclusion into question by noting that the pattern of performance observed could be explained by shifting attention from location to location across, rather than within, trials (Posner et al 1980). Subsequently a different group of researcher found evidence for a split focus of attention by placing a target in one of two locations and interpreting the data on the basis of the assumption that the benefits of attention will decreases as the size of the area attended increases (Castiello and Umilta 1992). Response times increased as the areas attended increased, and the occurrence was independent across the two locations. They interpret this as evidence in favor of multiple attentional foci. They also reported RT distributions and found unimodal distributions that argue against a switching strategy. However critics argued that this analysis could not rule out a switching strategy and noted that the results may serve as a special case since there was always one location in each hemifield. If attention is truly capable of being implemented over multiple locations simultaneously, it should not be necessary to have the two locations in different hemifields.

Much of the skepticism regarding demonstration of multi-focal attention in selective attention is similar to competing claims about how people are able to simultaneously track multiple objects. In contrast to the current MOT literature however, a wealth of recent selective attention paper have convincingly addressed the most of criticism such that it now seems clear that people are able to simultaneously divide attention to multiple locations under certain circumstances. For example, Heinze and colleagues (Heinze et al 1994) measured p1 amplitude in response to task irrelevant probes as an index of attention in different locations. When subjects were attending 2 of four locations, they found that if there was an intervening location between attended locations, p1 amplitude was just as high in this position as in the attended positions. This pattern of results is incompatible with the multiple spotlight view. In response, Kramer and Hahn asked subject to perform essentially the same task: judge whether two targets in cued positions were the same or different (Kramer and Hahn 1995). The targets were in opposite hemi-fields and were separated by distractors that were either same or different than the targets. They found that the target/distractor relationship had no effect on performance, suggesting that the intervening distractors were not processed. This therefore implies that attention can be flexibly deployed and maintained on multiple discrete locations. Interestingly this effect held only when the targets and distractors did not have an abrupt onset but were revealed behind forward masks that disappeared. There was an effect of distractor congruency when all objects onset abruptly implying that subjects were unable to ignore the intervening stimuli in this condition. This is in line with the idea that it may take time to effectively 'split' the focus of attention and the system may be more susceptible to distractors in irrelevant locations before the spotlight has been implemented in the optimal shape (or shapes) (Yantis and Johnston 1990). Perhaps the most influential theory of split attentional foci is from LaBerge and Brown (LaBerge and Brown 1989), who cite evidence that attention may be implemented in multiple *gradients*. From this perspective, Kramer and Hahn's experiment may be excessively strict by asserting that split foci of attention are only demonstrated when the intervening stimuli have absolutely no effect on processing. Using gradients, it is only necessary to show that items that are between two targets are processed less effectively than either of the two targets. By probing subjects on their ability to identify targets in unexpected positions, Awh and Pashler (Awh and Pashler 2000) used this logic to demonstrate a split gradient of attention even when targets onset abruptly. Subjects identified 2 targets on each trial and the location of the targets was accurately cued on 80% of trials. On the remaining 20% of trials, one of the target locations was directly between cued locations. Performance in this location was much worse than in either of the cued locations. Given that this was the optimal location for a *single* spotlight, zoom-lens or gradient, this is striking evidence in favor of a split attentional focus. This effect disappeared when distractors and target masks were eliminated: performance in the uncued location was statistically equivalent to the cued locations. When there are no distractors, the normative strategy seems to be to orient attention in a broad focus rather than splitting the spotlight. This implies that splitting the spotlight has a computational expense such that the spotlight is only split when absolutely necessary. This may explain why it is sometimes difficult to find evidence in favor of a split spotlight: if the attentional system can figure out a way to do the task effectively without splitting the spotlight, the spotlight will probably remain unitary.

One limitation of these behavioral studies is that they have difficulty getting an unbiased measure of attention in the unattended locations. Measuring the amount of interference at intervening locations is an indirect measure. Having even a small percentage of trials where targets are placed in locations that subjects are not supposed to attend may encourage subjects to attend invalid positions, particularly if that task is not sufficiently difficult. The use of physiological measures such as fMRI and ERP have enabled researchers to assess attentional allocation without directly probing intervening locations. McMains and Somers (McMains and Somers 2004, 2005) demonstrated multiple spotlights of attentional selection in early visual processing areas (V1 and V2) by asking subjects to attend to a rapid serial visual stream of letters and numbers in 5 static locations arranged in an 'x' formation. On some trials, subjects were told to look for a match between the number in a single location in one of the corners of the 'x' and a target number. On other trials, the subjects were asked to simultaneously attend to two locations on opposite corners of the 'x.' Subjects showed a two distinct peaks of activation in early visual areas with activation in the fovea, which served as the intervening location in this experiment. A strict serial model predicts that it should take at least twice as long to identify two target, but threshold performance (d'=1) for one location was estimated to be 59ms when attending one location and just 67ms for two locations.

In some ways, searching for evidence for a split spotlight is similar to trying to find evidence in favor of a serial search mechanism. Just as a limited parallel model may be adapted to fit any data that seems to favor serial processing provided a liberal enough definition of 'parallel,' doubters of the idea of a split spotlight of attention can explain any data if they are willing to assume no limit for the speed of attentional switching. At some point, attention must be assumed to be moving so quickly that it may be thought of as 'functionally split' even if data that completely rules out attentional switching of infinite speed continues to prove elusive.

By using an electrophysiological measure of the allocation of attention, Matthias Müller and colleagues have been able to demonstrate that attention may be split for relatively long periods (3+ seconds) of time (Malinowski et al 2007; Müller and Hübner 2002; Müller et al 2003). By flickering items in different locations at different rates, the group has used frequency-coded steady-state visual evoked potentials (SSVEPs) as a way of measuring attentional allocation for multiple locations simultaneously. Subjects are asked to monitor 2 of 4 locations across the horizon of the visual field in search of simultaneous presentation of a target letter as letters in all 4 locations quickly cycle every 181ms. SSVEPs generate a fundamental frequency at each of the flicker rate and by demodulating the overall waveform, the authors are able to derive independent waveforms for each of the four flicker rates. Critically, the peak to peak amplitude of this waveform increases with attention: that is, if a subject is attending an object flickering at 20.3Hz the peak to peak amplitude of the 20.3Hz waveform will be larger than if the subject is attending a 15.2Hz object (Müller et al 1998). Accordingly, they found that the when subjects were asked to monitor two contiguous regions of space the SSVEP amplitude was increased for the two locations relative to the unattended regions. When asked to attend two non-contiguous regions the SSVEP amplitude was higher for both attended locations than for the unattended region between the two attended regions (Müller et al 2003). In this study, the two non-contiguous regions were always in opposite hemifields. A follow-up experiment replicated the result within a single hemifield only when the to be ignored region was in the upper hemifield. When the to be ignore region was in the lower region, SSVEPs were equivalent between the attended position in the lower region and the unattended (Malinowski et al 2007). In both experiments, performance conformed to the electrophysiological results. In fact, subjects in Müller's study (2003) were slightly better at identifying targets when they were in non-contiguous regions than when they were next to one another. They suggest that this may connote a difference between the sustained division of attention necessary in these experiments and the more transient split of attention necessary to perform the previously discussed cuing papers. However, the behavioral benefit for non-contiguous regions may also be driven by the fact that non-contiguous regions were confounded with hemisphere in this experiment. It is also interesting to note that the electrophysiological data reported in this experiment begins one second after the trial begins. Although the data from the first second of the trial is not reported, if the data from this time period does not support the idea of a split attention spotlight, it may imply that it takes a significant amount of time for the spotlight to be split. To my knowledge, there are no studies that have examined the time-course of splitting the spotlight electrophysiologically.

Hemispheric Effects of Selection and Tracking

Although there is overwhelming evidence that attention may be simultaneously split between two locations, almost all demonstrations of this effect involve dividing attention into one location on each side of the vertical meridian. **Every paper that has**

compared the ability to divide attention between versus within a given hemisphere have found that the ability to divide attention within a single hemisphere is either greatly weakened or non-existent. As previously discussed, Muller and colleagues found strong electrophysiological and behavioral evidence of a split spotlight when the locations were in different hemifields (2003). But, the same group found that when all locations were within a single hemifield subjects appeared unable to ignore an irrelevant location in the upper hemifield but were capable of doing so in the lower hemifield (Malinowski et al 2007). When Awh and Pashler cued two vertical locations performance was drastically worse than with locations on either side of the vertical meridian. Although there was still evidence in favor of a split spotlight, the size of the effect was greatly reduced (and nonexistent with nonalphanumeric targets). Although McMains and Somers (2005) found partial evidence in favor of a multi-focal attention when they asked subjects to attend to two RSVP streams in a single hemifield, there was also attentional modulation of intervening areas. This is in contrast to their previous demonstration of multi-focal attention across hemifields and the authors suggest that 'it may be easier to split attention across the hemispheres than to split within a hemisphere. It also suggests that there may be limit to the spatial resolution of attentional splitting,' (pg 682).

Hahn and Kramer (Hahn and Kramer 1998) replicated their earlier work (1995) within a single hemifield: demonstrating no interference from distractors in locations that intervened the target location in a single hemifield. Once again, this effect was not present when all items onset simultaneously after the location cue. One limitation of this study is that the eye-movements were not monitored. Given that the targets appeared after a 150ms cue for 100ms with no mask, it is possible that the subjects were foveating the target position by the end of the trial thereby drawing into question whether this is a demonstration of a unilateral split of attention. These studies also confound the presence or absence of an abrupt onset with difficulty. In both studies, an abrupt onset is avoided by revealing the targets by taking away pieces of the location cue. This location cue also serves as a forward mask, resulting in slower, less accurate performance in these conditions. Perceptual load theory has shown that task difficulty may have a large effect on the distribution of visuo-spatial attention (Lavie 1995; Lavie and Tsal 1994). This theory states that irrelevant distractors tend to be processed only when the task is easy. When the task his hard, presumably the processing limit is met and the less information outside the targets is processed. This may explain why some simple tasks have shown evidence of late-selection (Duncan 1980), while other more difficult tasks tend to favor early-selection (Kahneman and Treisman 1984). As Kraft and colleagues (Kraft et al 2005) have noted, this suggests that multi-focal attention is more likely to exist for difficult tasks while a single all encompassing spotlight might be used in more simple tasks. In this light, the confounding of onset and difficulty in Hahn and Kramer's experiments is more problematic. Kraft's group had subjects identify two of four letters in locations that were either one or two hemispheres. They also varied the difficulty of the target discrimination and found equivalent performance for contiguous and non-contiguous locations, but only when the task was difficult. When the task was relatively easy, subjects were slower when the targets were in non-contiguous regions than adjacent locations. They also found that performance was slower with two non-contiguous locations in the same hemifield than adjacent positions irrespective of task difficulty. This data is inconsistent with both the unitary and multi-focal view of attention. Kraft supports a modified model that postulates that attention can be split across hemifields but forms a unitary focus within a single hemifield and that additional attentional resources are available when attention is divided across hemifields.

Converging evidence for the idea of dual attentional systems that are confined to their respective hemifields was recently extended from selection (Sereno and Kosslyn 1991) to multiple object tracking. When subjects are asked to track objects within or between hemifields, there is a large benefit for tracking between two hemispheres (Alvarez and Cavanagh 2005). They found that almost twice as many objects could be tracked in two hemifields as can be in a lateralized display. These findings of decreased performance in a single hemifield are in contrast to demonstrations of equivalent bi-lateral and uni-lateral performance in visual search and memory storage and improved performance in searching bi-lateral arrays for split-brain patients relative to control subjects (Duncan et al 1999; Luck et al 1989). Finally, Delvenne (Delvenne 2005) found that VSTM was equivalent for items across and within a single hemifield for colored squares, but found that spatial memory was worse within a single hemifield than with bi-lateral presentation. While significant, this effect was not nearly as large as the one reported by Alvarez & Cavanagh (2005). All of this seems to imply that the ability to hold multiple locations in some form of memory is impaired when the locations are in a single hemifield. However, it is not yet clear whether the deficit found by Alvarez and Cavanagh is a manifestation of hemispheric limitations on the number of items that may be simultaneously selected, number of items that may be tracked or both. As previously pointed out, objects that are not initially selected cannot be tracked so, it's not clear whether this deficit is due to selection or tracking. In any case, this finding has important implications for the role of attention in MOT. First, it implies that if MOT is accomplished by grouping of objects rather than multi-focal attention (Yantis 1992), the grouping of objects must be much more difficult across hemifields than within a single one…and the deficit is so severe that Alvarez & Cavanagh suggest that if grouping is taking place it must be occurring separately in each hemifield. The data is completely inconsistent with a single, rapidly moving spotlight unless there is a substantial cost for shifting the spotlight from hemifield to another (Eriksen and Yeh 1985). Rather it suggests that there must be at least 2 attentional foci but cannot address the presence or absence of more than 2 attentional foci.

Divided Attention and MOT

The ability to track 4 items simultaneously has been used to argue that attention may be simultaneously split into four locations (Pylyshyn and Storm 1988). In light of the spatial attention literature this claim is dubious. In particular although there is strong evidence that attention may be split across hemifields, every paper that asked subjects to split the spotlight within a single hemifield has found evidence in favor of multi-focal attention either reduced or nonexistent. Given that tracking 4 items would necessarily involve splitting the spotlight in at least one visual hemifield, from this perspective it seems unlikely that subjects are capable of simultaneously selecting all targets with independent attentional foci. One might argue that it is not necessary to select all targets simultaneously in a MOT task because subjects are typically given 2 seconds to select the targets. This should be more than enough time to select what are essentially pop-out targets. However, if the attentional system is barely capable of simultaneously selectively attending 2 static locations in a single hemifield, what chance does the system have of simultaneously attending to 4 randomly moving objects? This question may explain why MOT is so interesting to attention researchers: almost all theories of visual attention in 1988 (and even today) would predict that people should not be capable of tracking 4-5 randomly moving objects, yet as has been demonstrated again and again, we are.

MODELS OF MOT

There are a number of different models that have been proposed to explain the ability to track multiple objects that I will overview briefly below. In some cases the models have been modified in subsequent papers and I will attempt to portray the most current version of the models available. The models include **attentional switching, preattentive indexes (FINSTs), grouping** and **multifocal attention.**

Perhaps the most intuitive model is **attentional switching**: where subjects rapidly switch from one target to another in a serial fashion. Two of the earliest and most influential MOT papers attempted to model this method of tracking and essentially refuted it as a tenable possibility given the extreme speed of switching necessary in order to successfully track any more than two objects (Pylyshyn and Storm 1988; Yantis 1992). The essence of the simulation was that the only thing that distinguishes targets from distractors in a standard MOT task is the position and motion of an object: the proposed serial mechanism is one that continually updates changing characteristics by sampling object locations. They assume that attention moves in an analog fashion from the position of one target to the next such that the distance traveled and time to travel are directly related. If a distractor is closer to the supposed location of a target than the target, it is assumed that an error is made and the distractor's position will incorrectly be encoded as a target position. The serial model therefore predicts that the number of objects, speed of objects and duration of tracking will all decrease accuracy because each manipulation increases the probability that a target will be confused with a distractor. Pylyshyn and Storm concluded that they could rule out a strict serial model based on unrealistically fast movement of the attentional spotlight, but could not rule out a mixed model that involved resource-limited parallel processing. Yantis performed a similar simulation and estimated that the spotlight velocity would need to move between 150 and 200 degrees/s to mimic actual subject performance. Furthermore this assumes that once the spotlight is in a location, it is able to instantaneously select the closest item. As Yantis pointed out, this is highly implausible and velocity necessary for the serial model to compete would have to be even higher than the stated estimate (Yantis 1992). To put this in context, Hallett estimated that the maximum velocity of smooth eye-movement to be about 100 degrees/s (Hallett 1986). It bears mentioning that both of these models seemed doomed to failure from the start--both search for a target in one of the least likely places for a moving target to be: where it *was* rather than where it was going. More recent experiments have shown that subjects are quite sensitive to trajectory information and velocity cues for the targets that they track (Fencsik et al 2006; Suganuma and Yokosawa 2006 …but see {Keane, 2006 #5904). This information could be used to estimate where an object is going rather than reducing it to X and Y coordinates that will have changed by the time the information has been encoded. If trajectory information is used to aid tracking, one would predict that increasing the tendency of an object to randomly change direction would make tracking more difficult. It could also help explain why tracking is more difficult in smaller areas: smaller areas mean more object collisions, meaning more trajectory changes thereby decreasing the amount of time between samples necessary to track effectively (Intriligator and Cavanagh 2001). Although it still seems unlikely that a strictly serial model could explain MOT, there is increasing evidence that at least some portion of the task is serial as will be discussed later.

Similar to the switching models, Pylyshyn's **FINSTs** model depends on indexes for each target, but once attached to a target these indexes (Fingers of INSTantiation) are thought to stick to the target, automatically updating without effort or attention (Pylyshyn 1989; Pylyshyn et al 1994b; Pylyshyn 2004, 2006; Pylyshyn and Storm 1988; Sears and Pylyshyn 2000). The MOT paradigm was actually created to test this theory. It has been subject to a great deal of criticism over the specifics of the model and as a result it has undergone a great deal of revision since it's inception. In the first stage of tracking, visual indexes are assigned to targets on the basis of bottom-up salience. This process is thought to be automatic and effortless, the only limitation being a limit of about 4 items due to architecture of the visual system. The mechanism is thought to be quite primitive and Pylyshyn refers to mechanism as part of the "early vision system" (Pylyshyn 1989). The underlying idea is that in order to understand a visual scene, the visual system must be able to simultaneously reference more than one item and that the associated pointer system allows multiple items to be perceived in unison (Pylyshyn et al 1994a). The index mechanism is thought to be separate from attention such that the references do not encode anything about the items that they index (such as identity) other than location. The indexes are "sticky," meaning that they automatically stick to whatever item they were instantiated to perceive without any attentional effort (Pylyshyn and Storm 1988). This is in stark contrast to the phenomenology of tracking: the task seems very attentionally demanding. Although Pylyshyn still maintains that tracking is automatic and effortless, more recently he has admitted that the task may be effortful because the indexes may need to be refreshed to prevent decay and the subject must stay vigilantly on task (Pylyshyn et al 1994a). A strict interpretation of the model also predicts perfect performance as long as the number of objects is below roughly 4. In fact, performance is not perfect at set size three and performance also decreases markedly as the duration of tracking increases (Oksama and Hyönä 2004). The FINSTs model explains this by noting that the preattentive model may be prone to "leaking" (Pylyshyn et al 1994a). In some ways, the FINSTs model has served as the punching bag for many of the MOT papers that have followed as they point out obvious discrepancies between the model's predictions and actual results such as performance decreasing with increased tracking time (Oksama and Hyönä 2004), faster tracking speeds (Liu et al 2005), interference with very general cognitive tasks like tone monitoring (Alvarez et al 2005) and working memory (Fougnie and Marois 2006). Even Pylyshyn has demonstrated that the selection mechanism is not preattentive by showing that indexes can be assigned through focused attention when necessary (Pylyshyn and Annan 2006). As Scholl has pointed out, it is now quite clear that attention is an important part of MOT, the question is whether any part of tracking is automatic (Scholl in press).

As an alternative to the Pylyshyn's model, Yantis (Yantis 1992) suggested the **grouping model.** This model assumes a single focus of attention that tracks the position of a single higher-order object (an ever-changing polygon) that encompasses each of the targets. He assumes two stages to the tracking process: an initial stage of group formation based on gestalt principles and a group maintenance stage that is effortful and demands attention. This maintenance process is assumed to be similar to mental rotation. Although he was able to convincingly demonstrate that manipulating the ease of grouping can have strong effects on task performance, he was not able to demonstrate that grouping is *the mechanism* that enables tracking multiple independent targets. For instance, he asked one group of subjects to attempt to group targets and while the other group was given not explicit instructions. Grouping subjects were better than the uninformed subjects for the first two blocks of the experiment, but the two groups were equivalent by the fifth and sixth block. Clearly, grouping is helping the subjects, but it is also clear that that the subjects that were not told to group are able to track multiple objects (albeit not as effectively as the grouping subjects). Does this mean that these subjects were spontaneously grouping on the trials where they were effectively tracking and that by the end of the experiment they were grouping on every trial? It is not at all clear that this is the case. Furthermore, it is very difficult for this model to explain Alvarez and Cavanagh's (2005) finding that people are able to track twice as many objects in bi-lateral arrays than unilateral. One explanation would be that grouping is easier across hemifields than within a hemifield but there is some evidence that low-level perceptual interactions (illusory contours) are stronger within hemifield (Pillow and Rubin 2002). It would be interesting to empirically test the effect of grouping within and across hemifields.

There are a number of different variations of **attentive tracking** model but the theme that unites them is that attention is divided so that multiple targets may be attentively tracked simultaneously. The most basic version of attentive tracking is multi-focal tracking: where a separate focus of attention is placed on each target. Maintaining and updating separate foci of attention is effortful. Therefore, this model correctly predicts that as duration increases task difficulty will increase (Horowitz et al 2007; Oksama and Hyönä 2004). In an extreme case, subjects were asked to track object for 10 minutes while periodically being probed as to whether certain objects were targets or distractors. In the absence of feedback, performance declined monotonically, from effectively tracking 3 objects in the beginning of the trial to 1.5 objects by the end of the trial. This is a strong argument against purely automatic tracking (Horowitz et al 2007).

The prior review of spatial cueing seems to suggest that it would be very difficult to simultaneously split attention over 4 or 5 locations, yet little work has been done to understand the mechanisms that would allow subjects to use multiple foci of attention in an MOT task. One exception is an oscillatory neural model of MOT that proposes a two-stage oscillatory model (Kazanovich and Borisyuk 2006). The majority of current neural networks that try to model attention can be thought of connectionist models that employ some sort of winner-take all strategy reminiscent of Duncan and Desimone's biased-competition model (Tsotsos 1995)(Itti and Koch 2000, 2001). These models can be thought of as modeling location-based attention; the connections must be recomputed each time an object changes position. Kazanovich and Borisyuk argue that oscillatory neural networks are more suitable to object-based attention since these models are primarily concerned with phase-frequency space irrespective of location.

In the first stage of their model, each object is assigned a specific oscillatory frequency label. Information about the object is coded via synchronous firing similar to the mechanism proposed by Raffone and Wolters (Raffone and Wolters 2001) to bind features to objects. Although the oscillatory label varies with time, in the second stage an attentional subsystem is assigned to each label enabling the system to differentiate targets from other identical objects. The processing in this model is purely parallel, but there is a limited phase space where different oscillators may operate simultaneously: increasing the number of objects increases the likelihood of inadvertent temporal synchrony. Increased movement (speed) makes it less likely that there will be time to fully process synchronization, leading to more errors. One limitation of the model is that while Oksama & Hyönä's found that performance decrement interacted with trial duration, the model predicts a linear decrease with increased duration. Some of this may be due to differences between how the model and human track objects. If a human is asked to track more objects that he is capable of tracking, he may elect to track a manageable subset of targets, while the model would try to track all items regardless. Although there is a growing literature of papers that relate increased neural synchrony to more accurate performance, nobody has looked at synchronous firing in MOT yet. This model predicts that parallel central operators fire synchronously to enable to MOT, but is very vague as to where these operators are located making a difficult theory to confidently confirm or refute currently.

Kahneman, Treisman and Gibbs (Kahneman et al 1992) proposed that tracking objects is accomplished via "object-files" that accumulate information about the objects as they move and change. This can be thought of as a specific type of attentive tracking. Unlike Pylyshyn's visual indexes, object files are thought to bind featural information beyond location to an object. According to feature integration theory, attention is necessary to bind features to objects (Treisman and Gelade 1980). The object file model of MOT can be thought of as the implementation of object based attention over time. From the object-based perception perspective, if multiple objects can be simultaneously attended, then information about each object would have to be updated independently through object files. The capacity for objects files is thought to be somewhere between 4 and 8 objects and it is thought to relate to visual working memory (Oksama and Hyönä 2004) (Kahneman et al 1992) .

Object-Based Attention

Broadly speaking, the purpose of attention is to select information that is relevant for behavioral goals. While the majority of this paper has been concerned with location-based attention where items in certain locations are given a competitive advantage over other items, it has been demonstrated that feature-base and object-based mechanisms can be employed to facilitate behavior (Yantis and Serences 2003). In fact, as outlined above, while the evidence for splitting the spotlight over different locations is quite mixed, all of these demonstrations have arguably dealt with location-based attention. A dominant theme of the object-based attention literature is that there is a benefit for processing information that may be grouped under the umbrella of a single object. One of the most striking demonstrations of this effect is that when subjects were asked to quickly identify two attributes of a pair of superimposed objects, there was a significant benefit when identifying two attributes from the same object rather than one from each. For example, subjects were worse at localizing a gap on a “C” and then the orientation of a superimposed line than identifying the orientation and texture of the line (Duncan 1984).

Vecera and Farah (Vecera and Farah 1994) raised the possibility that this result could also be explained by spatial, rather than object-based, selection. Furthermore in a later experiment, a larger within target benefit was found when the targets were superimposed than when the two were separate (Kramer et al 1997). Awh and colleagues demonstrated compelling evidence that these results may all be explained by acknowledging that attention is not unitary: there is a larger benefit when targets are superimposed because it allows both spatial and object-based attention to facilitate selection (Awh et al 2001). They found that when subjects knew what attributes they would be tested on, there was a substantial within object benefit and an effect of the distance between attributes. However, when told what to report after presentation of the objects, the effects of spatial attention disappeared and the within object benefit remained on the second attribute that was probed. This suggests that space-based and object-based attention are two distinct processes and that object-based attention may have a different, slower time-course than object-based attention. It appears that in these paradigms, the information-processing load is a function of the number of objects rather than the number of features. This is also true in the visual working memory domain, where increasing the number of features of an object does not increase the difficulty of detecting a change in any of the features (Luck and Vogel 1997) as long as increasing the number of features does not also increase the difficulty of perceiving a change (Awh et al., 2007).

While most of the object-based attention literature has been concerned with proving that sometimes attention is driven primarily by object attributes rather than location, more recently some papers have attempted to use object-based attention to learn more about what makes an object. Along these lines, several researchers have attempted to be better understand object-hood by the extending the ‘same-object advantage’ found by Egly, Driver and Rafal (Egly et al 1994). Briefly, these studies have found same-object advantages for probes within parallel lines (Avarahami 1999) as well as uniformly, but not non-uniformly connected objects (Watson and Kramer 1999).

The existence of object-based attention is important for understanding MOT because it is an indication that attention may operate on level above space alone; perhaps it is this special type of attention that makes it possible for people to do something that seems prohibitively difficult based on most of the spatial attention literature: track 4 or 5 objects simultaneously. In fact, there is evidence that object-based attention may be used even in cases where it is the disadvantage of the subject. In particular, when asked to track one end of a line, performance was much worse than when subjects were asked to track a single object that used an identical motion pathway (Scholl et al 2001). It seems that the subjects automatically tracked the entire line, which moved in a very complicated pattern, rather than the relevant part of the line. While spatial cueing paradigms have been used to demonstrate that object-based attention is sometimes move important than location-based attention, here subjects seemed unable inhibit an apparently automatic object-based attention mode of tracking. Another demonstration of the power of object-based attention in MOT occurs when the objects to be tracked are not cohesive (vanMarle and Scholl 2003). When subjects were asked to track objects that essentially poured from one location to another, performance was much worse than when they were asked to track boxes using the same speed and trajectory files. Control experiments (such as constantly morphing objects and objects that behave similar to a Slinky) suggested that the critical element that makes pouring objects more difficult is lack of cohesion. It appears that people are much worse at tracking non-cohesive substances than rigid objects that are moving at the same speed, once again emphasizing the role of object-based attention in MOT tasks. Finally, using the dot-probe technique, (discussed in more detail later) Alvarez and Scholl (Alvarez and Scholl 2005) found evidence that attention is concentrated at the center of lines that are being tracked and that the bias towards the center of the object increases as the length of the line increases. Probe detection was much higher for short lines than long lines. The ability to detect probes on the endpoint of the tracked lines decreased as the length of the line increased. Interestingly, this pattern of results held true for distractor lines as well. This seems to imply that attention naturally focuses on the center of objects and flows outward, even when the object in question is not being tracked. These studies demonstrate that MOT is a very powerful way to assess the temporal dynamics of object-based attention in ways that are not possible using simple cueing procedures.

Empirical Tests of the Models

Although many of the results relevant to differentiating between these models have been mentioned above, I will now briefly outline several experiments that test some of the predictions made by the models above. While much of the early MOT literature is devoted to characteristics that effect tracking difficulty, there has been a recent movement to use individual differences and dual-task paradigms to better understand the mechanisms that underlie MOT.

One of the most theoretically interesting aspects of MOT is that object identity does not appear to be automatically bound to object location (Pylyshyn 2004). Pylyshyn numbered each target at the start of every trial, then asked subjects to identify specific targets and found that subjects were terrible at doing so despite being capable of identifying the object category as either target or distractor. It seems they were able to track the set of targets without keeping track of the individual identities. As Scholl (in press) has pointed out, this result is in opposition with the FINSTs visual indexing model: the visual indexes are thought to serve as a reference from an object's identity to its location. In the context of Pylyshyn's original metaphor, these results are the equivalent of successfully tracking objects with a finger to refer to a particular object, but then not knowing what finger is pointing to which object. The finding also casts doubt upon the idea that multiple object-files are used to accomplish the task. One of the dominant characteristics of object-based attention is that features of a single object are bound together. This does not appear to be the case in MOT. Rather, the fact that group but not individual identities are encoded seems to imply that tracking is accomplished via an either multi-focal tracking or serial switching mechanism wherein no information differentiates one target from one another once they are all identical.

Multiple Identity Tracking

More recently Horowitz and colleagues (2007) found that tracking unique cartoon animals made the task easier, but there was still a cost when the subjects were asked to identify an item as a specific target rather than as a part of the target group. One critical difference between this experiment and Pylyshyn's experiment on object identity is that while Pylyshyn's group asked subjects to recall an arbitrary label for each target from the beginning of the trial, Horowitz's experiment used unique objects. Object identity was only available during the selection phase of Pylyshyn's experiment while it was available throughout the trial in Horowitz's experiment. The stark difference in performance seems to support Scholl's (2007) idea of *tracking in the present.* According to this theory, in a typical tracking task it is not necessary to store spatiotemporal trace of where an object has been because the only information that is necessary to identify an item as a target is that the same object was a target just a moment before. It may therefore be maladaptive to store individual identity information since it does not aid the primary task. The fact that object identity information does appear to be bound to specific targets in Multiple Identity Tracking (MIT) tasks appear to support this view and one interpretation of the data is that a single system is able to simultaneously bind target identity and location together. Although capacity in MOT is typically thought of as the number of items that can be tracked simultaneously, there is some evidence that the number may vary based on fidelity of the resolution necessary to successfully track an item (Alvarez and Franconeri 2007; Shim et al in press). It then follows that adding identity information to an object increases the information load, leading to lower tracking capacity. An alternative explanation is that there are two systems that work in concert during MIT tasks: one that tracks location information and an identity location binding system that requires focal attention (Wheeler and Treisman 1999). One piece of evidence in favor of the two-system explanation is that pairing each target with an identical distractor resulted in a reduced capacity in the standard condition but no effect in the specific identity condition (Horowitz et al 2007). Furthermore, when distractors are eliminated, the apparent capacity increases for the standard task, but ability to identify specific targets was unaffected. Clearly, if attention is thought of as a mechanism that resolves ambiguity the fact there is very large increase in tracking capacity further indicates that attention is an important part of the standard tracking task. An alternative explanation to this, however, it that the absence of distractors is confounded with the total number of items on the screen thereby making it ambiguous why there is an increase in tracking capacity. Previous studies have shown that apparent capacity decreases as the density of object increases (Intriligator and Cavanagh 2001) and as a function of the minimum distance between targets and distractors and targets and other targets (Shim et al in press).

Individual Differences

Although often treated as error variance and largely ignored, a great deal can often be gained from taken from closely examining individual differences (Cronbach 1957). By examining the individual variability in seemingly unrelated tasks, it is possible to use this approach to further constrain potential theories of underlying mechanisms. In particular, by running hundreds of subjects through a battery of tests, Oksama and Hyona (2004) were able to show that there is a great deal of variability in tracking ability and that this variability is significantly correlated with visual working memory and task switching. These correlations were quite low (task-switching r=.21; Corsi r=.22), but in a second experiment subjects were asked to perform a MIT task and both correlations were much stronger (task-switching r=.41; Corsi r=.40). Interestingly, operation span did not correlate significantly with MOT performance but did with MIT (r=.28) and mental rotation was negatively correlated with MIT performance (r=-.44). These findings contrast with Yantis's **grouping model**, which predicted a strong relationship between tracking ability and visuospatial processing ability. Subjects in this study were members of the Finnish air force and were subject to a strict selection criterion that excluded all but the highest 10% based on performance on standardized test. Given the extremely restricted range, it may be that these correlations would be substantially different in normal population. In the MIT experiment, air force subjects were once again used, but there was no pre-selection based on standardized tests. It is tempting to conclude that this resulted in stronger correlations, but the stronger correlations may also be due to the differences in tasks. According to Horowitz et al (2007), the need to bind identity to location may employ working memory thereby explaining the stronger relationship between the task and working memory. Unfortunately, the different populations used in these studies make it impossible to confidently evaluate relation of the mechanisms employed in MOT and MIT. It would be very interesting to see how the correlation with WM is effected by the constraints of the task. It is nonetheless important that visual WM and task switching correlate with MOT. The authors feel that this data supports a mixed model that includes visual WM to encode the locations of the targets and attentional switching from one target to another but only when deemed necessary based on a high level attentional mechanism. They suggest a parallel tracking system, vulnerable to decay and interference, that is buoyed by a serial system that refreshes target information as necessary.

Recently they have more fully articulated the specifics of their model, called MOMIT (model of multiple identity tracking) (Oksama and Hyönä in press). In regards to MOT, the most important aspects of the model are that spatial indexes are stored in VSTM and that visual attention continuously moves to reactivate this information in a serial fashion. Based on data they collected, they then created a formal model with two free parameters: binding capacity and speed of refresh. The model was very good at predicting results across three set sizes and speeds and yielded realistic estimates of roughly 4 items for binding capacity and 250ms for speed of refresh. As far as I can tell, Oksama & Hyönä's 2004 paper is the only MOT paper to seriously examine individual differences in an attempt to better understand tracking. This is an area that is ripe for follow-ups, from determining whether the strength of the correlation between tracking and WM increases when identity information is required, to examining whether an independent measure of selection (as measured by visual search or enumeration) correlates with tracking activity. This methodology may be useful for further constraining the mechanisms of tracking by manipulating tracking the requirements of a task and examining how correlations with well established measures such as WM are affected.

Dual-Task Paradigms

While Pylyshyn's original conception of MOT is that it is carried out by a primitive, preattentional mechanism, most of the more recent theories assume that attention is involved while disagreeing over what form of attention is necessary at what stage of processing. One way to address this question is to examine the effect on tracking performance when it is part of a dual task procedure.

Noting that MOT and VWM both have a limit of about 4 items, Fougnie and Marois (Fougnie and Marois 2006) asked subjects to track objects while holding items in VWM. They found that the amount of interference was smaller in this version of the task than when subjects were asked to perform two VWM tasks. They conclude that there are "distinct capacity limits for attention and working memory." There are a number of problems with this conclusion. First of all, by ‘attention’ the authors actually mean multiple-object tracking capacity. As this paper should make clear, MOT and attention are related but they are certainly not the same thing. Second, the main result of the paper (more interference for VWM-VWM than VWM-MOT) may be driven by the design used. In the VWM-MOT task, subjects encoded items into VWM, then did a tracking task, were then asked to respond to the MOT task and finally responded to the VWM information (same or different). In the VWM-VWM task, subjects encoded VWM1 then VWM2, then responded to VWM1 followed by VWM2. This design invites interference by asking subjects to respond in the same order that the object that the objects appeared. A critical component of change detection that is often overlooked is the internal comparison of the initial representation of information during the response phase (Awh et al 2007). Therefore, asking subjects to respond in the order of initial appearance invites output interference; it would be interesting to see if the level of interference for the VWM-VWM task would decrease if it used the same response order as the VWM-MOT task (where output interference should be decreased). Regardless of the order of response, this method of testing invited output interference since the subject must always decide upon a response while holding a memory load, the refer back to the memory load. A more direct way to probe processing capacity interference would be to randomly ask about either the first task or the second task. In this case, the subject must hold on to the information necessary to complete both tasks, but there should be less output interference with only one response. Despite these limitations, the final conclusion of this paper is perhaps not all that surprising: doing the exact same task twice leads to more interference than doing two tasks that are not the same. This implies that VWM involves some processes that are distinct from MOT and are subject to distinct capacity limitations but it also shows that there is a substantial amount of overlap in the two tasks.

One of the most surprising results in the MOT literature is that people seem capable of pausing an MOT trial (for roughly 333ms), performing a visual search task and then resuming the MOT trial (Alvarez et al 2005). The finding suggests that subjects are capable of storing the locations of the objects in spatial memory while attention is focused on search. This suggests that if there is a single attentional resource that underlies both mechanisms, it can be switched from MOT to search very efficiently. Although subjects were worse when asked to do both tasks than when there was no task in the blank interval where the visual search would have taken place, the cost of adding this second task was no greater than when (in a different experiment) subjects were asked to track while simultaneously doing an auditory tone monitoring task. Interference was much higher when the subjects were asked to do two versions of the same task in the dual task procedure. This suggests that while auditory tone monitoring and visual search appear to lead to interference on a general level, neither task appears to directly involve mechanisms that are vital to tracking. They also showed that it is possible to the search through and track spatially overlapping stimuli and to track items while searching through a set of non-overlapping stimuli. In each case, there were significant dual-task costs, but performance was better than would be expected if the tasks were mutually exclusive and their analyses suggested that the cost was due to a limitation on the central executive. To explain their results, the authors favor a parallel access model where memory of the location of each of targets is updated in parallel using attention. When a second task is added to tracking, it occupies some portion of attention, resulting in longer time between refreshing the locations of the targets and leading to more errors. The final conclusion is that tracking and search do not continuously draw on the same attentional resource but there is clearly some overlap. The question that remains is: what resources are shared by the two tasks and what resources are completely independent?

One way to better understand this question may be to introduce a third task: scene memory. Recently Wolfe and colleagues (Wolfe et al 2007) found that scene memory (the ability to identify a scene as having been previously viewed) was much worse when subjects were asked to perform a visual search during the initial presentation of the scene than when they were asked to perform an auditory tone monitoring task. On the other hand, when subjects performed an MOT task while several scenes were presented, performance on scene memory was no worse than when they performed a baseline central executive task (Junge et al, in press). Scholl suggests that visual search and scene memory are primarily concerned with identifying *what* the target is while MOT is primarily concerned with *where* the targets are located (2007). In this light it follows that there is more interference in the visual search/ scene memory task because they are both *what* tasks while the lack of interference in the MOT-search task is due to the fact that it a *where* does not interfere with the *what* task. Unfortunately, it is difficult to explain other well-known results using this logic: search is impaired when subjects are given a spatial working memory load, but not when given a nonspatial version of the same task (Oh and Kim 2004)(Woodman and Luck 2004). Although the spatial working memory task obviously requires more *where* information, it leads to more interferece with visual search than the nonspatial version of the task. These data suggest that visuospatial working memory and visual search require access to a common system for representing spatial locations. To further confuse the issue, as previously stated, nonspatial WM interferes with MOT (Fougnie and Marois 2006) (although not as much as two nonspatial WM tasks interfere with one another). However, unlike the other studies there is no baseline procedure in this experiment to determine if the amount of interference is greater than for an attentional task that interferes only at the executive level. As shown in the chart below the evidence from these dual task experiments seems to contradict itself from the What vs. Where perspective. One explanation is that dividing attention into two categories (What vs. Where) is too broad a distinction. That being said, it would be interesting to if MIT, which presumably relies on the '*what*' processing stream more than MOT, would interfere with visual search.

A special type of the dual task experiment is the dot-probe task. This technique has been used to infer the locus of attention during a number of visual search tasks (Cave and Zimmerman 1997; Cepeda et al 1998; Klein 1988) and has recently been used by a number of MOT researchers as a clever way to index the distribution of attention during the MOT task. The measure assumes that the ability to detect a faint probe may serve as an indication of the availability of attentional resources at a specific location. It has been used to provide evidence of inhibition of old items during visual search (Klein 1988; Watson and Humphreys 2000). Klein suggested that this mechanism would lead to more efficient search, but others have disputed this claim (Horowitz and Wolfe 1998) and the result has been difficult to replicate (Wolfe and Pokorny 1990).

The dot probe technique has been adapted to MOT by asking subjects to monitor search for a subtle probe item on some proportion of trials while simultaneously tracking. In some versions of the task (Alvarez and Scholl 2005) subjects are asked to respond to the probe as quickly as possible while in others the subjects are whether the probe was present each trial after identifying the tracked targets (Pylyshyn 2006). One question that remains unanswered for MOT studies that employ the dot-probe is the effect of the dot probe on performance. **There have been no studies that have compared MOT performance with and without the dot-probe.** A very strict interpretation of the FINSTs model might predict that the dot-probe would have no effect on tracking because tracking is a primitive, preattentional task. Every subsequent model has involved some form of attention during the tracking task and must therefore predict that adding an additional attentional task to perform simultaneously would reduce performance. One hint of this interaction is that a task (auditory tone monitoring) designed specifically to *not* interfere with visual attention resources necessary to track resulted in a marked decrease in tracking performance (Alvarez et al 2005). This is presumably due to interference on more centralized level of attention where modality is irrelevant. Given that the dot probe task is a demanding visual attention task, it is reasonable to assume that adding this task to a MOT task results in at least as much interference as auditory tone monitoring and visual search.

If the dot probe task does take attentional resources away from the tracking task, it is reasonable to ask whether adding this task changes how attention is distributed. To take an extreme example, if I ask you to track multiple objects while also searching for faint probes and the probes are always on distractors, you might expect that subjects would start paying more attention to distractors. The distribution of attention during tracking is one of the fundamental questions about MOT that has not yet been answered. In an attempt to address this question, Pylyshyn asked subjects to track 4 targets while simultaneously monitoring the screen for brief probes that occurred on half the trials. Probe detection was higher for probes that occurred on targets than distractors, but it's not clear if this is due to increased attention on the targets or decreased attention on the distractors. In an attempt to disentangle this problem, probes also occurred in empty space, where probe detection was even higher than when on a target. This may be due to the fact that probes in space are more arresting than probes on either a target or distractor because the object beneath the probe essentially acts like a meta-contrast mask. As Pylyshyn notes, "the problem of controlling for masking effects is ubiquitous in studies of probe detection where the difference between detection of probes on objects and in empty space is of interest." (pg6). Although some have addressed this problem by adding elements to the background that are physically similar to target and nontarget items and probing these background elements (Cepeda et al 1998), Pylyshyn chose to obtain a baseline measure of probe detection with no tracking as a method of circumventing this problem. This enabled him to essentially perform a multiple regression to predict performance if probe detection were equivalent at all locations. One problem with this prediction is that no attempt is made to verify the accuracy of the prediction. Given that both probe detection and tracking employ attention, it very likely that there is an interaction between the two, meaning this correction may be overly simplistic.

According to the resultant probe detection performance that has been statistically adjusted for baseline, target and space are treated equivalently with attention to distractor locations relatively inhibited. These results seem to suggest that a different mechanism is being employing in MOT than visual search. A number of dot-probe studies have shown that probe detection is faster and more accurate when it is in a target location than distractor location (Cave and Zimmerman 1997; Cepeda et al 1998). Here, if the baseline correction is to be believed, there is no evidence that attention is on the targets when targets are being tracked. Rather, attention seems primarily concerned with distractors; Pylyshyn suggests that inhibition is an important stage of scene segmentation. However, all of these conclusions rest upon the validity of the baseline correction procedure; if it is not a valid, a simple and more parsimonious explanation may explain the data: more attention is paid to targets than distractors. Although there is additional evidence for object-based inhibition of moving targets from the dot-probe literature (Ogawa et al 2002), it would be interesting to see if other, less invasive measures of attentional distribution found evidence for inhibition of the non-targets.

A simple way to avoid the questions of how attention is affected by a dual-task situation is to make the dot probe irrelevant to the task. Neural measure such as ERPs allow researchers to index attention to irrelevant probes. This technique has shown that both the P1 and N1 are enhanced when the probe is in the location of a previously displayed target relative to distractor locations (Luck and Hillyard 1995). The authors suggest that the P1 enhancement represents suppressed processing at nontarget locations while the N1 enhancement represents enhanced processing at the target location. The electrophysiological dot-probe technique has not yet been adapted for moving displays, but data from stationary visual search makes a number of clear predictions. According to all current theories of MOT, attention is either continuously split such that processing of all targets is enhanced or target position must be repeatedly updated. In either case, there should be an enhanced P1/N1 complex on targets relative to distractors. Disentangling whether this difference is affected by inhibition of the distractor once again requires that the space problem be addressed. One way to do so would be to have probes in stationary distractors that are otherwise identical to targets and distractors (Cepeda et al 1998). Pylyshyn and Ogawa's model of tracking via enhancement of targets and suppression of distractors would predict a smaller N1 for distractors than background. Importantly Pylyshyn has pointed out that static background positions differ from targets and distractors because they do not move (2006). It would be interesting to see if these predictions are verified.

GENERAL DISCUSSION

If we accept that attention is an important part of multiple-object tracking, we may be able to use this task to learn more about how attention operates in the real world. There are many common situations where successful completion of a task depends on dividing attention between multiple dynamic locations over time, such as monitoring traffic on crowded day and keeping track of your kids in a public pool. Under what circumstances is attention truly divided and how is this division accomplished? Multiple-object tracking studies allow researchers to address this question, but the answer is still unclear. One of the difficulties with using this paradigm to address the question of divided attention is that it is a complicated, multifaceted task. The ability to simultaneously select targets at the beginning of each trial may rely on a completely different mechanism than keeping track of the targets once they start moving. While the grand majority of the MOT literature has been concerned with the sustained tracking period while assuming that the people are able to initially select multiple targets with few errors, the spatial cueing and visual search literatures suggest that this process may quite difficult. For instance, Alvarez and Cavanagh's (2005) striking finding that tracking capacity is lower when tracking within, rather than between, a hemifield may be due to limitations in the ability to select multiple items in a single hemifield initially rather than anything to do with tracking. The two tasks are by nature embedded in one another such that there is no direct way to assess whether the inability to correctly identify what items were tracked is due to an error during selection or tracking. This is supported by the fact that location-based working memory (Delvenne 2005) and the ability to divided attention over multiple locations are both decreased in a single hemifield (Kraft et al 2005). An implicit assumption of most MOT tasks is that 2s is enough time to select any number of targets. One of the central suggestions of this paper is that simultaneous selection of multiple targets in distinct locations may serve as the bottleneck that determines that apparent capacity of the number of items a subject can track. Regardless of the duration of the selection period, it is unlikely that 20 targets could be selected individually and simultaneously. Pylyshyn assumes that during the selection phase blinking targets are automatically selected, but attentional capture due to object onset has been shown to have a limit of about 4 items (Yantis and Johnson 1990). Furthermore, it is likely that there are substantial individual differences in the number of items that may simultaneously capture attention. Pylyshyn and Annan (2006) have recently tried to address this problem by directly manipulating the difficulty of selection, but hopefully more work will be done to assess how the processes are related.

That being said, while it is dangerous to ignore the selection period of MOT, the idea that attention can be over a sustained period as people track multiple objects seems to be a striking demonstration of sustained divided attention and this finding alone may have driven the growing interest in this paradigm. Pylyshyn has thought of MOT as occurring via a primitive mechanism that automatically indexes the location of all targets in parallel, but ultimately, this conception of tracking has been overturned by the ever-increasing weight of evidence against it. For instance, the marked decrease in apparent capacity as tracking duration increases implies that there must be some element of serial processing that occurs during the tracking period rather than an automatic updating process that could proceed forever without error.

Although the purely serial model of multiple object tracking was dismissed in the first MOT paper (Pylyshyn and Storm 1988), this model of tracking is in many ways a straw man. Just as it is unconvincing when a model with too many free parameters effectively predicts behavioral data, it's not surprising that Pylyshyn and Storm's model failed because it essentially did not have enough parameters: the only information the model used was a table of each target's most recently sampled location. This model encodes nothing about variables that are known to effect behavioral tracking performance, such as ease of grouping (Yantis 1992), speed of motion (Fencsik et al 2006) and trajectory information (Horowitz, in press). An extreme example of the limitations of this model is that the model would be just as bad tracking three targets (amongst distractors) aligned in a straight line and slowly moving across the screen at a set velocity throughout the trial as tracking three targets the moved randomly. Adding these missing parameters may breathe life into serial models of MOT. **The general direction of the MOT literature seems to be towards a relatively simple system with very complicated rules about how to allocate finite resources (weakly parallel**

tracking supplemented by serial switching based on information gleaned from grouping, trajectory, and distractor information) rather than a system with almost unlimited resources and a very simple method of implementing these resources (parallel tracking based on location alone). As a result, even simple tasks such as auditory tone monitoring interfere with MOT because MOT is dependent upon the resources of a central executive to allocate limited attentional resources in the most efficient way. Tracking is correlated with a range of tasks such task-switching (Oksama and Hyönä 2004) and working memory that are associated with attentional control (Vogel et al 2005). All of this seems to indicate that rather than an automated system that sails along without much effort once it is set up, tracking seems to require sustained attentional control in order to constantly update target locations into some sort of durable representation.

Future Directions

While it is relatively simple to demonstrate convincing evidence that a process is parallel, it is very difficult to prove that a process is purely serial as different versions of the parallel model can account for almost data (Townsend 1990). Therefore it may not be all that surprising that there is converging evidence that tracking is not purely parallel. Some of the most recent models of MOT have concluded that MOT is accomplished through a hybrid mechanism that is a mixture of parallel and serial processing (Oksama and Hyönä 2004, in press; Scholl in press). One of the most pervasive findings in the recent visual attention literature is that attention is not a unitary phenomenon: it adapts to accomplish the task at hand in a variety of different ways. To extend this logic, it may be more accurate to think of attention as a group of mechanisms that all enable the organism to accomplish a task in face of distracting or irrelevant information.

Recent evidence has shown that the speed with which people are capable of confidently tracking items decreases as the number of items to track decreases such that people are apparently able to track one object moving at roughly 16 degrees per second and 8 items when the items move quite slowly (<1 degree per second) (Alvarez and Franconeri 2007). The authors interpret this to be strong evidence that tracking is a resource-limited attentive process rather that what they call a fixed architecture model. The speed data supports a resolution limited conceptualization of tracking because a slot model should not by affected the difficulty of tracking items as long as the number of items to be tracked is within the fixed capacity of the system: therefore the speed to track 1 item should be equivalent to track 3 items for most people. Alvarez and Franconeri hypothesize that the number of objects that may be tracked is governed by a flexible resource that adjusts the amount of resources used on each item to accommodate the difficulty of particular task. Therefore 2 fast items should take more resources than 2 slow items even though both are probably below what would normally be considered 'capacity.' This argument mirrors a similar situation in the visual working memory literature and in both cases that argument can be thought of as a slot versus resolution explanation for capacity limitations (Alvarez and Cavanagh 2004; Awh et al 2007).

Awh and colleagues have recently reported strong evidence that the same number of items is represented in working memory regardless of the complexity of the object (2007). The critical observation was that the observed reduction in apparent capacity with increased information load was confounded with increased comparison difficulty when testing what items were held in memory. Further, they noted that people who were good at representing many objects in working memory were not necessarily good at noticing small changes in the objects in memory: that is the number of 'slots' for storage was not correlated with the fidelity of the representation. This suggests that the resolution-limited tasks may tap into a different system than tasks where the primary difficulty is the number of items to be maintained. The idea of at least two separate systems that underlie working memory representation was further solidified in fMRI, where different areas of the brain have been shown to apparently represent different types of information in a simple change detection task. While activation in the **inferior** inner parietal sulcus (IPS) increased monotonically as the number of items increased until capacity was reached regardless of item complexity, activity in the **superior** IPS seemed to represent the complexity of the objects and reached asymptote representing fewer than 4 complex items (Xu and Chun 2006). Previous neuroimaging studies have shown that some of the same areas show increased activation as the number of items to track increase (Culham et al 2001; Jovicich et al 2001). It would be very interesting to see if the pattern of activation in MOT mirrors that of visual working memory. This would further solidify the idea that at least two mechanisms with different types of capacity limitations underlie the ability to track multiple items.

Target Individuation

An important caveat to the general approach of this paper is that the term 'selection' may mean different things in different contexts. For example, imagine two tasks: one where a subject must identify whether each trial has an "H" or and "L," and another where the task is to determine whether a given trial contains an "H." Is the mechanism for initial selection the same in both tasks? It has been my assumption throughout this paper that it is. Evidence from the non-human primate literature suggests that similar mechanisms are being employed, but to my knowledge the two tasks have never been directly compared in a single study in this literature. However, recently, Mazza and colleagues (2007) found that when subject were asked to identify or localize a target, the initial selection activity (the N2pc) was identical in both amplitude and latency. A subsequent lateralized component (very similar to the CDA) showed more contralateral activity for the identification trials.

A more relevant question to this paper is whether the same selection mechanism is employed when subjects are searching for multiple targets (as in the majority of the experiments discussed in the divided attention section of the paper) as when multiple targets must be selected so that they can be tracked. Recent work in our lab tentatively verifies this assertion. When subjects were asked to count the number of lateralized targets in an initial display identical to lateralized MOT experiments, we found that the selection activity (operationalized as the N2pc amplitude) increased from one to three targets, but did not increase from 3 to 5 targets. This all seems to suggest that the N2pc is a powerful index of the initial selection process that is a necessary antecedent to encoding information in variety of tasks. As such, the fact this component appears to have strong capacity limitations suggests that an initial selection bottleneck may help explain a wide range of behavioral results.

Conclusions

Research on multiple-object tracking is thriving, as interest in the areas and the number of papers published on the subject seems to grow exponentially. Part of the reason for the growing interest in the paradigm is it's surface level ecological validity. We can all think of instances where it is necessary to keep track of several objects simultaneously. One critical difference between tracking and most other attentional tasks is that it requires sustained attentional processing rather than a transient response to a single target. Although this paper has noted many similarities between aspects of the MOT paradigm and other well-known attentional paradigms such as visual search and spatial cueing, perhaps MOT is greater than the sum of its parts in terms of understanding how attention works in the real world.

If we think of MOT in terms of a two-stage process (selection and tracking), both stages are obviously important to success in the task, but communication between the two stages is also critical. As in many real life situations, the crucial component of success in this task may be the ability to allocate resources in timely, efficient manner rather that the total amount of attentional resources available to an individual. Hopefully, by understanding the attentional underpinnings of the MOT task, future research will be able to use this task to better understand how attention allows us to carry out complicated tasks such as MOT in the real world.

Thinking about MOT as an interactive amalgamation of multiple tasks may allow us to better understand the apparent disconnect between papers that suggest that primary capacity limitation in MOT is the number of objects to be tracked (e.g. (Oksama and

Hyönä 2004; Pylyshyn and Storm 1988)) or the resolution necessary to track items (Alvarez and Franconeri 2007; Shim et al in press). One clear prediction is that when the speed of objects is increased it may lead to an increase in activation in one area (possibly superior IPS) while increasing the number of objects to track may lead to an increase in another area (possibly inferior IPS).

An interesting question that this model raises is why tracking within is a single hemifield is so much more difficult than tracking across both hemifields (Alvarez and Cavanagh 2005). This effect may be due to a limitation in the ability to divide attention—a mechanism critical to the stage-one system. As outlined previously, there is a variety of evidence for the idea that it is more difficult to divide attention within a hemifield that across hemifields (Kraft et al 2005; Malinowski et al 2007). Yet, it is not currently clear whether the ability to localize targets is similarly affected. If not, it predicts that speed manipulations should not have as strong a laterality effect as the number of objects. Finally, if MOT taps into the same two mechanisms as the change detection paradigm, we might expect that an individual's ability to track many objects is unrelated to the precision with which the objects are tracked: someone who is good at dividing attention may not necessarily have high precision for judging the location of targets.

CHAPTER II

NEURAL MEASURES OF INDIVIDUAL DIFFERENCES IN SELECTING AND TRACKING MULTIPLE MOVING OBJECTS

This chapter was previously published with Edward K. Vogel in the *Journal of Neuroscience*.

INTRODUCTION

Common tasks such as driving a car in traffic are dependent upon our ability to simultaneously attend multiple objects as they move about in the visual field. This ability is known to be highly limited such that most individuals can track only about four moving objects simultaneously (Pylyshyn and Storm, 1988; Scholl et al., 2001; Cavanagh and Alvarez, 2005). Multiple object tracking is thought to require at least two components of visual attention: a transient selection process that initially determines which items will be tracked, and a sustained process that keeps an updated representation of each object as it moves amongst identical distractors (Yantis, 1992; Alvarez and Cavanagh, 2005; Pylyshyn and Annan, 2006). Although previous imaging and neurophysiological studies have reported neural correlates of both selection (Woodman and Luck, 1999; Buschman and Miller, 2007) and tracking (Culham et al., 1998; Culham et al., 2001; Jovicich et al., 2001) these two mechanisms have typically been studied in isolation. Moreover, the relationship between these types of activity and the capacity limitations that constrain attentional tracking has not been demonstrated. As a result, it is still unclear whether capacity limits in tracking are due to limitations of initially selecting multiple targets amongst distractors, sustaining attention to the moving targets, or some combination of these two factors.

Recent neuroimaging studies of attentional tracking have reported that the Intra-parietal sulcus (IPS) and the Superior Frontal Sulcus (SFS) show significant load-dependent activations, such that as the number of items tracked increased, BOLD activation in these areas also increases (Culham et al., 1998; Culham et al., 2001; Jovicich et al., 2001). However, it is still currently ambiguous what these load-dependent activations actually reflect. For example, increases in tracking load are necessarily accompanied by increases in task-general processes such as effort and arousal, which makes it difficult to determine whether the increasing cortical activity is the result of more attended object representations or simply due to the subject expending more effort when tracking more items. A further ambiguity of these studies regards which component of attention underlies these load effects: is it driven by the initial selection of the targets, or does the activity reflect the sustained attention to the items as they move about the visual field? Because the previous neuroimaging studies of tracking all used FMRI, the poor temporal resolution of the technique makes it difficult to disentangle the quick sequence of attentional events in this task.

In the current study, we sought to establish distinct electrophysiological measures of target selection and sustained attention during a tracking task as a means of determining which of these components of attention is the principal limiting factor in tracking performance. To do this, we recorded event-related potentials (ERPs) from subjects while they performed a multiple object tracking task (MOT) in which they were presented a bilateral array of objects and were instructed to attend a subset of objects in a single hemifield. The advantage of this bilateral stimulus design is that it allows us to isolate the lateralized effects of attention from the bilateral perceptual response evoked by the onset and motion of the stimuli in the display. There are several candidate ERP components that have been observed in lateralized attention tasks that may play a role in both the selection and sustained attention to the moving targets in a tracking task. In terms of initially selecting the targets, we expect to observe an N2pc component, which is a transient contralateral negative wave appearing at approximately 200ms post-stimulus over posterior electrode sites (Luck and Hillyard, 1994; Eimer, 1996). This component has been shown to reflect the selection of targets amongst distractors in visual search tasks and has been localized to generators in extrastriate cortex, including V4 and posterior portions of inferior temporal cortex (Luck et al., 1997; Hopf et al., 2000; Hopf et al., 2002; Hopf et al., 2006). Moreover, it appears to be functionally equivalent to another component labeled the "early directing attention negativity" (EDAN) (Harter et al., 1989; Van Velzen and Eimer, 2003). At more frontal electrode sites, it is also possible that we would observe an ADAN (anterior directing attention negativity), which is a transient negative wave (350-500ms) that is thought to reflect control signals in prefrontal cortex involved in orienting attention towards the general location of an upcoming target (Harter et al., 1989; Nobre et al., 2000; Simpson et al., 2006).

In terms of sustained attention during tracking, there are two known lateralized components that have been shown to be sensitive to the orienting of attention towards a single hemifield. The first component, the LDAP (late directing attention positivity; e.g., Hopf and Mangun, 2000), is highly similar to the ADAN but appears over more posterior and temporal electrode sites and has a positive voltage. Like the ADAN, it also appears to reflect the orienting of spatial attention towards a hemifield following a centrally presented spatial cue in anticipation of an upcoming target. Though it is a sustained wave, this component does not appear to reflect attentional processing of the targets per se because it is not sensitive to the task demands imposed by the targets, and typically has expired prior to target onset (Hopf and Mangun, 2000). By contrast, the second component, the CDA (contralateral delay activity) appears to be a good candidate for sustained attention to targets during tracking because it has been shown to be finely sensitive to the number of objects that are currently being maintained in visual working memory as well as being sensitive to the capacity limits of this system (Vogel and Machizawa, 2004; Vogel et al., 2005; Jolicoeur et al., 2006; McCollough et al., 2007; Mazza et al, 2007; Woodman & Vogel, 2008). This component is a sustained negative wave over posterior contralateral electrode sites, and likely stems from a source in the lateral intraparietal sulcus in the parietal cortex (Todd and Marois, 2004, 2005; Xu and Chun, 2006).

While the precise role of visual working memory during attentional tracking tasks is currently unclear (see e.g., Fougnie and Marois, 2006), there are at least two lines of evidence that suggest that similar mechanisms likely underlie the performance of each type of task. First, there is considerable evidence that maintaining object information in visual working memory requires sustained spatial attention to the locations of the remembered items (e.g., Awh et al., 2000). Second, the capacity of attentional tracking (~4 items) is highly similar to the capacity of visual working memory (3-4 items) and at least one study has found evidence that an individual's memory capacity positively predicts his or her tracking capacity (Oksama and Hyönä, 2004). On the basis of these previous findings, it appears highly plausible that similar capacity-limited mechanisms underlie performance of both tasks, and thus we expected that the CDA component would be observed while subjects sustained attention upon the moving targets during the tracking task.

MATERIALS AND METHODS

Participants

Neurologically normal participants (12 In Experiment 1, 15 in Experiment 2, 18 in Experiment 3, 33 in Experiment 4, and 18 in Experiment 5; Age range 18-31) from the Eugene, Oregon community gave informed consent according to procedures approved by the University of Oregon institutional review board.

Stimulus Displays and Procedure

All stimulus arrays were presented in regions subtending 5.1 X 6.0 degrees (or 4.2 X 4.9 in the small area condition of Experiment 2) that were centered 3.2 degrees to the right or left of a central fixation cross (see Figure 1a). The inner boundary of each movement area was lateralized 0.5 degrees of visual angle to the left or right of fixation to minimize the impact of small movements of eye position. Each trial began with stationary squares that subtended .38 X .38 degrees of visual angle and were displayed in both left and right regions for 500ms. A subset of the squares was red in one hemifield, and green in the other; the remaining items were black. In each experiment, half of the subjects were asked to track red squares while the other half were instructed to track the green squares. In Experiments 1, 2, and 4 there were 8 total squares in each hemifield while Experiments 3 and 5 had 10 items to insure that at least 50% of the boxes were distractors in each trial. After 500 ms, targets (red and green items) changed to black and all items began to move for 1500ms. When motion stopped, one square was drawn in red in one hemifield and another became green in the opposite hemifield which initiated a 2000ms response window for the participant. The probed square was one of the original targets on 50% of trials and was a randomly selected distractor within the hemifield on the remaining trials. Each participant completed 240 trials per condition in the first experiment, 200 in the second experiment, 160 in the third experiment, and 224 in the final two experiments.

Motion Parameters

In Experiments 1 and 2, the direction of motion varied randomly and the boxes bounced off the border of the viewing area, but not off of each other (brief occlusion possible). The speed of motion varied from .25 to 1.86 degrees of visual angle/ second with an average of about 1 degree per second. Motion trajectory was linear and changed at random intervals or when the object made contact with (invisible) outer barrier of the viewing area. Several of these parameters were modified slightly in Experiments 3 and 4. In particular, the size of the squares was increased to .7 degrees and the squares bounced off (no occlusion) of each other when they made contact. Furthermore, the average speed in these experiments was increased to 1.58 degrees/ second. These changes made no observable difference in the ERP data or behavioral performance between experiments.

In Experiment 5, in a separate behavior-only session participants were asked to track 3,4 or 5 objects that were distributed across both hemifields for 10 seconds following a 500ms cue that was identical to the cue in previous experiments. In the ERP session of this experiment, subjects performed a tracking task that was identical to that used in Experiment 4.

Measuring Tracking Capacity

We used Scholl's (2001) formula to derive the effective number of objects tracked: $M = n(2P-1)$. Where M is effective number of objects tracked, n is number of targets and P is the empirically observed proportion of correct answers.

Electrophysiological Recording and Analysis

Event-related potentials (ERPs) were recorded in each experiment using our standard recording and analysis procedures, including rejection of trials contaminated by blocking, blinks or large (>1 degree) eye-movements (see (Vogel et al., 1998; McCollough et al., 2007). We recorded from 22 tin electrodes mounted in an elastic cap (Electrocap International) using the International 10/20 System. 10/20 sites F3, FZ, F4, T3, C3, CZ, C4, T4, P3, PZ, P4, T5, T6, O1 and O2 were used along with 5 non-standard sites: OL midway between T5 and O1; OR midway between T6 and O2; PO3 midway between P3 and OL; PO4 midway between P4 and OR; POz midway between PO3 and PO4. All sites were recoded with a left-mastoid reference, and the data were re-referenced offline to the algebraic average of the left and right mastoids. Horizontal electrooculogram (EOG) was recorded from electrodes placed approximately 1 cm to the left and right of the external canthi of each eye to measure horizontal eye movements. To detect blinks, vertical EOG was recorded from an electrode mounted beneath the left eye and referenced to the left mastoid. Subjects with trial rejection rates >25% were excluded from the sample.

Contralateral waveforms were computed by averaging the activity recorded over the right hemisphere when subjects tracked items in the array at the left side of screen. Contralateral tracking activity was measured at posterior parietal, lateral occipital, posterior temporal, parietal and occipital electrode sites as the difference in mean amplitude between the ipsilateral and contralateral waveforms. We used two measurement windows: 200-300ms after the onset of the items for the N2pc analyses, and 800-1200ms (300-700ms after motion onset) for the tracking analyses. Differences in scalp topography were tested by normalizing the data for each component following the procedure described by McCarthy & Wood (1985) and testing for the interaction between electrode position and time window (i.e., 200-300ms vs 800-1200ms). The EEG and

EOG were amplified with a SA Instrumentation amplifier with a bandpass of 0.01-80Hz and were digitized at 250 Hz in LabView 6.1 running on a Macintosh.

Eye Movements

Any trials containing either a blink or eye-movement were excluded from further analysis. The horizontal electrooculogram (EOG) for Experiment 1 is plotted as a function of the cued hemifield in Figure 2-2. There was a small but significant tendency for eye position to drift towards the attended side during the latter half of the trial ($p<.05$). Though, the magnitude of this deviation was not influenced by the number of targets being tracked, nor was it related to tracking performance. Further, this deviation from fixation was quite small: the mean amplitude of this EOG activity was 2.5 µV, which corresponds to an eye-movement of less than 0.16 degrees of visual angle from the fixation point (Hillyard and Galambos, 1970). Given that the area that the boxes moved within was lateralized by a minimum of more than 0.5 degrees from fixation, it is unlikely that these small drifts in fixation affected our data.

RESULTS

Experiment 1: ERP Correlates of Selecting and Tracking Moving Objects

On each trial, subjects were presented a bilateral array containing six squares in each hemifield (see Figure 2-1). For the first 500ms of each trial (cue period), the objects were stationary with a subset of the items in a given hemifield drawn in red (the targets) and the remaining items drawn in black (the distractors). Green items appeared at the start of each trial on the unattended side. These items were photometrically isoluminant and equal in number to the red target items on each trial. Half of the subjects tracked red items while the others tracked green. After 500ms, the red and green items changed to black and all of the objects began to move amongst each other in random directions within the hemifield for 2 seconds; at that point, the items stopped moving and one item turned red. Subjects were instructed to attentionally track the targets and pressed one of two buttons to indicate whether the final red item was one of the targets or not. We time-locked the ERPs to the onset of the cue array and recorded throughout the duration of the trial so that we could observe both the transient selection of the targets during the cue period as well as the sustained attention response during the tracking period. In Experiment 1, we asked subjects to track one, two, or three targets on each trial so that we could determine whether the activity was modulated by the number of tracked items.

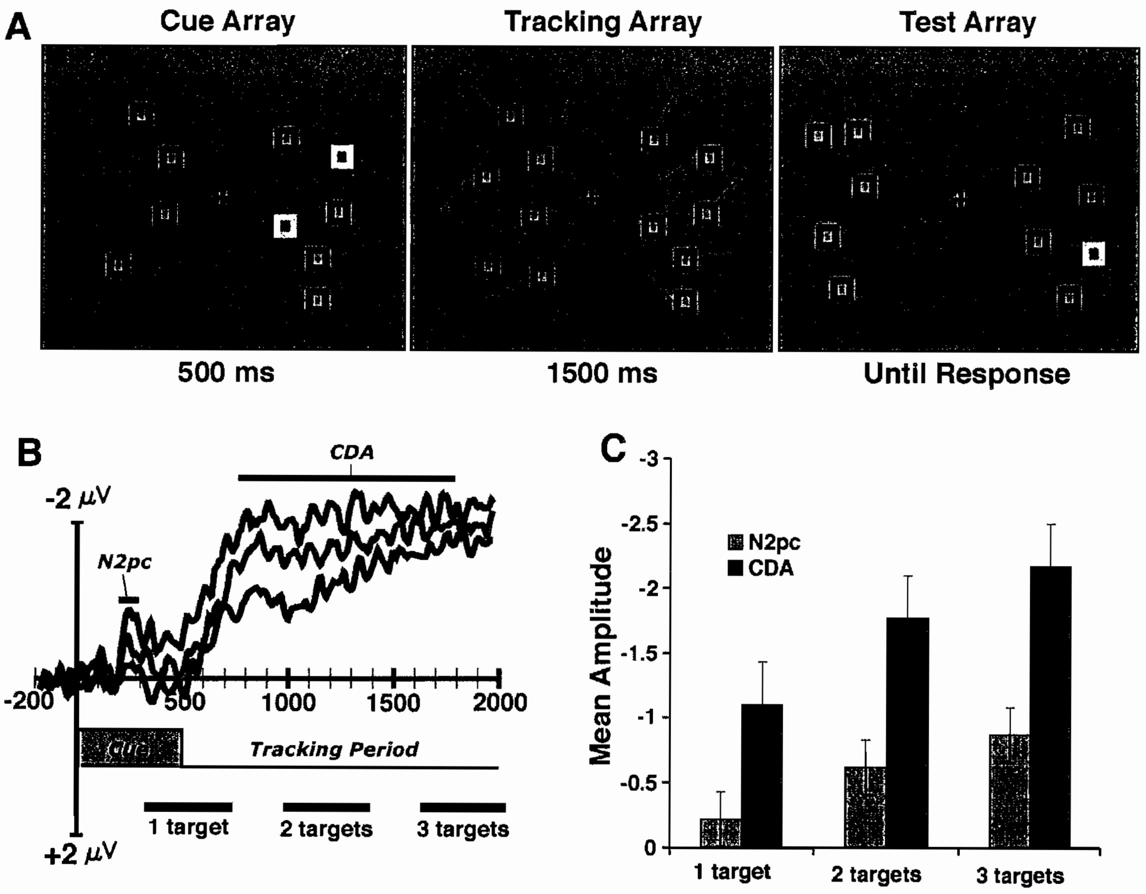

Figure 2-1:Experiment Paradigm and Experiment 1 Results (A) ERP multiple object tracking task. Participants tracked either red or green boxes while maintaining central fixation. In each experiment, the number of total objects (including distractors) was held constant while the number of target boxes varied across trials. (B) ERP difference waves (Contra minus Ipsi) for Experiment 1 from the average of posterior electrode sites (PO3/PO4; P3/P4; OL/OR; T5/T6). Negative voltage is plotted upwards. Note that all ERP waveforms in this and subsequent figures reflect correct trial performance. (C) Mean amplitude during the selection (200-300ms) and tracking periods (800-1200ms) as a function of the number of target items.

200ms following the onset of the cue array, we observed a transient negative-going wave over the hemisphere that was contralateral to the attended hemifield. This activity was followed by a larger and sustained contralateral negative wave that began shortly after the tracking period started and persisted throughout the course of the trial until the test was presented. As shown in Figure 2-1, the amplitude of both of these waves was strongly modulated by the number of target items; increasing the number of targets resulted in substantial increases in amplitude (3 targets > 2 targets > 1 target; all p's < .01). Moreover, the amplitude of this activity was highly sensitive to whether or not the subject performed the tracking task correctly: both waves showing large, significant decreases in amplitude on error trials relative to correct trials (both p's < .01). This indicates that both waves reflect processes that are necessary antecedents to correct tracking performance. Figure 2-2 shows the distribution of these waves across each of the lateral recording sites. The transient activity during the selection phase was primarily centered over posterior electrodes with a maximum over lateral occipital electrodes (OL/OR). During this selection period there was no significant lateralized activity observed over frontal electrodes (F < 1). The sustained activity during the tracking period was more broadly distributed over the posterior electrode sites with a maximum over posterior parietal electrodes (PO3/PO4). This activity was also observed over frontal electrode sites (F3/F4), though the contralateral effect at these sites was not significantly modulated by the number of tracked targets (F=2.3, p >.10).

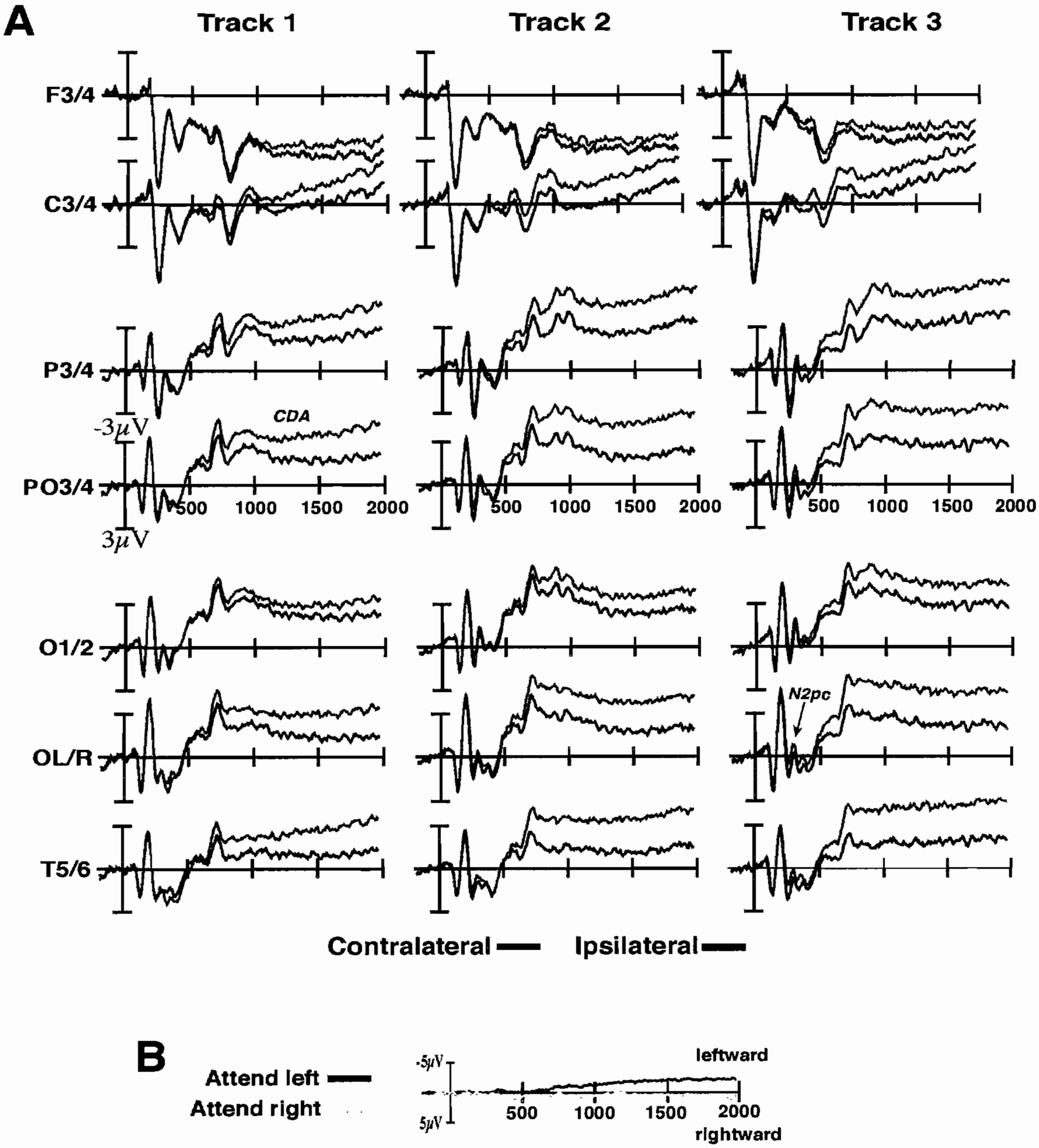

Figure 2-2: Contralateral and Ipsilateral Waveforms(A) Contralateral and ipsilateral activity in response to the three tracking loads in Experiment 1 across all frontal, parietal, and occipital electrodes. Waveforms were time-locked to the initial appearance of targets and motion began at 500ms. (B) Grand-averaged horizontal EOG waveforms for attend left and attend right trials.

The transient wave during the cue period appears to be the N2pc wave which, as described in the Introduction, has previously been shown to reflect the selection of targets amongst distractors in visual search tasks (Luck et al., 1997; Hopf et al., 2000; Woodman and Luck, 2003). By contrast, the large sustained wave during tracking appears to be the contralateral delay activity (CDA) that we and others have shown reflects the number of active object representations held in visual short term memory (VSTM) (e.g., Vogel and Machizawa, 2004). Together, the N2pc and CDA waves appear to index two critical components of attentional tracking: the initial selection of the target objects during the cue period (N2pc); and sustained attention towards the target representations as they move about the hemifield (CDA). Although the N2pc and the CDA were both modulated by the number of targets, we found that these two waves have distinct scalp distributions yielding a highly significant electrode position by time window (200-300ms vs 800-1200ms) interaction ($p<.01$, See Methods): with the N2pc showing a more ventral distribution than the more dorsal CDA. This finding supports a previous demonstration of distinct scalp distributions for these two components in the context of a working memory task (McCollough et al., 2007). Together, these results suggest that while there appears to be a tight coupling between object selection and sustained attention towards the targets, they may reflect the output of distinct cortical areas.

Experiment 2: Spatial Extent of Attention or Number of Objects?

Although the amplitude of both the N2pc and CDA in the first experiment increased as a function of the number of targets, it is possible that this increase is simply due to the required spatial extent of the target area rather than reflecting the increasing number of targets selected and tracked during the trial. That is, as the number of target items increases, there is also potential for a corresponding increase in the area of the attentional window or "spotlight" that encompasses the targets and this may be what caused the increases in amplitude in the first experiment (e.g., Eriksen and St. James, 1986; Hillyard et al., 1998). To test this alternative, in the second experiment we directly manipulated the amount of area required to track the targets. Subjects tracked two or three targets that either encompassed a large area or a small area within the hemifield. We found that while the amplitudes of both the N2pc and CDA were again significantly modulated by the number of targets (both p's < .01), there was no significant effect of area on amplitude for either wave (both F's < 1; see Figure 2-3). We did however find a significant effect of area on behavioral tracking performance, where performance in the small area conditions was significantly poorer (~10%) than in the large area conditions (p < .01). These results are consistent with previous studies that have shown that displays with a high density of items result in more difficult tracking and poorer performance (e.g., Intriligator and Cavanagh, 2001). It also helps to confirm that our manipulation of area was substantial enough to observe a significant behavioral effect. Indeed, the lack of an amplitude modulation by area also argues against the hypothesis that the amount of general effort or difficulty required to track more targets is the cause of the observed increase in amplitude. That is, despite the small area condition being significantly more difficult than the large area condition, there was no concomitant rise in amplitude for either the N2pc or the CDA. Nonetheless, it is important to note that the apparent dissociation between behavioral performance and CDA amplitude in this experiment may be due to a limitation of our measure. In particular, it is possible that poorer behavioral performance in the small area condition is due to the subjects inadvertently tracking distractor items that were mistaken, or swapped, for target items during the course of the trial due to the closer proximity of targets and distractors. This scenario would lead to a decrease in behavioral performance because the wrong items were being tracked. However, it would predict no change in CDA amplitude because the same total number of items are being tracked on the trial. Specifically, the limitation of this component is that it provides an index of the number of objects currently being tracked irrespective of whether or not they are targets.

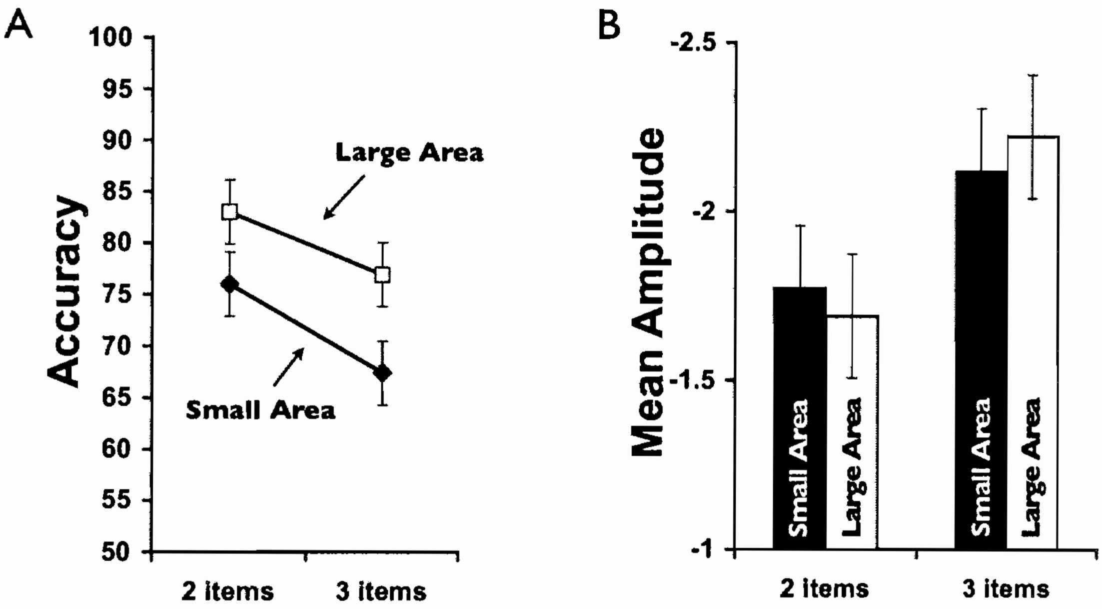

Figure 2-3: Experiment 2 Results (A) Behavioral performance in Experiment 2 showing significant main effects of both area and number of items. (B) Mean amplitude of CDA activity in Experiment 2. While there was a significant main effect of number of targets, area had no significant effect on amplitude of either the N2pc or the CDA.

Experiment 3: Sensitivity to Behavioral Tracking Limitations

The results of the first two experiments are consistent with the proposal that the amplitude of both the N2pc and the CDA reflects the number of targets being selected or tracked, respectively. However, to strengthen this claim it is necessary to demonstrate that this activity is indeed sensitive to the known behavioral performance limitations associated with attentional tracking. Therefore, in the third experiment we measured these two waves under a task condition that is likely to exceed the subject's tracking capacity so that we could determine whether this activity is sensitive to these performance limitations. Indeed, this has been a significant limitation of previous neuroimaging studies examining tracking-related load effects because they have not tested whether the observed activity continues to increase when the number of targets exceeds capacity. In addition, by examining a wider range of target array sizes, we can begin to examine whether these two types of activity are sensitive to differences across individuals in tracking ability. In this experiment, subjects tracked one, three, or five targets on each trial. In this experiment, all trials contained 10 items so that 50% of the items were distractors when subjects tracked 5 items. We divided subjects into high capacity and low capacity groups on the basis of a median split of their behavioral tracking capacity (see Methods). Figure 2-4 shows the N2pc and CDA waves for each target array size for the high and low capacity groups. As can be seen in the figure, both groups showed an increase in amplitude for both the N2pc and the CDA from one to three targets (low capacity: both p's $< .05$; high capacity: both p's $< .001$). However, the two groups diverged greatly when tracking five items. The amplitude for the high capacity group when tracking five items remained equivalent to that of tracking three items (N2pc: $F<1$; CDA: $p > .15$). Thus, when given more items than they could track, the high capacity subjects appeared to be able to continue to track their limit of objects (i.e., ~3 items). However, for the low capacity group, the track five amplitude decreased significantly below the three item level and was equivalent to that of tracking a single item (N2pc: $p < .001$; CDA: $p < .05$). While the precise cause of this amplitude decrease is currently unclear, it does appear to reflect a consistent pattern across all subjects dependent upon their specific tracking capacity. That is, there was a significant negative correlation between an individual's tracking capacity and the amount of decrease between three targets and five targets ($r = -.60$ N2pc; $r = -.56$ CDA; both p's $< .01$), such that as tracking capacity increased the amount of amplitude drop decreased. In summary, the results of Experiment 3 provide further evidence that the amplitude increases of the N2pc and CDA are the consequence of the number of items that are currently being selected or tracked. In particular, these results demonstrate that the amplitude is not simply driven by the amount of cognitive load required to perform the task because the amplitude of each component reached an asymptotic limit at roughly 3 items, even though the amount of cognitive load continued to increase when the subjects attempted to track 5 items. Thus, the properties of these neural mechanisms appear to be finely sensitive to the known capacity limitations associated with attentional tracking.

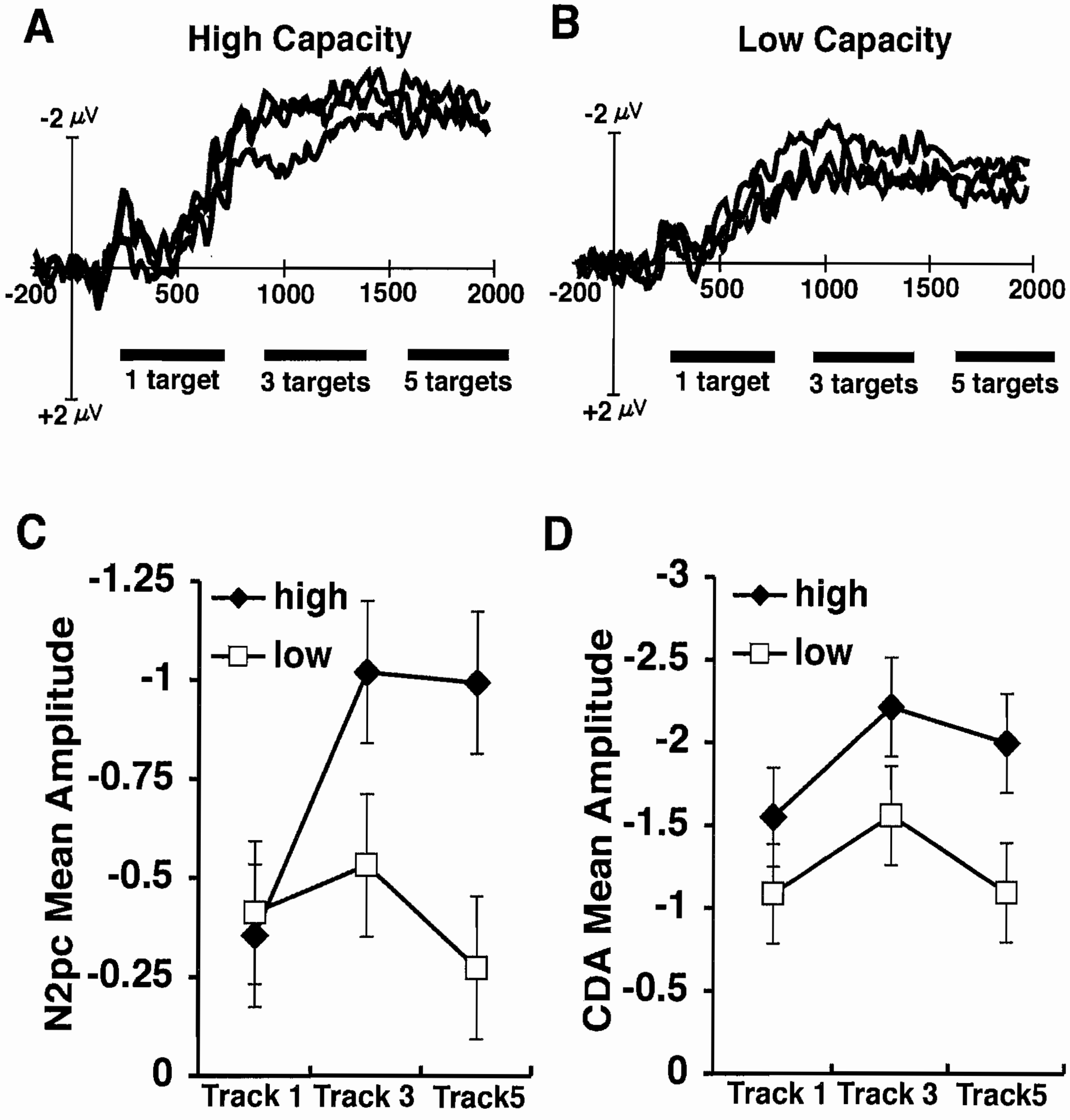

Figure 2-4: Experiment 3 Results ERP difference waves for correct trials in Experiment 3 divided between high capacity (A) and low capacity individuals (B) on the basis of a median split of tracking performance. Mean amplitude (in microvolts) of the N2pc (C) and the CDA (D) for the high and low capacity groups across the three target array sizes.

Experiment 4: Predicting Individual Differences in Tracking Capacity

The results of Experiment 3 indicate that the amplitude of both the N2pc and the CDA are highly sensitive to the tracking capacity limitations that constrain performance in this task because it reaches a limit at tracking three targets and is also finely attuned to individual differences in tracking capacity. However, this sensitivity to individual differences was not restricted to the response to supracapacity target arrays, but was also observed in the size of the increase in amplitude from one target to three targets. This resulted in a highly significant interaction between group (high vs low) and number of targets (1 vs 3) (N2pc: $p < .001$; CDA: $p < .01$), with a larger increase from one to three targets for the high capacity group than for the low capacity group. The smaller difference in amplitude between one and three targets for the low capacity group suggests that the one-target arrays consumed a larger proportion of available capacity than for the high capacity group, resulting in a smaller increase to three items. Paired t-tests support this assertion because the difference between the high and low groups was not significant in the track 1 condition (p's > .15) but the difference between these two groups was highly significant in the track 3 condition (N2pc: $p < .005$; CDA: $p < .01$).

We tested the robustness of this relationship by running an additional group of subjects in the one and three target conditions and combining this data with all of the subjects from the previous experiments so that we could have a large sample (N=63). Figure 2-5 shows the amplitude of both waves for tracking one or three targets divided between high capacity and low capacity subjects. From the figure, there are two apparent differences between the high and low capacity groups: first, the high capacity group tends to have overall larger amplitudes for each wave; and second, the high capacity group shows a larger rise in amplitude from 1 to 3 items than the low capacity group. This pattern of effects was confirmed in an Analysis of Variance (ANOVA), yielding significant main effects of group (both p's $< .05$) and number of targets (both p's $< .001$), as well as a significant interaction between group and number of targets ($p < .01$). Although high capacity subjects tend to have higher overall amplitudes (irrespective of number of targets), this factor is only a fairly weak to moderate predictor of an individual's tracking capacity (N2pc: $r=.22$, $p <.10$; CDA: $r=.31$, $p <.05$). By contrast, we found that the rise in amplitude from one target to three targets was a much stronger predictor of an individual's tracking capacity (N2pc: $r=.70$, $p < .001$; CDA: $r =.48$; $p < .001$). Importantly, these strong correlations persisted even when we partialled out the relationship between overall amplitude and tracking capacity (partial r's = .68 and .41 for N2pc and CDA, respectively). Thus, it appears that it is the amount of differentiation in amplitude between increasing numbers of targets that may be most predictive of an individual's tracking capacity. We also found that the rise in N2pc amplitude from one to three targets was strongly correlated with the rise of the CDA ($r = .72$, $p < .001$) which further indicates that there is a tight coupling between these measures of object selection and sustained attention. However, because of this strong relationship, we also calculated partial correlations for both the N2pc and CDA effects (i.e., rise from 1 to 3 targets) so that we could measure each wave's unique contribution to predicting tracking capacity. Although the N2pc effect remained a strong predictor of tracking capacity when the contribution of the CDA effect was removed (partial $r = .59$, $p < .001$), the CDA effect was only a weak predictor of tracking capacity when the N2pc effect was removed (partial r = .09; ns). Importantly, these effects were not simply due to more variability in the CDA than the N2pc. Measurements of the reliability of each component revealed that both components were highly stable within subjects, and that the CDA actually had a higher reliability than the N2pc (Cronbach's alpha = 0.74 for the N2pc; 0.94 for the CDA). Consequently, these results demonstrate that while neural indices of both target selection (N2pc) and sustained attention (CDA) can serve as strong neurophysiological predictors of attentional tracking capacity, it is the selection process that explains most of the unique variance in tracking capacity across individuals.

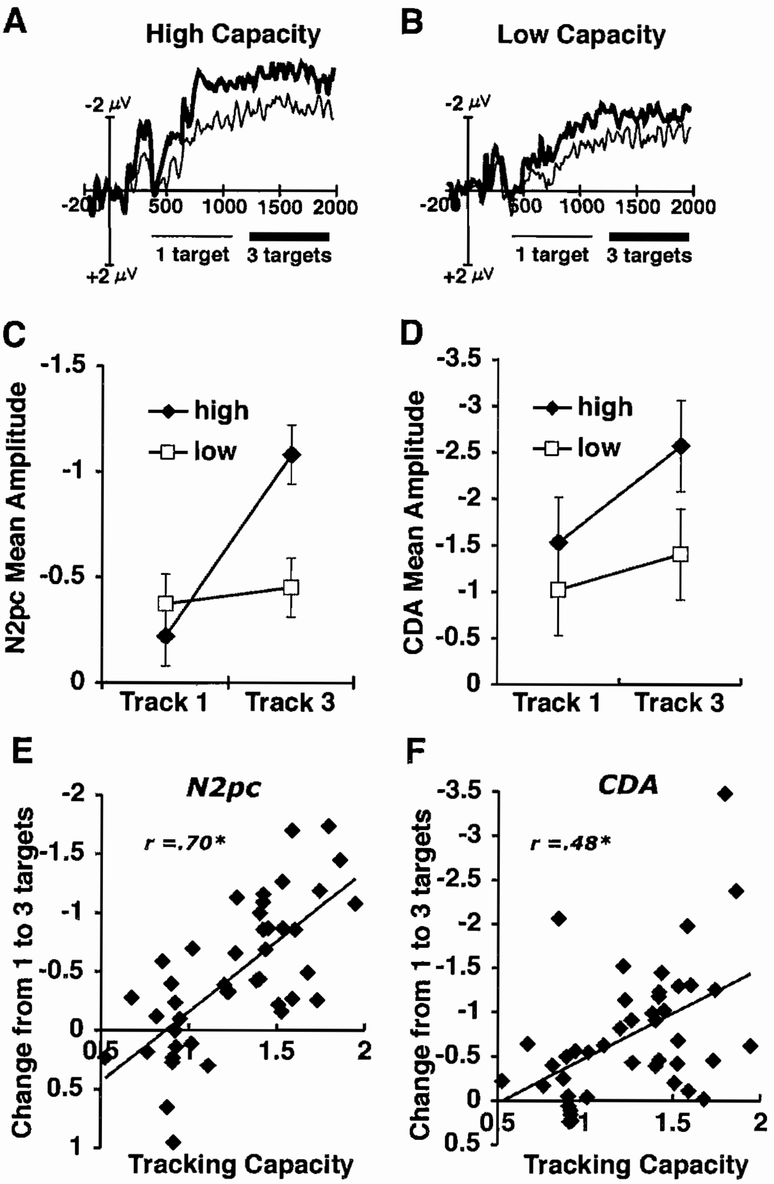

Figure 2-5: Experiment 4 Results (A, B) ERP difference waves for high and low capacity subjects in Experiment 4. (C, D) Mean amplitudes of the N2pc and CDA waves across high and low capacity groups. There was a significant interaction between group (high/low) and number of objects for both waves (p < .01). (E, F) Correlation between an individual's tracking capacity and the difference in amplitude (in microvolts) between one and three objects for the N2pc and the CDA. Note that tracking capacity in our single-hemifield experiments was generally 2-3 items: lower than most previous tracking capacity estimates, but consistent with Alvarez & Cavanagh's (2005) demonstration of lower capacity estimates when tracking items in a single hemifield.

Experiment 5: Limiting Factor for Tracking Capacity: Selection or Tracking?

Our observation that how efficiently an individual initially selects the target items strongly predicts their overall tracking capacity is somewhat surprising because selection occurs well before tracking (i.e., motion onset) even begins. In this regard, one could argue that there must always be a strong relationship between selection and tracking performance because subjects can track only the targets that were appropriately selected in the first place. However, there are likely to be many processes that contribute to an individual's overall tracking capacity depending upon the specific nature of the tracking task that is being used to estimate capacity (vanMarle and Scholl, 2003; Oksama and Hyönä, 2004; Alvarez et al., 2005; Liu et al., 2005; Pylyshyn and Annan, 2006). Indeed, our behavioral estimate of tracking capacity may actually load heavily on the selection stage because the subjects were required to hold fixation while selecting a subset of targets amongst distractors within a single hemifield. Moreover, it is possible that there is a somewhat weaker contribution of sustained attention activity in our behavioral measure because our tracking period is relatively short (i.e., 1.5 seconds) compared to previous studies that tend to use longer periods of tracking (e.g., 8-10 seconds).

In the final experiment we tested whether these two neural predictors of tracking capacity would be sensitive to a change in the relative contributions of selection and sustained attention by assessing each component's (i.e., N2pc and CDA) ability to predict an individual's tracking capacity in a "whole field" tracking task with a longer duration. More specifically, subjects were tested in two separate sessions. In a behavior-only session, subjects were asked to track 3, 4, or 5 target items amongst distractors that were spread across the entire visual field ("whole-field") and they tracked these items for 8 seconds. In a separate ERP session, subjects performed a single hemifield tracking task that was identical to that used in Experiment 4. We estimated each subject's "whole field" tracking capacity on the basis of performance in the behavior-only session, and used this estimate as a predictor of his or her N2pc and CDA effects that were measured in the single hemifield ERP tracking task. In a "whole field" tracking situation, the difficulty of target selection should be reduced because the subjects could freely view and select the targets across the entire display. In contrast, the difficulty of sustained attention should be raised because of the substantial increase in how long the targets needed to be tracked continuously. Consequently, we would expect that the N2pc effect should now become a weaker predictor of "whole field" tracking capacity; simultaneously, we expect that the CDA should become a stronger predictor of tracking capacity as the limiting factor in task performance shifts from selection to sustained attention. As shown in Figure 2-6, we observed that while the correlation between the N2pc difference effect and whole field tracking capacity was considerably weaker than we observed previously ($r = .31$, $p < .07$), the CDA difference became a much stronger predictor of tracking performance ($r = .72$; $p < .001$). Again, the N2pc and CDA effects were strongly correlated ($r = .52$, $p < .05$). Moreover, when we partialled out the contribution of the N2pc effect, the relationship between the CDA effect and tracking capacity remained strong (partial $r = .69$; $p < .01$); Conversely, the N2pc was no longer predictive of tracking capacity when the CDA effect contribution was removed (partial r = .10; ns). Thus, in this "whole field" tracking context, it is our index of sustained attention that explains most of the unique variance in attentional tracking capacity across individuals.

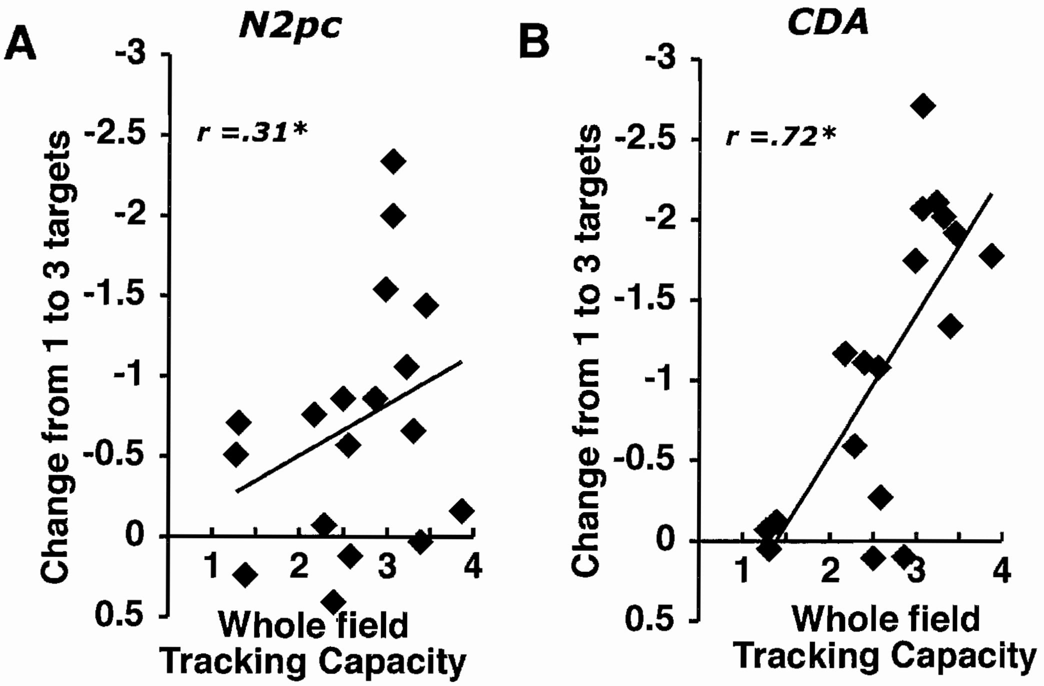

Figure 2-6: Experiment 5 Results Correlations between an individual's Whole Field tracking capacity and the rise in amplitude from 1 to 3 targets for the N2pc (A) and the CDA (B). Tracking capacity was estimated by averaging behavioral performance across all set sizes (3, 4 and 5).

DISCUSSION

Overall, these results indicate that we have isolated neural measures of the target selection and sustained attention processes that underlie our limited ability to track multiple moving objects. Indeed, by measuring the amplitudes of the N2pc and CDA waves we could determine how many targets were being selected or tracked during a trial as well as being highly sensitive to a given subject's specific tracking capacity.

Moreover, these two neural measures allow us to finely index what the primary limiting factors for performance are on a given measure of tracking capacity. Under difficult selection conditions, variability in the N2pc effect strongly predicts tracking performance. Whereas, when selection is less taxing but the targets must be tracked for longer durations, it is the variability in the CDA that strongly predicts tracking performance. However, under both situations, we found that it was the amount of separation in amplitude between different numbers of targets (i.e., rise from 1 to 3 targets) that was the primary predictor of tracking ability. Consequently, these results suggest that individual differences in tracking performance may be primarily determined by how efficiently the visual system can individuate the targets from one another as well as from the distractors (Sears and Pylyshyn, 2000; Intriligator and Cavanagh, 2001; Ogawa et al., 2002; Vogel et al., 2005; Suganuma and Yokosawa, 2006).

The results of this series of experiments also have significant implications regarding the neural systems that underlie the attentional mechanisms involved in selecting and tracking moving objects. For example, we have found that the amplitude of the N2pc provides a reliable index of the number of targets being selected, but strongly follows the limits of attentional tracking capacity. Considering that the N2pc is thought to be generated in V4 and posterior portions of inferior temporal cortex (e.g., Hopf et al., 2006), these results suggest that selective attention effects in these regions may show similar sensitivity to capacity limits. This is consistent with the viewpoint that attention effects in these regions may reflect processes that help to individuate targets from distractors (e.g., Desimone and Duncan, 1995; Kastner and Ungerleider, 2000).

Moreover, recent work has suggested that the attentional “spotlight” can be split to two noncontiguous locations simultaneously without also being allocated to the intervening space (Awh and Pashler, 2000; Müller and Hübner, 2002; Müller et al., 2003), and that areas of extra-striate cortex show distinct focal activation patterns under split-attention conditions (McMains and Somers, 2004, 2005). Thus, it is plausible that similar attentional mechanisms underlie our current N2pc target selection effects and these demonstrations of split attentional foci. If this is the case, we would predict that the attentional capacity of the observer would impose an upper limit on the number of locations that could be simultaneously selected.

The response of the CDA during tracking also suggests implications regarding the underlying neural mechanisms involved in sustaining attention towards targets. The primary candidate neural source for the CDA is the IPS, which stems from the fact that previous work has shown that this area was modulated by the number of items that are being tracked (e.g., Jovicich et al., 2001), as well as the finding that this region shows highly similar patterns of BOLD activation during working memory load manipulations. Like the CDA, the IPS also reaches asymptotic activity levels for memory loads of approximately 3 items, and is sensitive to individual differences in working memory capacity (Todd and Marois, 2005). Thus, the finding that the CDA shows parallel responses during attentional tracking and visual working memory tasks suggests that cells in the IPS may actually facilitate the processing of both tasks. In this regard, the IPS may reflect a smart, but limited-capacity pointer system that helps keep individuated representations of objects actively maintained in working memory tasks and spatially updated in attentional tracking tasks. While the current results are highly consistent with the interpretation that similar neural mechanisms may underlie the capacity limits of both types of tasks, they are still insufficient to resolve this particular question because we have not directly compared the neural activity during visual working memory and attentional tracking tasks in the same subjects. However, the present results appear to provide an experimental approach for addressing this question in the future.

Conclusions

Our limited ability to divide attention so that we may keep track of multiple moving objects is a central limitation within cognition, and is thought to underlie our performance of a wide assortment of common tasks. Moreover, an individual's tracking capacity has been shown to be positively related to performance on a broad range of high-level cognitive functions, including measures of fluid intelligence (Oksama and Hyönä, 2004). The present results demonstrate strong and robust neurophysiological predictors of individual differences in attentional tracking capacity. Thus, they provide an initial link between this fundamental cognitive limitation and the two primary stages of neural activity that facilitate attentional tracking.

CHAPTER III

ATTENTIONAL ENHANCEMENT DURING MULTIPLE OBJECT TRACKING

This work was previously published with Andrew McCollough, Todd S. Horowitz and Edward K. Vogel in *Psychonomic Bulletin & Review.*

INTRODUCTION

One of the more dramatic demonstrations of attention to multiple foci is the multiple object tracking task (MOT, Pylyshyn & Storm, 1988). The subject is presented with an array of identical objects and told to follow a subset of target objects as all of the items move independently for several seconds or minutes. Intuitively, this is a challenging task, yet most people can track 3-5 objects under typical conditions. Our goal in this study was to determine how spatial attention is allocated during this task. In particular, we sought to establish a hierarchy of the allocation of attention to various elements of the display (i.e., targets, distractors, and background) so that we may begin to characterize the mechanisms by which attention facilitates tracking.

Spatial attention is thought to act through a combination of mechanisms that both enhance the processing of relevant information and suppress the processing of irrelevant information (e.g. Posner & Dehaene, 1994). These two mechanisms are generally distinguished by comparing the processing of attended and unattended information to an attention-neutral baseline condition. Attended stimuli typically show *enhancement* relative to baseline, while unattended stimuli show *suppression*. The preferred technique of assessing the rôle of spatial attention during tracking tasks has been the *dot-probe* method (Alvarez & Scholl, 2005; Feria, 2008; Flombaum et al., 2008; Pylyshyn, 2006; Pylyshyn et al., in press), which has been widely used to infer attentional distribution in visual search tasks (Cave & Zimmerman, 1997; Cepeda, Cave, Bichot, & Kim, 1998; Klein, 1988). In this technique, subjects must detect small, low contrast probe dots presented at various locations while simultaneously performing the MOT task. The assumption is that probes should be detected most readily at attended locations and should be more likely to be missed when presented at unattended locations.

Using the dot-probe technique, Pylyshyn (2006; Pylyshyn et al., in press) compared detection performance for probes on targets and distractors with a neutral baseline condition in which probes were presented in empty space within the display. He found that detection was highest for empty space probes, while target probes were detected more frequently than distractor probes. Pylyshyn attributed this unexpected superiority for empty space to a low-level masking effect for probes on objects. To control for this masking effect, he also asked subjects to detect probes in the display without the requirement to track targets and found that they were much better at detecting probes in space than on moving items. Using performance on this task to reinterpret probe detection in the tracking task, he concluded that probe performance was equivalent for targets and empty space, but impaired for probes on distractors. This pattern of results suggests that the primary role of spatial attention during MOT is to suppress distractors. Surprisingly though, it suggests that the tracked targets are not enhanced by attention, which contrasts strongly with the spatial attention literature that typically observes a combination of enhancement and suppression attention effects (Hillyard et al., 1998; Hopf et al., 2006; Luck, 1995; Moran & Desimone, 1985). One way to interpret these data would be to conclude that attentional enhancement is simply not involved in tracking moving targets. However, the absence of evidence is not evidence of absence. The aim of this paper is to test the alternative hypothesis that this lack of evidence for attentional enhancement of targets during tracking is a consequence of how attentional allocation in MOT has been measured.

The absence of evidence for an attentional enhancement of tracked targets may suggest that the attentional mechanisms that facilitate tracking are distinct from those involved in spatial attention. However, we argue that the dot-probe approach is not ideal for assessing the spatial distribution of attention in MOT, particularly target enhancement. Accurate probe detection relies upon the subject's awareness of the probe, which requires complete processing of the probe to the level of report. Considering that most previous demonstrations of target enhancement in spatial attention tasks have been shown to occur at fairly early (~100ms) perceptual stages of processing (Hillyard et al., 1998; Luck, 1995), the dot-probe approach may not be sufficiently sensitive to detect enhancements that occur at such an early stage. Furthermore, the dot-probe technique itself may influence the distribution of attention in MOT. Subjects are in a dual task situation where attentional resources must be shared between tracking and probe detection. Subjects cannot ignore distractors and empty space entirely, because task-relevant probes will be presented at these locations. Thus, detection performance for dot probes may tell us more about the strategies subjects use to achieve both tasks simultaneously than it does about attention distribution in the primary task (MOT).

In the present study, subjects have a single task: tracking targets. We present probes at various locations, but instead of asking the subject to detect them, we measure the electrophysiological response to these *task-irrelevant* probes. We measured the P1 and N1 components of the event-related potential (ERP). These are early (~75-150ms) visual-evoked responses that reflect initial perceptual processing in extrastriate cortical areas (Heinze et al., 1994a; Hillyard et al., 1998). Both components have repeatedly been shown to be acutely sensitive to the allocation of spatial attention, even when the evoking stimulus is task-irrelevant (Heinze et al., 1990; Vogel, Luck, & Shapiro, 1998). Moreover, the P1 and N1 attention effects have been shown to be sensitive to both enhancement of attended information and suppression of unattended information. In particular, Luck (Luck et al, 1994; Luck, 1995) found that the P1 to items at unattended locations was suppressed relative to neutral conditions. Conversely, the N1 to items at attended locations was enhanced relative to neutral conditions. Together, these previous results indicate that the P1 and N1 responses to task-irrelevant probes provide an ideal index for measuring both attentional enhancement and suppression in MOT at an early perceptual stage. If target positions are attentionally enhanced, we should expect larger

P1/N1 responses to probes on targets than to distractors or empty space. If distractors are suppressed, we should expect a decreased P1 response to distractors relative to empty space.

As Pylyshyn (2006) noted, finding an appropriate neutral baseline condition is a difficult problem for the dot-probe technique. It may be easier to detect empty space probes because they are not masked by item contours. Therefore, we also measured the ERP response to probes presented within stationary objects placed at random positions within each quadrant of the display (see also Pylyshyn et al., in press). Aside from not moving, these objects were identical in appearance to the moving items, so that stationary probes would be equally subject to contour masking[1]. Thus, we had two neutral baseline conditions: empty space and stationary objects.

Subjects maintained central fixation while tracking two targets among four moving distractors and four stationary objects for 6.33 seconds (see Figure 3-1). At the end of the trial, all movement ceased, one object became red and the subject judged whether or not it was a target. During the tracking period of each trial, eight task-irrelevant white square probes were briefly flashed at variable intervals. These probes could appear randomly on a target, a distractor, in empty space, or on a stationary object.

---

[1] Although contours for stationary distractors may not be identical to moving items due to motion-defined contours, our results indicate that probes in empty space elicited a smaller electrophysiological response than probes on distractors or targets.

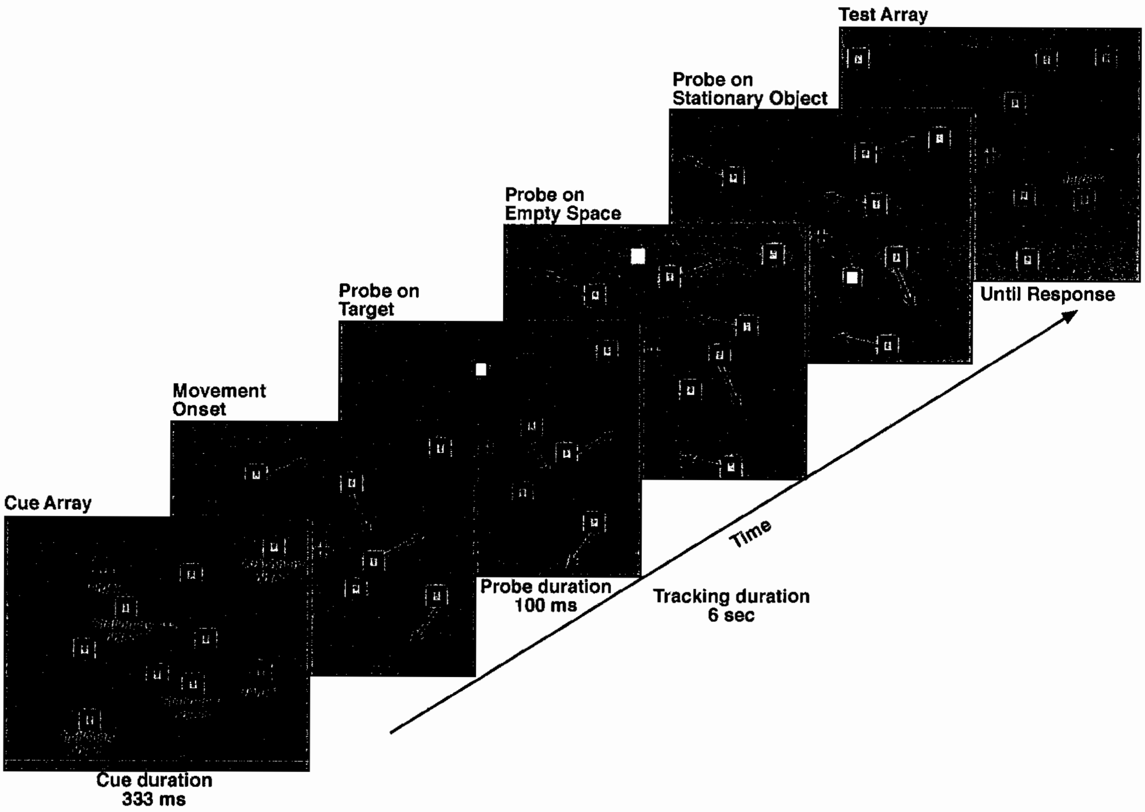

Figure 3-1: Experimental Paradigm The sequence of events in our MOT task. At the start of each trial, targets were identified as red (striped in the figure) amongst black squares. During the trial, target and distractor items moved in random directions. Approximately every 633ms, a task-irrelevant probe appeared on a target, moving distractor, stationary object or empty space. At the end of each trial subjects categorized a single red item as either 'target' or 'non-target' with a button press.

METHOD

Participants

Thirty-one participants (19 female, age range 18-31) from the Eugene, Oregon community completed the experiment for monetary compensation. Three participants were excluded because of excessive eye movements (see below), leaving a total of 28 subjects in the sample.

Stimuli and Procedure

Each participant completed 12 blocks of 30 trials each (360 total trials). Each trial included two of each type of probe: *target*, *distractor*, *stationary object* and *empty space*, for a total of 720 probes per type. All items were empty boxes subtending approximately 0.5 degree of visual angle (°). Items moved along random trajectories at a constant velocity of 1°/s. Motion was constrained within an invisible 17° x 17° box centered on the screen. Items were allowed to collide and reflected from each other at their angle of incidence with no momentum exchange.

At the start of each trial, all items were stationary. Two of the ten items were red, designating them as targets. After 333 ms, the targets turned black and began to move, along with four of the eight distractors. During the trial, white probes appeared at varying intervals with a minimum inter-probe interval of 633ms and a duration of 100ms. After 6333ms, all motion ceased, one item became red, and the participant responded as to whether or not this item was a target. The red item was equally likely to be a target or a moving distractor.

Recording and Analysis

Electroencephalographic (EEG) activity was recorded from 20 tin electrodes mounted in an elastic cap (Electrocap International). In addition to the standard

International 10/20 System sites, four additional sites were used: OL and OR, positioned midway between O1 and T5 on the left hemisphere and O2 and T6 on the right; POz, located on the midline between Pz and O1-O2, and PO3 and PO4, located halfway between POz and T5 on the left and POz and T6 on the right. All sites were recorded with a left-mastoid reference, and the data were re-referenced offline to the algebraic average of the left and right mastoids. The horizontal electrooculogram (EOG) was recorded from electrodes placed approximately 1cm to the left and right of the external canthus of each eye to measure horizontal eye movements. In order to detect blinks and vertical eye movements the vertical EOG was recorded from an electrode mounted beneath the left eye and referenced to the right mastoid. Probe events containing artifacts (ocular, movement, or amplifier saturation) were discarded. Subjects with artifact rejection rates in excess of 25% were excluded from the sample. Three subjects were excluded from further analysis using this criterion. EEG and EOG were amplified with an SA Instrumentation amplifier with a bandpass of 0.01-80 Hz and were digitized at 250 Hz in LabView 6.1 running on a Macintosh.

RESULTS AND DISCUSSION

Behavioral Tracking Performance

Tracking performance was quite good (mean percent correct: 88%, SD = .08). We transformed accuracy to effective tracking capacity, $m = n(2p-1)$, where n is the number of targets (e.g. 2), and *p* is percent correct (Scholl, 2001). Mean *m* was 1.52 objects (out of a maximum possible score of 2), with substantial inter-subject variability (SD = 0.3).

ERP Responses to Probes

Figure 3-2 shows ERPs time-locked to probe onset across the four probe conditions. The two early spatial attention-sensitive components of interest can be clearly seen. The initial positive wave (P1) displays a narrowly-focused scalp distribution, maximal over occipital electrodes. This is followed by the more broadly-distributed negative wave (N1) which is maximal at central electrodes. For further analysis, we defined P1 amplitude as the mean amplitude from 100-150ms following probe onset at an occipital pair of electrodes (OL/OR). We similarly defined N1 as the mean amplitude from 125-185ms following probe onset at central electrode sites (Cz, C3, & C4). As seen in Figure 2B, both of these components were strongly modulated by probe type, yielding a significant effect of probe type on amplitude (**P1** F(3,81)=9.93, p<.001, **N1** F(3,81)= 23.44, p<.001).

For both components, amplitude was highest for target probes, followed by distractors and empty space, and was lowest for stationary objects. Subsequent paired t-tests revealed significant differences between target probes and all other probe types (**P1**: t(27)= 3.36, 4.65, 3.01; **N1**: t(27) 4.13, 6.42, 6.89, all p<.007). Furthermore, N1 amplitude to distractor probes was greater than either of the baseline probe types (**stationary object** t(27)=3.01, p>.006; **empty space** t(27)=3.23, p>.004). However, while P1 amplitude to distractor probes was greater than to stationary objects (t(27)=3.33, p<.004), it was not reliably different from responses to empty space (t(27)=.75).

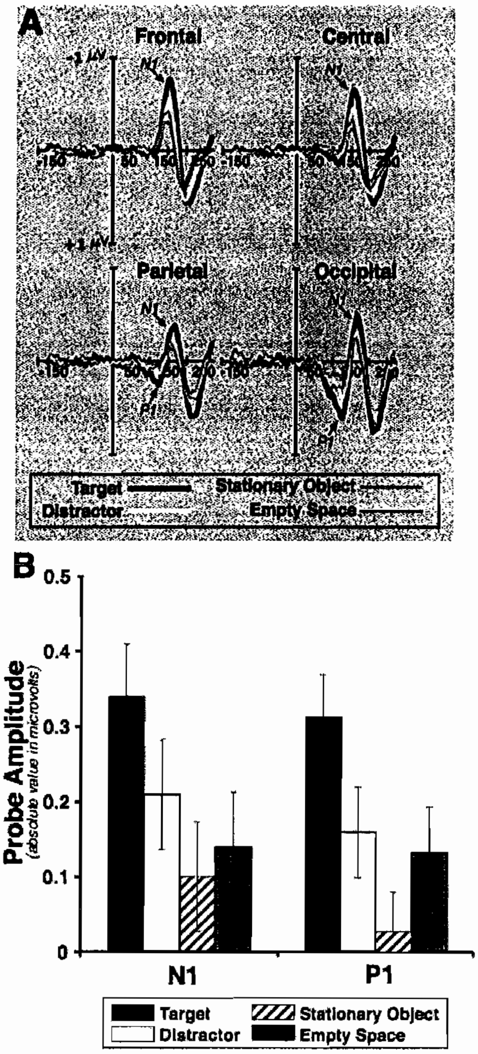

Figure 3-2: Electrophysiological Results Electrophysiological response time-locked to probe onset. The frontal, central and parietal waveforms are grouped averages of three electrodes at those sites, while the occipital waveform is the average response from the OL and OR electrodes. 2B: Absolute value of mean amplitude for the N1 and P1. P1 amplitude is a positive-voltage wave observed from the occipital sites 100-150ms post-stimulus. N1 amplitude is a negative-voltage wave observed from the central electrode sites 125-185ms post-stimulus. Error bars represent standard error of the mean.

Relationship to Tracking Performance

Are these electrophysiological effects simply correlated with attentional allocation or are they related to performance? To answer this question, we took advantage of the inter-individual variance in tracking and attempted to predict P1/N1 amplitude on the basis of tracking performance. We performed a median split of the ERP data based on the subjects' tracking performance and analyzed ERP amplitude as a function of group (i.e. good trackers vs poor trackers) and probe type. N1 amplitude was highly sensitive to tracking performance. As can be seen in Figure 3a, the primary difference between the two groups was in the relative amplitudes to targets and distractor probes, with good trackers showing a much larger difference between these two conditions than poor trackers (see Figure 3b). We found a significant interaction between group and target vs distractor probes (F(1,26)=6.24, p=.019). Importantly, we looked at correlations across all subjects to verify that this effect was not an artifact of the median-split procedure. Before doing so, we calculated the reliability of each measure using a split-half correlation procedure. The reliability for these measures were as follows: behavioral performance (r = .83), average N1 response (r=.89), response to target probes (r=.67), and the difference between target and distractor responses (r=.65). Figure 3c shows the correlation between the target-distractor difference in N1 amplitude and tracking capacity (*m*), which was highly significant (r=.43, p=.024; when corrected for attenuation, r= 0.59). However, it was not the case that good trackers simply had larger N1 amplitudes for all probes: neither overall N1 amplitude irrespective of probe placement (r=.08) nor target amplitude alone (r=.17) were significantly correlated with tracking ability. Similarly, the difference in amplitude between target probes and the two baseline probe types were not significantly correlated with tracking performance (r=.09 and r=.19, for empty space and stationary object, respectively), suggesting that the treatment of background space is the same for all subjects irrespective of tracking ability. In sum, these results indicate there was less attentional differentiation between moving distractors and targets for poor trackers than for their more skillful counterparts.

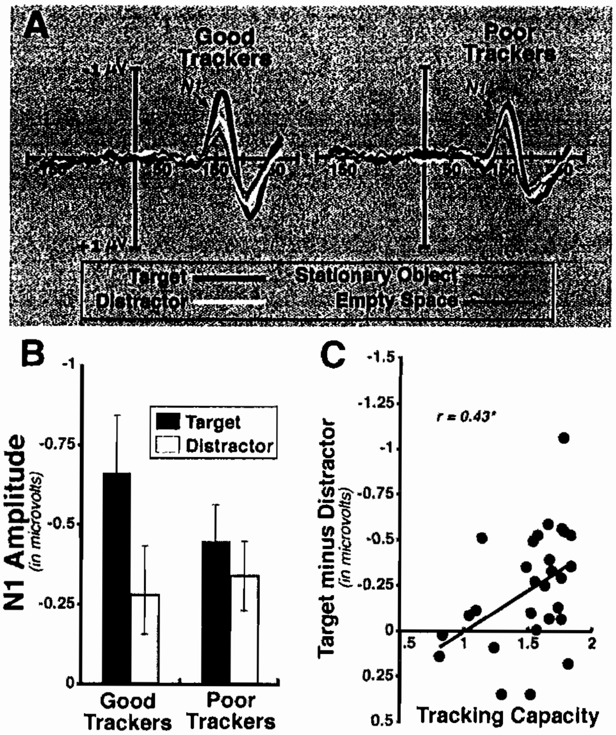

Figure 3-3: Individual Differences in Electrophysiological Data Electrophysiological response from central electrodes to probes for good trackers and poor trackers. Subjects were divided on a median split based on behavioral accuracy. Electrophysiological response to probes on stationary objects did not vary as a function of tracking accuracy. 3B: Mean amplitude from the central electrode group in response to probes on targets and distractors for good and poor trackers. 3C: Scatterplot between behavioral tracking ability (tracking capacity) and the difference between the response to target and distractor probes. In both 3B and 3C, the target response is larger relative to the distractor response for good trackers than poor trackers. One very accurate subject showed a much larger target-distractor difference than all other subjects. If we remove this subject, the correlation remains significant (r=.40, p=.038).

GENERAL DISCUSSION

What is the role of spatial attention during MOT? On the basis of results from the dot-probe paradigm, Pylyshyn (2006; in press) suggested that while attention suppresses distractors, the tracked targets are not enhanced by attention. On this distractor suppression model, we would expect equivalent ERP responses for probes on targets and on the background. However, we observed a substantially different hierarchy of attentional allocation: targets showed the greatest response, with weaker responses to the distractors, and the weakest responses to the background or stationary objects. Thus, our results provide strong evidence in favor of attentional enhancements of the targets during tracking. However, we found no evidence that the distractors are suppressed below the level of the background at least when measured at this early level of perceptual processing.

Previous work using spatial attention manipulations has indicated that the P1 component is indeed sensitive to the suppression of information at unattended locations (e.g., Luck et al, 1994). Thus, the absence of a suppression effect in the present study is unlikely to be due to a lack of sensitivity to suppression mechanisms. Nonetheless, these results certainly do not rule out the possibility of distractor suppression at all levels. Indeed, the behavioral evidence consistent with distractor suppression during MOT has been replicated in a number of studies and appears to be a robust and reliable effect (Flombaum et al., 2008; Pylyshyn, 2006; Pylyshyn et al, in press). How can we integrate the current results favoring target enhancement with the previous literature favoring distractor suppression? One possibility is that, while the P1/N1 response reflects attention at early, perceptual stages of processing, the behavioral measures reflect distractor suppression at later post-perceptual stages. If this formulation is correct, we would expect that post-perceptual ERP components (e.g., N400, P3) should show distractor suppression effects (for a related line of reasoning see Vogel, Luck & Shapiro, 1998). Another possibility is that distractor suppression reflects a strategy subjects adopt to deal with the dual-task demands of tracking targets while detecting probes. While we cannot distinguish between these alternatives with our current data set, this is a fruitful topic for further research. One caveat to the distractor suppression interpretation of existing MOT dot-probe studies is that the designation of enhancement or suppression is always made relative to the empty space baseline, and these studies typically find that probe detection in the absence of a tracking task is higher for empty space than for moving objects (Pylyshyn, 2006; Pylyshyn et al, in press). One finding that is very clear and consistent with the current results is that probes on target locations are always reported at a much higher rate than distractor probes.

During an attentional tracking task, we observed modulations of the visual-evoked P1 and N1 components that closely resemble those observed in standard spatial attention tasks (Heinze et al., 1994b; Mangun & Hillyard, 1991). While the attentional modulations of these components may be similar, it is certainly plausible that distinct mechanisms may be facilitating MOT and conventional spatial attention tasks. In particular, while spatial attention tasks generally require attention to be focused on a cued location in anticipation of a single upcoming target, MOT would appear to require object-based attention (Alvarez & Scholl, 2005; Drew & Vogel, 2008; Scholl, Pylyshyn, & Feldman, 2001; vanMarle & Scholl, 2003). Nonetheless, both location- and object-based attention appear to produce similar modulations of the perceptual response to task-irrelevant probes. For example, Martinez et al (2006) used a task-irrelevant probe ERP technique while subjects performed a variation of the Egly et al (1994) object-based attention task, and found that the P1 and N1 were enhanced for probes presented at the attended portion of an object. Importantly, they also found that the P1 and N1 were larger for probes on the unattended portion of the attended object than they were for probes on an unattended object that was equally distant from the attended region, indicating that the benefits of attentional allocation extended throughout the object.

Using a novel method of assessing spatial attention during MOT, our current results also help us to understand why individuals differ in tracking ability. We found that the difference between good and poor trackers was not the overall amplitude of the response to probes at the attended location, nor was it the treatment of nonmoving stimuli. The key difference in our data was the relative amounts of attention allocated to targets and distractors. We found that tracking performance improved as the difference in amplitude between probes on targets and distractors increased. One straightforward interpretation of this result is that poor trackers were more likely than good trackers to inadvertently track one or more distractors, leading to a smaller average difference between target and distractor responses. Although we did not find direct evidence that poor trackers paid significantly more attention to distractors than targets, it is possible that we failed to see such a relationship due to the fairly large number of distractors in the display. That is, given that there were 4 moving distractors, if a subject inadvertently began to track a particular distractor, we had only a one in four chance of probing that particular item on that trial. Future experiments will be necessary to more clearly determine whether these subjects directly allocate more attention to distractors. Nonetheless, the present results indicate that behavioral tracking performance is related to the relative amounts of attention allocated to targets and distractors. Thus, the current results are similar to our recent work examining the relationships between working memory capacity and the ability to prevent salient but irrelevant information from being stored in memory (Vogel, McCollough & Machizawa, 2005). Thus, the present results add to the growing body of evidence that the ability to selectively prevent irrelevant information from being attended is an important correlate for success in both visual working memory and MOT (Kane & Engle, 2003; McNab & Klingberg, 2008; Vogel, et al., 2005).

CHAPTER IV

ATTENTION TO OBJECTS AND MOTION DURING MULTIPLE OBJECT TRACKING

This chapter was written in collaboration with Todd S. Horowitz, Jeremy Wolfe, and Edward K. Vogel.

INTRODUCTION

Despite phenomenal experience to the contrary, the visual attention literature has shown that capacity limitations constrain the amount of visual information that we are able to process at any given moment to about four items. Experimentally, when visual information abruptly disappears for more than about 300ms (long enough for iconic memory to fade), people are typically able remember about 4 independent items (Jiang et al., 2000; Vogel et al., 2001; Xu, 2002). This task is thought to index the capacity of working memory, a cognitive construct thought to underlie the ability to maintain information in a durable form for short periods of time (Cowan, 2001; Vogel et al., 2001). While attention and WM have typically been thought of as separate, more recently researchers have noted many similarities between the two constructs. For example, Cowan (2001) has conceptualized working memory as the active portion of long term memory that is currently the focus of attention and a growing number of studies have shown that an individual's WM capacity predicts performance on a wide variety of attention tasks (Engle, 2002; Kane and Engle, 2002; Unsworth et al., 2004).

Recently, Fougnie and Marois (Fougnie and Marois, 2006) explored the connection between VWM and attention using a dual task experiment where participants were asked to maintain a number of items in memory while performing one of two tasks: either an additional WM task or a multiple object tracking task. In the MOT task, people are asked to track a subset of target items in a field of identical distractor items as all objects move about a field randomly. People are typically able to track between 4 and 5 items simultaneously (Pylyshyn and Storm, 1988). Although the authors found that the VWM task interfered with an additional VWM task more than the MOT task, there was clear evidence for interference between VWM and MOT tasks. The locus of this interference effect, however, is unclear. One possibility is that both tasks rely on the same capacity limited space for representing individual items and the interference observed was a result of competition for the same limited resource. Still, while both tasks necessitate maintaining the representation of as many targets as possible, the MOT task requires each target's location to be continuously updated throughout the trial duration so the interference may not be perfectly additive.

Lateralized versions of the VWM and MOT tasks yield a strikingly similar electrophysiological response: a contralateral negativity that is broadly distributed over posterior electrode sites and increases as the number of targets (Vogel and Machizawa, 2004; Vogel et al., 2005; McCollough et al., 2007; Woodman and Vogel, 2008). In both tasks, amplitude of the component does not increase for set sizes above behavioral capacity and the amplitude of the component is not sensitive to difficulty manipulations that do not affect the number of items the subject must attend (Drew and Vogel, 2008). In the current set of experiments, we sought to use this component to further investigate the relationship between tracking and visual working memory.

The fMRI literature is instructive in understanding the surprisingly similar response evoked by these two seemingly dissimilar tasks. During visual working memory tasks, a number of studies have shown that activity in the interparietal sulcus increases with the number of items that must be encoded (Linden et al., 2003; Xu and Chun, 2006; McNab and Klingberg, 2008), and reaches asymptote when the behavioral capacity is exceeded (Todd and Marois, 2004, 2005; Xu and Chun, 2006). The fMRI literature on MOT is considerably smaller, but the papers appear to converge on increased activation from a relatively stable group of areas during tracking (Culham et al., 1998; Culham et al., 2001; Jovicich et al., 2001; Howe et al., 2009). These papers generally compare passive viewing of moving stimuli to active viewing (tracking) and find a network of areas are more active during tracking including, FEF, SPL, IPS and MT+. Two papers (Culham et al., 2001; Jovicich et al., 2001) varied the number of targets the subject tracked and compared areas that were more sensitive to the load manipulation (tracking an increasing number of objects) than to the task manipulation (active tracking of targets compared to passive viewing of the moving stimuli). Both studies found that activity in IPS increased as the number of targets increased. Given the fact that activity in this area increases as a function of set size during both VWM and MOT tasks, activity in this area may reflect a pointer system that devotes an attentional focus to each of the tracked targets (Howe et al., 2009). The fact that both tasks appear to employ a similar region to focus attention on target locations implies that the strikingly similar behavioral capacity limitations in the two tasks may be driven by the processing capacity of the same pointer system. If this is the case, the strong prediction is that activity in IPS should reach asymptote when behavioral capacity is reached in an MOT task.

Howe and colleagues (2009) also hypothesize that area MT+'s role in the tracking task is to represent the location of the objects. Interestingly, when Jovovich and colleagues asked participants to track 0, 2, 3, 4 or 5 targets, there was only a marginal linear increase in MT+ activity as tracking load increased, but showed a large increase from 0 to 1 item (2001). Similarly, Culham and colleagues (2001) found that the task effect (tracking > passive viewing) was larger than the load effect in MT+. Critically, the visual stimulation during the tracking interval was identical in all 5 conditions. This suggests that MT+ is predominantly responding to *attention to motion* and is only weakly affected by the number of targets or the difficulty of the task. Accordingly, when Howe and colleagues (2009) contrasted a moving MOT display to a stationary display where the participants were simply asked to memorize the original location of the targets, MT+ activity was much larger during the moving display.

This set of results suggests that in terms of neural mechanisms, tracking and VWM tasks both engage a mechanism that is sensitive to the number of target in a given trial and appears to emanate from near the IPS. Further, the two tasks differ in the amount that they engage area MT+. This area appears to be primarily driven by the need to attend to motion and update target positions rather than the mere presence of motion and responds weakly to target load manipulations.

EXPERIMENT 1

In this experiment, we sought to directly compare the electrophysiological correlates of the lateralized VWM and MOT tasks within a single group of subjects (see Figure 4-1). On half of the blocks participants tracked 1 or 3 items, while in others they held 1 or 3 items in memory. In both cases, the initial selection period was 500ms and was followed by a 1500ms interval when subjects either tracked items as they moved randomly about the screen, or maintained the object information across a delay interval. At the end of each change detection trial, the items from the selection period reappeared and participants were asked to categorize the items as either 'same' or 'different' with a gamepad controller. In tracking trials, one item was filled in red and participants were asked to judge whether the item in question was originally red or not ('same' or 'different' than the original color). In both cases, the correct answer was ''same' on 50% of trials. Although we held the number of objects constant across the two tasks, memory performance was better than tracking performance in this experiment (VWM accuracy: 91%; MOT: 85%; $t(12)=3.09$, $p<.01$).

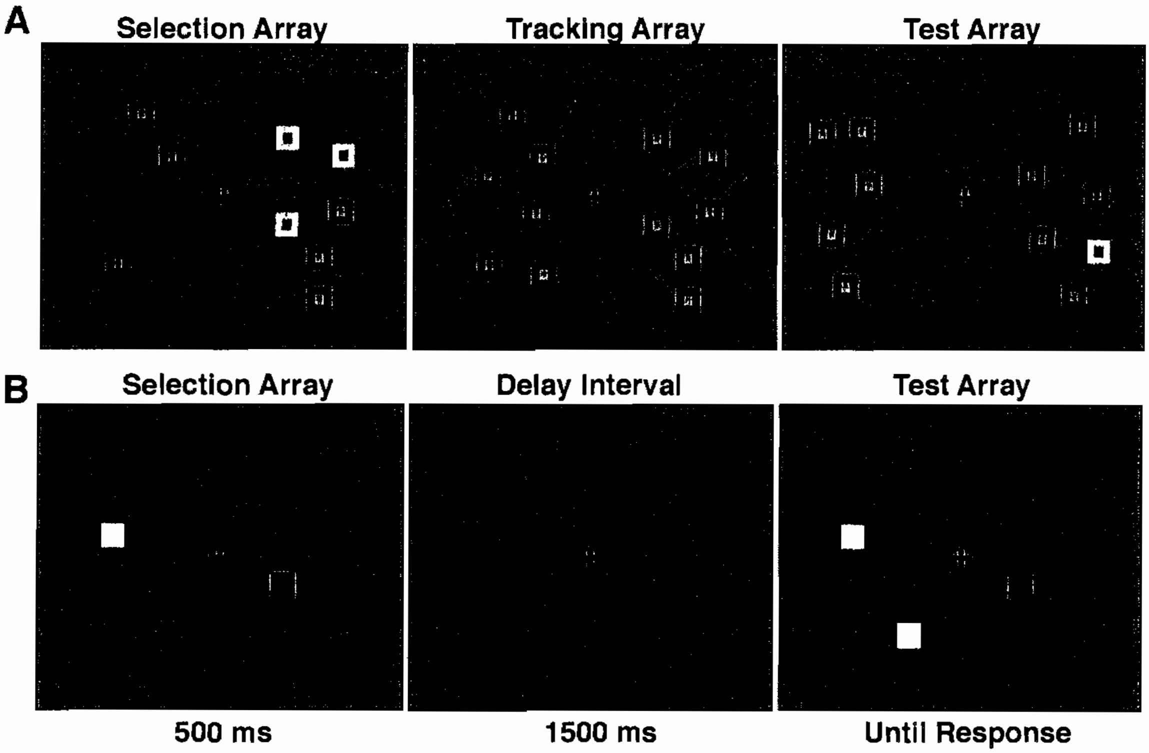

Figure 4-1: Experiment 1 paradigm

In this paper as in previous work, we will focus on lateralized components by defining electrode pairs as either contralateral or ipsilateral with respect to the side of the screen the participants were asked to covertly attend on a given trial. Next, we averaged the response across a set of 5 electrodes (P3/4, PO3/4, Ol/OR, O1/O2 and T5/6; see methods) and the side of the screen that was attended on a given trial (See Figure 4-2). Finally, by subtracting ipsilateral activity from contralateral activity we arrive at a difference wave that represents the average response. Examining this waveform for the 4 conditions in the experiment, two differences between the activity evoked by the tasks are clear: an overall increase in amplitude for the tracking trials, and a decrease in amplitude roughly 1000ms after stimulus offset for both memory conditions but neither tracking condition. To quantify these differences, we analyzed mean amplitude in two time periods: an early period (500-800ms) prior to the observed amplitude decay in the memory conditions, and a later time period (1600-1900ms) after the decay had taken place. In the memory task, there was a significant Time by Object number interaction (F(1,12)13.90, p<.004) with main effects for number of objects (F(1,12)=10.03, p<.009) and time window (F(1,12)=9.80, p<.01). The interaction appears to be driven by the fact that there is a clear set size effect for VWM task early on in the trial (t(12)=4.51, p<.002), and this effect is no longer significant later in the trial (t(12)=1.02, p=n.s.). Although the time and number of objects did not interact in the tracking task (F(1,12)=.01, p=n.s.), there was a significant main effect for number of objects (F(1,12)=17.81, p<.002) and time period (F(1,12)=21.53, p<.002). Amplitude for three objects was significantly higher than one object in both time periods (Early: F(1,12)=4.13, p<.002; Late: F(1,12)=3.89, p<.003). This is a striking finding: although the same amount of information must be maintained during the latter portion of the VWM trials, the differential contralateral activity decreases while activity in the tracking task increases. We suspect that this dissociation between the tasks has to do with consolidation of information during the VWM task that is not possible during the tracking task, but more work needs to be done to solidify this claim.

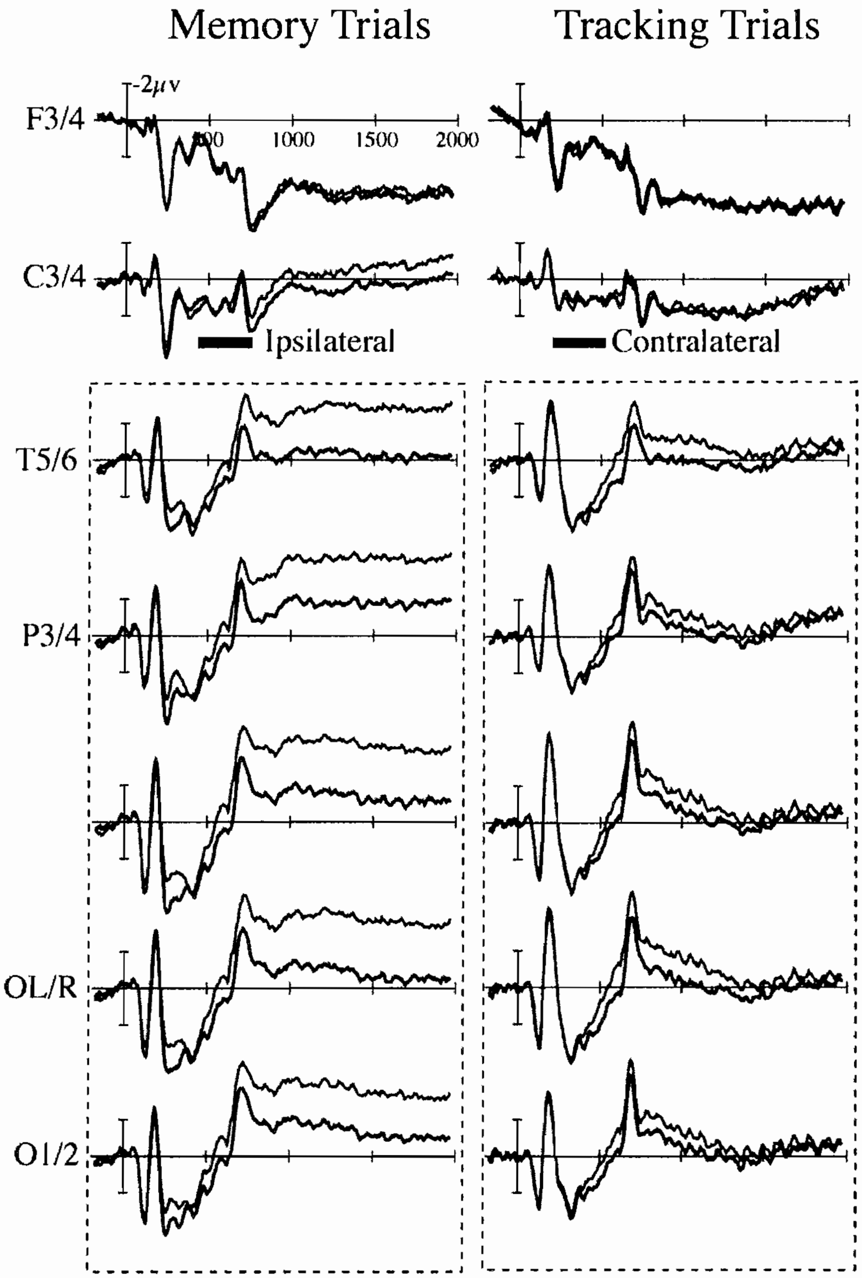

Figure 4-2: Experiment 1 Contralateral Waveforms

We a found a more subtle effect by examining amplitude for the two tasks during early time window. Although there is a main effect for number of items (F(1,12)= 28.53, p<.001) and task (F(1,12)= 5.13, p<.05), the two factors do not interact (F(1,12)=.509, p=n.s.). In the later time window, amplitude in the memory task decreases, leading to an interaction (F(1,12)= 7.81, p<.05) with main effect for number of objects (F(1,12)= 12.45, p<.005) and trial type (F(1,12)= 13.41, p<.005) (see Figure 4-2 & 4-3).

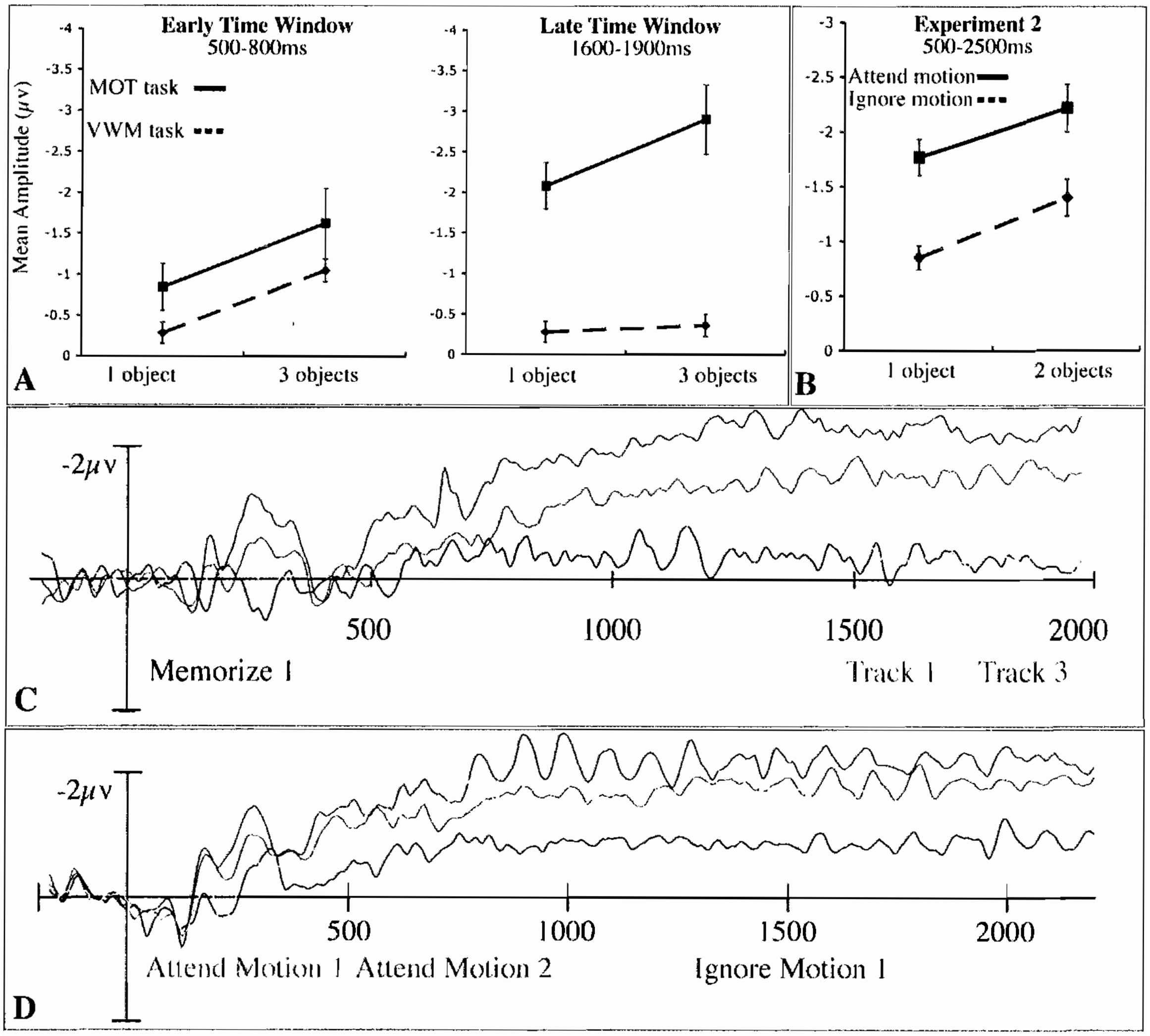

Figure 4-3: Experiment 1 & 2 Results

Amplitude in the tracking task is higher than the memory task even in this early period, when memory amplitude was maximal. In general, over many experiments in our lab, we have found that the CDA in tracking experiments tends to be larger (~2μv) than VWM experiments of comparable difficulty (~1μv). In Experiments 2 and 3, we manipulated both behavioral relevance of motion (Experiment 2) and the presence or absence of motion (Experiment 3) to better understand what this amplitude increase can tell us about the neural mechanisms that underlie tracking.

EXPERIMENT 2

By focusing on the observed differences in evoked contralateral amplitude during these two tasks, we hope to better understand how the tasks differ cognitively. One possibility is that the differences are driven by the differences in difficulty across the two tasks. In Experiment 1, the tracking task was more difficult than the memory task, so it may be the case that amplitude for the contralateral component simply increases with task difficulty. Previous work in our lab has shown that amplitude in both the memory task (Ikkai et al., in prep) and the tracking task (Drew & Vogel, 2008) is unaffected by difficulty manipulations, but in order to rule out this possibility, in the current experiment we ensured that difficulty for the two tasks was identical. Another possibility is that the mere presence of motion leads to a larger difference in contralateral and ipsilateral activity. In the fMRI literature, the typical method for localizing area MT+ is to contrast areas that show more activity during passive viewing of moving stimuli than viewing of static stimuli (Tootell et al., 1995; Tootell et al., 1997). Given that this area is retinotopically organized (Huk et al., 2002), the literature would predict a differential contralateral increase in area MT+ during the motion trials of Experiment 1 relative to the memory trials. Finally, two studies in the MOT literature have shown an increase in MT+ activity for attended motion as compared to passively viewed motion (Culham et al., 2001; Jovivich et al., 2001).

In Experiment 2, we attempted to test both of these hypotheses by keeping difficulty constant across the two tasks and holding the visual stimulation between the two tasks identical while manipulating the task set. The stimuli in this experiment were 4 lateralized, 2-armed pinwheels. In one block of trials, participants were asked to track either one or two arms of the pinwheels. When tracking two arms, the arms were always on different pinwheels on the same side of the screen so that it was always necessary to differentiate between a target arm and nearby distractor arm. After a 500ms selection period, the pinwheels started to rotate, changing direction and speed randomly so as to necessitate attentive tracking. At the end of each trial, one bar of the pinwheel was illuminated and the participant identified the bar as tracked or not tracked. In the other blocks of the experiment, the participants were asked to memorize the color of either one or two bars on the pinwheels during the selection period. During the delay interval, the color-less pinwheels rotated randomly using the same motion parameters as the tracking block. At the end of each trial, one bar was colored and the participant identified it as either same or different as it's original color. We used a set of 7 equiluminant colors that varied from red to green to increase the difficulty of the memory task.

The primary question in this experiment was whether the presence of motion would lead to the differences we observed between the response elicited by the tracking and VWM task in Experiment 1. In a 2x2 repeated measures ANOVA over a time window that encompassed the majority of the trial (500-2500ms), we found a significant effect of both number of targets (F(1,15)=39.4, p<.001) and trial type (F(1,15)=28.27, p<.001), but the interaction was not significant (F(1,15)=.482, p=n.s.). As can be seen in Figure 4-3, CDA amplitude during the tracking trials was significantly larger than memory trials for the duration of the trial. Given the difficulty in the two tasks was equivalent (Tracking 76.8% correct, Memory 76.8% correct; F(1,15)=0.00, p=n.s.), the main effect we observed appears to have been driven by the difference in task demands. In the tracking blocks, it was necessary to attend to the motion of the pinwheels, while in the memory blocks the motion was completely irrelevant to the color-memory task. Unlike Experiment 1 though, we did not observe a decrease in CDA amplitude during the memory trial (500-1500ms amplitude= -1.17µv, 1500-2500 amplitude = -1.12µv, t(15)=.94, p=n.s.). This suggests that the presence of irrelevant motion was responsible for the stability of the CDA. It may be that irrelevant motion in an attended position necessitates more active maintenance of visual information than when there is no competing visual information in that location. In the absence of visual stimulation, the ipsilateral hemisphere may be able to assist in the maintenance process leading to a decrease in the CDA.

This pattern of results suggests that the need to attend to motion leads to a large increase of contralateral amplitude, which we have termed Contralateral Attention to Motion Activity (CAMA). We believe that this activity is distinct from CDA activity, which is thought to serve as an index of the number of items that are actively being maintained in working memory. In this experiment, as in the early period of Experiment 1, we did not observe an interaction between number of targets and the task despite large main effects for both factors. If the CAMA is a simply an index of the number of items that are moving or who's position information must be updated, there should be an interaction between these two factors as tracking 2 items (or 3 items in Experiment 1) should necessitate more updating than tracking one item. Instead, we found evidence of an all-or-none effect where the contralateral amplitude increases a set amount when motion must be attended but this increase was unaffected by the number of targets. This can be observed subtracting memory activity from tracking activity in the appropriate set sizes (Figure 4-4). The logic for this subtraction is that the two conditions have the same number of items that must be attended and are indexed by the CDA, but differ in necessity to attend to motion, as indexed by the CAMA. Using the same logic, we created topographic maps of activity for the attended motion effect by subtracting memory activity from tracking activity and comparing this topographic map to the set size effect. We computed this map by subtracting the response for Track 1 item from Track 2 trials. Our ability to make strong conclusions about localization is restricted due to the inherent limitations of ERP localization and the fact that we used relatively low-density caps with 20 electrodes. Nonetheless, these scalp topography maps show a clear difference between the distribution of activity related to the attention to motion (the CAMA effect) and the set size effect. While the set size effect appears to be quite similar to scalp topography for the CDA with a relatively narrow focus on occipito-parietal electrodes, the CAMA is much more broadly distributed and appears to extend more anterior than CDA activity.

We found a similar pattern of results in Experiment 1 during the early tracking period (500-800ms): the updating effect is more broadly distributed and anterior, while the set size effect appears similar to Experiment 2 and previous work (McCollough et al., 2007; Jolicoeur et al., 2008). Given the difficulty of interpreting the underlying neural generators based on scalp voltage distributions, this apparent scalp topography difference should be interpreted with caution. The most important difference between the two effects is dissociation between the two components and target modulation. The CDA is sensitive to target set size whereas the CAMA is not. Furthermore, the all-or-none response we observed when motion had to be attended to complete the task mirrors results from two fMRI studies of MOT (Culham et al., 2001; Jovicich et al., 2001). In both cases, when passive viewing of moving items was contrasted with tracking conditions there was a larger increase in MT+ amplitude than when the number of targets tracked was manipulated.

In this experiment, we have identified two independent causes for the differences we observed between activity evoked by lateralized tracking and memory tasks in Experiment 1. It appears that the contralateral difference wave is more stable in the presence of motion even if the motion is irrelevant to the task at hand. Further, the large increase in amplitude in tracking as compared to memory tasks appears to be driven by the process of attending to motion.

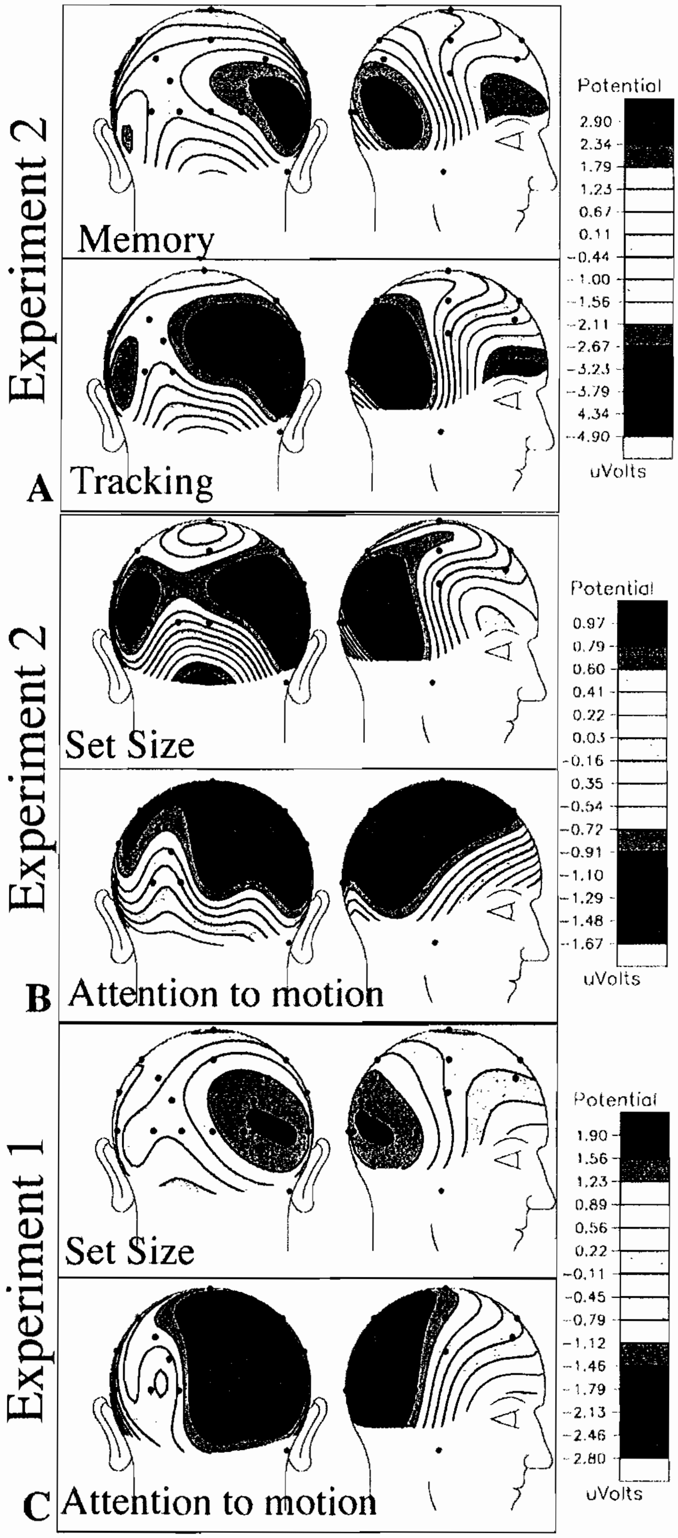

Figure 4-4: Topographic Maps From Experiments 1 and 2

EXPERIMENT 3

In Experiment 3, we aimed to extend the findings of Experiment 2 by further manipulating the presence or absence of motion. If attention to motion elicits a separate electrophysiological component than maintaining an item in working memory, then by transiently stopping and starting motion we should be able to turn this activity on or off without affecting the CDA-related activity. This may also allow us to estimate the time-course of the CAMA: when objects stop moving, how long will it be before this is reflected in the amplitude of the waveform? In this experiment, we asked participants to track 2 lateralized objects in 4 motion conditions: Normal, Pause, Stop and Never Move. On Pause trials all objects (including objects on the unattended side) stopped moving for 500ms and then began to move again. On Stop trials the objects stopped moving at the same point and never started moving again. In the Never Move trials, the objects never moved whereas all objects moved randomly throughout the trial in the Normal condition. Critically, all conditions were interleaved with identical initial selection periods of 500ms.

The data support the notion that differential contralateral amplitude decreases in the absence of attended motion. Amplitude for the 4 conditions was equivalent during the selection period prior to motion onset (200-300ms, F(3,33)=1.78, p= n.s.), but there was a significant effect of condition in all subsequent time windows (F(3,33)=10.2, 6.0 and 7.5 for the early, middle and late time periods respectively all ps<.003; see Figure 4-5). We used a priori paired t-tests to further probe these differences. In the early time period (1000-1500), before motion stoppage in the Pause and Stop conditions, the three moving conditions were statistically equivalent to one another (F(2,22)=1.38, p>.2) while the Never Move condition was significantly lower than the other three conditions (t(11)=3.10 (Normal), 3.02 (Pause), 3.96(Stop), all ps <.05). In the time period immediately following the stoppage of motion (1500-2000), amplitude for the Pause condition was significantly lower than amplitude in the Normal trial (t(11)=2.65, p<.05). In the final time window (2000-2500), after objects in the Pause condition began moving again, amplitude in this condition rose significantly higher than amplitude in the Never move condition (t(11)=3.53, p<.01) to a level equivalent to the Normal condition (t(11)=1.07, p>.3) while amplitude in the Stop condition was statistically equivalent to Never Move amplitude(t(11)=1.9, p>.05). In line with our predictions, amplitude in the Never Move condition follows a very similar pattern as the memory conditions in Experiment 1, slowly decreasing as the trial progresses. This is perhaps not surprising as a tracking trial without motion is equivalent to a location working memory trial.

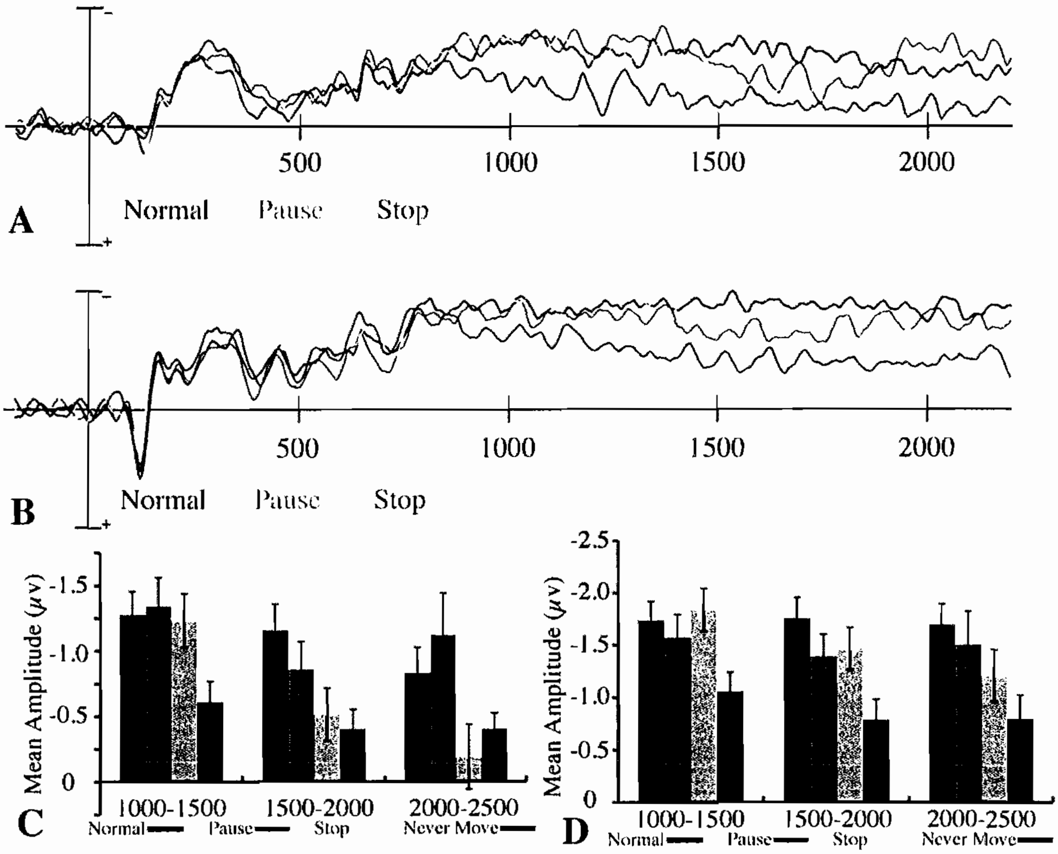

Figure 4-5: Experiment 3 and 3a Results

One concern with the comparisons between Experiments 2 and 3 is that the type of motion differs and as such may elicit a different pattern of electrophysiological responses thereby rendering any comparisons across the motion types less meaningful. To address this issue, we replicated the effects Experiment 3 using the spinning pinwheel stimuli from Experiment 2. We replicated the four conditions from Experiment 3 in Experiment 3a. In each condition, the participant was asked to track two bars. Although overall CDA amplitude was higher in the rotating pinwheel version of the experiment (mean amplitude for the Normal condition = -.78, -1.7 for Experiment 3, 3a respectively t(12)=3.54, p<.0005), the pattern of data across the two experiments is strikingly similar. As in experiment 3, there was a significant effect of trial type in each time window after the selection period (F(3,33)=20.43, 27.75 and 17.78 for the early, middle and late time periods respectively all ps<.001), but no effect of condition during the selection period (F(3,33)=1.73, p=n.s.). No Move amplitude was lower than the other 3 conditions in the early time window (t(11)=6.25 (Normal), 5.65 (Pause), 6.61(Stop), all ps <.05) and Pause and Stop amplitudes were significantly lower than Normal amplitude during the middle time window following the initial cessation of motion (t(11)=2.56, 5.5 respectively both ps<.001). During the late time period after object began moving again in the Pause condition, amplitude rose significantly above No Move amplitude (t(11)=4.84, p<.005) so that it was equivalent to Normal amplitude (t(11)=2.09, p>.05) and Stop amplitude was equivalent to amplitude in the No Move condition (t(11)1.95, p>.05).

Time course of the attention to motion effect

The perception of animated motion is an inherently cognitive act as we compare previous object location to current object location and interpolate dynamic motion from one point to another. As such, perception of the stoppage of motion may not be a simple, automatic process. To estimate the latency of the attention to motion effect, we subtracted amplitude in the pause condition from amplitude in the normal condition in Experiments 3 and 3a (Figure 4-6). In both experiments motion stopped at 1182ms and began again at 1682ms. Although the timing of the motion stoppage in these experiments was identical, the type of motion (many small, randomly moving boxes or two large, rotating pinwheels) was substantially different. It was therefore quite surprising how similar the pause effect was for the two experiments. First, we used a 50ms sliding window analysis to estimate the latency of the observed effect. Using this coarse level of analysis, the two experiments showed a similar time-course with both showing a significant difference from 1525 -1875ms and Experiment 3a becoming significant 100ms earlier at 1425ms. We also computed the point at which 25% of the area under the curve was reached (a fractional area latency analysis) and found that the pause effect reached this point at 1577ms in Experiment 3 and 1558ms in Experiment 3a. The latency of this effect was statistically equivalent across the two experiments (independent samples t-test: t(22)=.868, p=n.s.) and we found a similar result using a fractional area peak latency measure (25% fractional peak latency for Experiment 3: 1625ms; 3a: 1532ms, t(22)=1.79 , p=n.s.). In sum, across two experiments using different types of motion, we found that the latency of the attention to motion effect was consistently between 300-400ms post motion stoppage. While numerous previous neuroimaging studies have demonstrated that attended motion leads to higher activity in area MT+ (among other areas), this is the first demonstration of an enhanced electrophysiological response to attended motion. We are not aware of any studies that have examined the latency of attention to motion effects in humans, but Seidemann & Newsome (Seidemann and Newsome, 1999) measured the unit response in area MT of the macaque and found a similar estimate for the time course of the attentional enhancement of a preferred motion direction. Here, the firing rate for preferred motion did not become significantly higher than firing rate for the null direction until approximately 250ms after motion onset. The attentional effects in these two experiments are very different: one apparently connoting that motion of a preferred direction is in an attended RF whereas the attention effect that we observed is related to the perception of abrupt motion stoppage. Still, this provides an important illustration that attention effects in area MT in of the macaque monkey have a similar time course to an attention to motion effect that we hypothesize to be emanating from the human analogue of area MT.

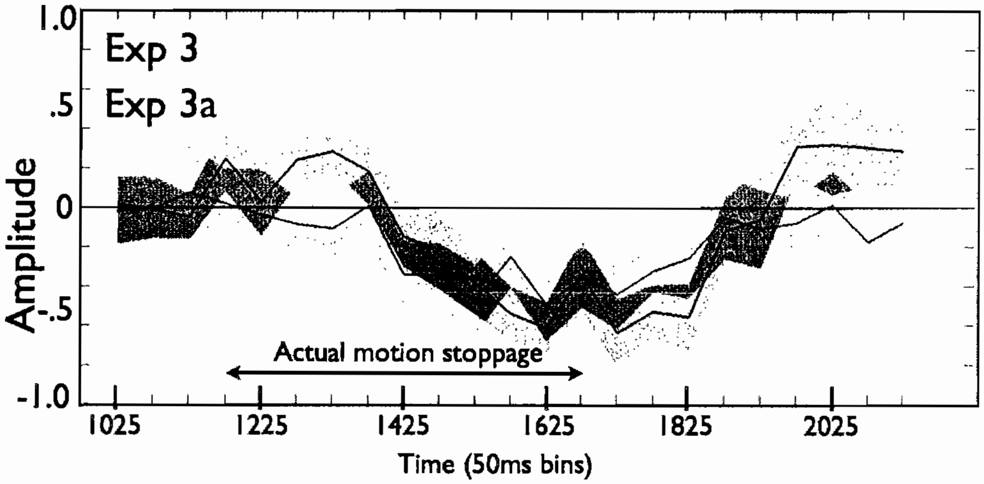

Figure 4-6: Time Course of Attention to Motion Effect

Experiments 3 and 3a show a remarkably similar pattern of results. Both indicate that in the absence of the necessity to update target information, contralateral amplitude decreases to a level that is equivalent to amplitude during a VWM trial with the same number of targets. Given the results from Experiment 2, we believe that this decrease in amplitude is due to a decrease in CAMA amplitude while the number of items that must be represented remains constant. Clearly, multiple object tracking is a complicated, multifaceted task, but the current set of results lend credence to the idea that two important aspects of the MOT task are a pointer system that indicates what items are targets and an attentional system that continuously updates the current location of these targets. When it is no longer necessary to update target information, the neural signature of this task becomes quite similar to a VWM task, where the pointer system alone is necessary.

DISCUSSION

Despite growing popularity as a paradigm to explore divided attention and object-based attention, the neural mechanisms that underlie MOT are not yet well understood. In the current set of studies, we have used what is known of the neural mechanisms that underlie VWM and attended motion to help us better understand how these tasks relate to MOT. Using lateralized versions of the VWM and MOT tasks, we found that a contralateral component evoked by both tasks is sensitive to the number of items that are being currently attended. We observed two main differences in the electrophysiological response: decay in amplitude during the maintenance period of the VWM task that was not evident in the tracking task, and an overall increase in amplitude during tracking relative to comparable VWM trials. In subsequent experiments, we determined that the decay of amplitude does not take place in the presence of irrelevant, unattended motion and that the main effect of amplitude appears to be driven by attention to motion. Using simple subtraction logic, we were able to isolate the activity related to attention to motion and found that, unlike the CDA, the component was not sensitive to the number of items being attended. Furthermore, scalp topography suggests that the CAMA is more broadly distributed and anterior than the CDA.

FMRI studies of MOT have shown a relatively consistent network of activity in areas such as SMA, FEF, SPL, IPS and MT+ (Culham et al., 2001; Culham and Kanwisher, 2001; Jovicich et al., 2001; Howe et al., 2009) (Culham et al., 1998; Culham et al., 2001; Jovicich et al., 2001; Howe et al., 2009). Unfortunately, due to the poor temporal resolution of this technique, it is unclear whether the activity in these regions is due to initial selection of targets, active tracking of targets or response selection. In an effort to avoid the response selection problem, Howe and colleagues did not ask for a response at the end of each trial, and found a very similar network but no activation in SMA. In each of these studies as well as the current study, participants were instructed to fixate during tracking. As participants generally move their eyes during MOT tasks, activation in FEF may be due to either saccade planning or inhibition (Fehd and Seiffert, 2008). Both studies that manipulated target load found that activation in IPS increased with increased load. Culham and colleagues found that activation in FEF, SPL and MT+ showed greater task activation (active tracking vs passive viewing) that load activation (activation that increased as the number of targets increased). Jovivich and colleagues (2001) found a similar pattern in FEF and MT+, but reported that SPL was load dependent. However, Howe et al., have suggested that the area defined as SPL by Jovivch was actually closer to IPS, which was load dependant in Culham et al. (2001) as well. Nonetheless, in the current study we found that CDA amplitude was sensitive to load manipulations, while the CAMA was sensitive to presence of attended motion and insensitive to a load manipulation. Furthermore, although the low-density ERP recordings in the current study bar strong statements about localization, using subtractive logic we attempted to isolate effects specific to attention to motion and increasing the overall number of targets. In general, the topography of the attention to motion effect was more broad and anterior than the activity related to attending an increasing number of targets. Taken together, our data is consistent with the idea that attention to motion leads to a categorically different pattern of activity than tracking or maintaining object information. To our knowledge this is the first account of an electrophysiological component that is sensitive to the presence or absence of attended motion.

In Experiments 3 and 3a, we were able to estimate the time-course of this effect and found that the attention to motion effect first became significant roughly 300-400ms post movement stoppage in both experiments. This estimate is in line with time course estimations made in the unit-recording literature for a different type of attention to motion effect (Seidemann & Newsome, 1999). While this effect is by definition an effect of preferential attention towards a specific direction of motion, it is less clear why we observe a decrease in amplitude during motion stoppage in the current study. Although the decrease in amplitude we observed is clearly related to attention to motion, it is not clear what aspect of attention to motion the effect connotes. Similar to the previously mentioned paper, the effect may be driven by attention to moving items and therefore decrease in the absence of motion. On the other hand, the effect may be specifically tied to the need to continuously update target information during the tracking interval. Our data cannot differentiate between these two interpretations of the data. One way to address this ambiguity would be to create a situation where it is necessary to update in the absence of motion. If the CAMA effect is driven by the need to update rather than attention to motion, updating in the absence of motion should result in a large contralateral increase in amplitude relative to a condition where the same information must be maintained, but not updated.

The relationship between working memory and updating

A multiple object tracking trial without movement is functionally equivalent to a location-based working memory trial. Perhaps not surprisingly, the two trial types evoke similar electrophysiological responses. In the fMRI literature, there seems to be a clear consensus that the IPS is an important area for both MOT and VWM tasks. In both cases, activity increases monotonically as the number of targets increases (Culham et al., 2001; Jovicich et al., 2001; Todd and Marois, 2004, 2005). The VWM literature has demonstrated that this activity ceases to rise once WM capacity is exceeded, but this result has not yet been extended to the MOT literature. Interestingly, when Howe et al. (2009), subtracted activity during static tracking (essentially VWM) trials from passive viewing of moving stimuli, the only area that was more active during the stationary task was posterior IPS (PIPS). This suggests that activity in this region codes for the number of items that are being actively attended regardless whether the items are moving or stationary. Anterior IPS (AIPS) activity did not differ in the stationary and passive tracking tasks, but was more active during active tracking than in passive viewing or stationary trials. This seems in line with Xu and Chun's (2006) finding that during a VWM task with simple or complex items, PIPS increased with the number of locations to be attended irrespective of the complexity of the items whereas AIPS was sensitive to the both the number and complexity of the objects. In the MOT context then PIPS would serve as a spatial index of what locations contain targets while AIPS seems tied to more complex computation necessary to update these location tags as the objects move. While previous work from our lab has shown that CDA activity behaves similar to PIPS activity during both MOT and VWM tasks, the current study clearly demonstrates that the one of the primary differences between the activity evoked by these two tasks is related to attention to motion. Importantly, we have found that activity related to attention to motion behaves much differently than attention to individual items. This activity appears to be an all or none response that is unaffected by the number of targets.

In exploring the difference between MOT and VWM, we found a number of clear distinctions in terms of electrophysiological response that we believe to the indicative of underlying differences in the computations that necessary to successfully perform both tasks. The tasks share a common requirement to index a number of targets, and we believe this process is reflected by the CDA this pointer system. However, the clearest difference between the two tasks is the need to attend to motion during MOT such that the current location of each target is continuously updated as they move. Our previous work has shown that individual differences in CDA amplitude are predictive of tracking ability: in short, individuals whose CDA amplitude does not rise from 1 to 3 items tend to be poor trackers (Drew & Vogel, 2008). While the pointer system appears to be a critical part of both VWM and MOT, the need to attend to motion such that target locations may be continuously updated differentiates the two tasks. We believe that the CAMA, a contralateral negativity with a broader, more anterior distribution that rides on top of the CDA during typical MOT tasks is an index of this process. Interestingly, although there was substantial variability in magnitude of the CAMA, differences in this activity do not appear to correlate with behavior. That is, good trackers did not appear to have a larger CAMA in experiment 2 or a exhibit a quicker decrease in amplitude in response to motion stoppage in Experiment 3. This suggests that this component may serve an index of whether motion is being attended or not rather than the quality of the motion representation that is processed. Future experiments will be needed to establish the functional role of this component during MOT, but the current study makes it clear that contralateral activity can be used as an online metric of attention to motion and that the time-course of this effect is similar to an attention to motion effect found in the unit-recording literature (Seidemann and Newsome, 1999).

MATERIALS AND METHODS

Participants

We analyzed the data of 13 subjects in Experiment 1, 16 in Experiment 2, 12 in Experiment 3 and 12 in Experiment 3a. Ages ranged from 18-28 and all participants gave informed consent according to procedures approved by the University of Oregon and were paid $10 for participation. All participants reported no history of neurological problems, normal color vision and normal or corrected-to-normal visual acuity.

Stimuli and procedures

Experiment 1. On half of the blocks participants tracked 1 or 3 items, while in others they held 1 or 3 items in memory. In both cases, the initial selection period was 500ms and was followed by a 1500ms interval where subjects either tracked items as they moved randomly about the screen, or maintained the object information across the delay interval. Each trial began with a 200ms arrow cue followed by an inter-stimulus interval that varied between 100 and 200ms. At the end of each change detection trial, the items from the selection period reappeared and participants were asked to categorize the items as either ‘same’ or ‘different’ with a game-pad controller. In tracking trials, one item was filled in red and participants were asked to judge whether the item in question was originally red or not (‘same’ or ‘different’ than the original color). In both cases, the correct answer was ‘different’ on 50% of trials. Order of tracking and change detection blocks was counterbalanced across participants. In both cases, the objects were squares that subtended .6 degrees of visual angle.

All the objects moved randomly throughout the tracking trials, bouncing whenever they made contact with other objects or the invisible motion bounding area (a 10.5 X 4.5 rectangle that was offset 2.1 degrees lateral to the fixation cross). Velocity and direction of motion also changed at random intervals during the trials. Average velocity was 1.6 degrees/second.

Experiment 2. There were 4 conditions in this blocked design experiment. Each trial began with a 500ms arrow cue that was followed by a 32ms inter-stimulus interval. During the tracking blocks, participants were asked to track one or two bars on lateralized spinning pinwheels (two perpendicular bars joined at the center of each bar; See Cavanagh & Alvarez, 2005) and to keep track of the cued bars as the spinners spun randomly for 2500ms. During the memory block, participants were asked to memorize the initial color of one or two bars on the spinners. We created a color set of seven equiluminant colors that varied smoothly between red and green, making this a difficult memory task. After the 500ms selection period, the cue colors disappeared and the spinners changed rotation speed and/or rotation direction at random intervals so that the motion was unpredictable. The average rotation rate was ~165 degrees/s. Participants were instructed to ignore the motion during the memory blocks and needed to track the rotation of the target bar in the tracking block. Similar to the change detection paradigm in Experiment 1, at the end of each memory trial, colors were replaced on the bars in the same position as in the beginning of the trial and participants were asked to judge whether the colors were 'same' or 'different.' In tracking trials, one bar on the attended side was illuminated red and participant had to identify it as either a target or distractor. Each bar was 2.9 degrees long with a width of 0.3 degrees. The pinwheels were arranged at the corners of a 5.6 x 5.6 degree box centered at the fixation cross meaning that each pinwheel was 1.34 degrees lateralized from the center of the screen at it's closest point.

Experiment 3. The lateralized tracking procedure from Experiment 1 was mimicked unless otherwise noted. There were 4 conditions in this experiment. In the 'Pause' condition all objects on both the attended and unattended sides were stationary for 500ms between 1182 and 1682ms in the trial, then began moving again. In the 'Stop' condition, all items stopped moving at the same point in time and never began to move again, remaining stationary until the end of the trial. In the 'No Move' condition, all objects remained stationary for the duration of the trial. Finally, in the 'Normal' condition, all the objects moved randomly for the duration of the trial. All trial types were interleaved, and were deliberately made to appear indistinguishable during the selection period of 500ms at the beginning of each trial.

Experiment 3a. The rotating pinwheel stimuli and motion parameters from Experiment 2 were mimicked. Unlike Experiment 2, there were two targets in each trial in this experiment and the targets and distractor bars were equiluminant red and green respectively. There was no explicit location cue in this experiment as the participants were simply told to attend the red bars and ignore the green. The conditions and timing from Experiment 3 was mimicked so that both experiments had the same 4 conditions: Stop, No Move, Never Move and Normal.

Electrophysiological recording and analysis

ERPs were recorded in each experiment using our standard recording and analysis procedures (McCollough et al., 2007; Drew and Vogel, 2008). We rejected all trials that were contaminated by blocking, blinks or large (>1 degree) eye movements. If more than 25% of trials were rejected for these reasons the participant's data was omitted from further analysis. In total, we excluded 7 of the 60 participants that participated in the study based on this criterion. There were 4 conditions in each experiment and participants completed 160 trials in each condition in all 4 experiments. All 4 experiments were divided into blocks that lasted roughly 5 minutes. In Experiment 1, the order of block type (memory or tracking) was counterbalanced across participants. Experiment 2 used a set order of ignore motion blocks followed by attend motion in an effort to avoid participants unnecessarily attended the irrelevant motion.

We recorded from 22 tin electrodes mounted in an elastic cap (Electrocap International, Eaton, OH) using the International 10/20 System. 10/20 sites F3, FZ, F4, T3, C3, CZ, C4, T4, P3, PZ, P4, T5, T6, O1 and O2 were used along with 5 non-standard sites: OL midway between T5 and O1; OR midway between T6 and O2; PO3 midway between P3 and OL; PO4 midway between P4 and OR; POz midway between PO3 and PO4. All sites were recorded with a left-mastoid reference, and the data were re-referenced offline to the algebraic average of the left and right mastoids. Horizontal electrooculogram (EOG) was recorded from electrodes placed approximately 1 cm to the left and right of the external canthi of each eye to measure horizontal eye movements. To detect blinks, vertical EOG was recorded from an electrode mounted beneath the left eye and referenced to the left mastoid. The EEG and EOG were amplified with a SA Instrumentation amplifier with a bandpass of 0.01-80Hz and were digitized at 250 Hz in LabView 6.1 running on a Macintosh. Contralateral and ipsilateral waveforms were defined based on the side of screen the participant attended on each trial. We computed a difference wave by subtracting ipsilateral activity from contralateral in each of the 8 paired electrodes (F3/4, C3/4, P3/4, PO3/4, T3/4, T5/6, OL/R, and O1/2). Finally, the resultant difference wave was averaged over a set of 5 occipito-parietal electrodes: P3/4, PO3/4, T5/6, OL/R, and O1/2. In computing the topographic maps in Figure 4-4, we collapsed across attend right and attend left trials by trading lateralized electrode sites for attend right trials such that the right hemisphere was always contralateral. Therefore, the topographic maps denote the average contralateral response on the right hemisphere and the average ipsilateral response on the left. Medial electrodes are simply the average amplitude during attend right and attend left trials. Each of the maps is a simple subtraction of amplitude in one condition from a different condition. The attention to motion effect was computed by subtracting average ignore motion amplitude from average attend motion amplitude. The set size effect was computed by subtracting Track 1 item amplitude from Track 2 items (or 3 items in Experiment 1) amplitude.

CHAPTER V

GENERAL CONCLUSIONS

One of the staples of cognitive psychology is that almost all papers begin with a connection to the real world. Visual search studies talk about airport security and medical screening. Working memory studies mention mental arithmetic or remembering a license plate number while dialing a number on your cell phone. These opening stanzas serve two functions: they give the reader a concrete metaphor that may help them understand why the experimenters are examining the idea in question and they inform the reader why the study might be important to people outside of the field. Invariably, these the connections are in the same direction: from the cognitive psychologist's abstract, reductionist world full of black and white boxes and neutral grey background, to the infinitely less controlled world that we all live in. This dissertation has followed the same basic blueprint: the end goal of all these studies on multiple object tracking is to better understand how people accomplish complex real world tasks like driving on crowded highways and keeping track of your children in a crowded playground. However, one of the goals of this dissertation was to edge slightly closer to ecological validity by applying a strong grounding in more basic attentional research to a relatively complicated task in MOT. Clearly, a better understanding of the neural underpinnings of tracking little black boxes as they move randomly about on a neutral grey screen is a long way from understanding what enables a person to keep track of the slow truck in front of him while merging into the fast lane to the left. But, hopefully by continuing down this path of building upon the research of predecessors towards more ecologically valid paradigms, we can move towards research that directly applies to real world issues.

In Chapter II, we adapted a known paradigm for studying visual working memory and attempted to apply it to MOT. The two tasks are similar in that both ask observers to select a variable number of targets at the onset of each trial, but differ in what the observer is then asked to do with this information. In VWM task, the information must simply be held for some period so that when subsequently queried about the target information, they can accurately retrieve or recognize the information. In the MOT task, the observer must update the location information for each target as the targets move randomly so that they are capable of identifying the targets again at the end of motion period. We were surprised to find that the activity evoked by these two tasks was quite similar: a large negative slow wave emanating from posterior electrodes sites that was larger at contralateral than ipsilateral sites. We found that, similar to the VWM paradigm that was the inspiration for this study, the contralateral-ipsilateral difference at posterior sites (the CDA) increased as a function of the number of targets on a given trial. Further, behavioral tracking ability was found to correlate with this component such that poor trackers tended to show a smaller difference in CDA amplitude when the tracking load was increased from 1 to 3 items than good trackers. This implies that poor trackers may have suffered at the task because they were unable to increase the number of targets they were able to effectively track as efficiently as good trackers. We also manipulated the difficulty of tracking while holding the number of targets constant by adjusting the area of motion. The difficulty manipulation did not affect amplitude, suggesting that the CDA is a marker of the *number* of targets that are currently being attended and is not sensitive to the amount of attentional resources that must devoted to each target.

Chapter III examined the role of attention during tracking. This study was in response to a number of studies in the MOT literature that have claimed that one of the primary roles of attention during MOT is to suppress or inhibit distractors. This effect was demonstrated (Pylyshyn, 2006) using what is known as 'dot-probe' technique and has since been replicated a number of times (Flombaum et al., 2008; Pylyshyn et al., 2009). In this technique, observers are asked to track object while simultaneously monitoring the display for brief probes that occur on a subset of trials. Probes could occur on targets, distractors or empty space. Detection of probes was taken as a measure of the locus of attention during the tracking task. Pylyshyn and colleagues initially found that probe performance was highest for empty space, then targets, with detection for distractors the lowest. Critically, this pattern of results is ambiguous with respect to attentional enhancement or suppression for the targets and distractors because the baseline condition (empty space) showed the highest rate of detection. However, Pylyshyn then asked observers to perform the same probe detection task in the absence of any tracking requirement and found that detection was higher on empty space than moving targets (presumably due to lateral masking). Using this data, Pylyshyn computed a corrected probe detection rate and found that performance for probes in empty space and targets was equivalent, with distractor performance significantly lower.

We wondered if this effect was an artifact of the dual task situation observers were placed in during these experiments. That is, asking the observers to keep track of two tasks at once may have changed the typical allocation of attentional during tracking in absence of an additional task. To address this issue, we asked observers to ignore probes while focusing on tracking. We recorded the electrophysiological response to the task-irrelevant probes as a function of their location. We found that the early visual evoked responses were largest for targets, with probes on distractors, empty space and stationary objects all equivalent to one another. There is a large literature that has linked modulations of these components to the focus of spatial attention (e.g. Heinze et al., 1990; Heinze et al, 1994; Hillyard et al., 1998). This pattern of results suggests that spatial attention enhances target locations during tracking, with distractors and empty space both being treated equally. We found no evidence of distractor suppression. Although this does not refute the previous finding of distractor suppression during MOT, it does draw into question the level of processing that manifested the previously observed effect. As the early attention mechanisms of spatial attention exhibit no evidence of suppression, perhaps the effect is due to a later effect such different thresholds for reporting a probe on items that are being tracked and those that are being ignored for the tracking task. Further work will be necessary to address this hypothesis.

In chapter IV, we directly compared lateralized versions of the VWM and MOT task. Although both tasks elicited a CDA component that was sensitive to the number targets on a given trial, there were two clear differences in the evoked activity for the two tasks:

1. In the VWM task the CDA decayed approximately 1000ms after offset of the targets while no decay was observed during MOT.
2. Amplitude of the CDA was much larger in the MOT task, even in the early period of the VWM task when amplitude was maximal.

We then manipulated the presence or absence of attended motion and found that during a VWM task in the presence of irrelevant motion, amplitude does not decay but amplitude in this task was still lower than amplitude in a difficulty matched tracking task. This suggests that the amplitude decay observed in typical lateralized memory tasks is due to the absence of motion (attended or unattended), while the amplitude increase we observe in tracking tasks is specifically tied to the need to attend to motion. Unexpectedly, we found that the amplitude increase related to attention to motion was not sensitive to target number manipulations and appeared to emanate from a more broadly distributed anterior region than the CDA. We have hypothesized that this component, which we have termed the CAMA (contralateral attention to motion activity), is due to MT+ activity. Several fMRI studies of MOT corroborate this claim. Specifically, activity in area MT+ shows a large increase in activity in the presence or absence of attended motion and is relatively insensitive to increases in target load (Culham et al., 2001; Jovicich et al., 2001). Furthermore, when MOT is contrasted with a static MOT trial where the objects never move and the observer must simply encode the original location of the targets, area MT+ is much more active in the presence of attended motion. This same contract showed that activity in the posterior IPS had an equivalent amount of activity for both normal and static MOT trials. This area is often associated with working memory representations and they interpreted this pattern of activity as evidence in favor of the idea that it is necessary to represent each target in working memory. It was therefore not surprising that when we manipulated the presence or absence of motion, we observed a rapid decrease in CAMA amplitude in the absence of motion. We found that static MOT trials elicited a pattern of activity that was strikingly similar to VWM activity. All of this seems to suggest that using electrophysiological recordings during MOT we were able to isolate two distinct processes that are both vital to this task: an indexing system that individuates each target and a continuous updating system that adjusts the current location of each index as the objects move.

Together, the 3 studies presented here demonstrate the utility of using ERPs to examine a complicated task in MOT. Through each study, the general approach was to take a known entity and apply this knowledge to a new question. Using this approach, our understanding of the neural mechanisms that allow observers to track multiple independent objects simultaneously has increased substantially. A dominant theme through all of the studies is that attention appears to play a number of different roles over the course of a single MOT trial. Chapter II showed that attention is necessary to initially select the target objects before they begin to move and that a similar mechanism is active during the tracking phase of the trial. Chapter III showed that spatial attention focuses on target locations during tracking, while not differentiating between empty space and distractor locations. Chapter IV showed that in addition to the attentional indexing that is evident in Chapter II during tracking, attention to task relevant motion appears to be a separate process that also operated during typical MOT trials. Although there is a tendency in the MOT literature to discuss the process of tracking as a unitary construct, the current study clearly demonstrates that this is not the case. If we are to continue to move forward in our understanding of this task, it will be important to acknowledge that different aspects of MOT map onto different types of attention.

REFERENCES

Alvarez G. A., Cavanagh P., (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science 15*, 106-111.

Alvarez G. A., Cavanagh P., (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science 16*, 637-643.

Alvarez G. A., 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., Horowitz T. S., Arsenio H. C., DiMase J. S., Wolfe J. M., (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology-Human Perception and Performance 31*, 643-667.

Alvarez G. A., Scholl B. J., (2005). How Does Attention Select and Track Spatially Extended Objects? New Effects of Attentional Concentration and Amplification. *Journal of Experimental Psychology: General 134(4)*, 461-476.

Avarahami J., (1999). Objects of attention, objects of perception. *Perception & Psychophysics 61*, 1604-1612.

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.

Awh E., Dhaliwal H., Christensen S., Matsukura M., (2001). Evidence for two components of object-based selection. *Psychological Science 12*, 329-334.

Awh E., Pashler H., (2000). Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance 26*, 834-846.

Buschman T. J., Miller E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science 315*, 1860-1862.

Castiello U., Umilta C., (1992). Splitting focal attention. *Journal of Experimental Psychology: Human Perception and Performance 18*, 837-848.

Cave K. R., Bichot N. P., (1999). Visuospatial attention: Beyond a spotlight model. *Psychonomic Bulletin & Review 6*, 204-223.

Cave K. R., Zimmerman J. M., (1997). Flexibility in spatial attention before and after practice. *Psychological Science 8*, 399-403.

Cepeda N. J., Cave K. R., Bichot N. P., Kim M.-S., (1998). Spatial selection via feature-driven inhibition of distractor locations. *Perception & Psychophysics 60*, 727-746.

Chelazzi L., Desimone R., (1999). Serial attention mechaisms in visual search: A critical look at the evidence. *Psychological Research, 62*, 195-219.

Chelazzi L., Miller E. K., Duncan J., Desimone R., (1993). A neural basis for visual search in inferior temporal cortex. *Nature 363*, 345-347.

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.

Cohen A., Ivry R. B. (1991). Density effects in conjunction search: Evidence for a coarse location mechanism of feature integration. *Journal of Experimental Psychology: Human Perception and Performance 17*, 891-901.

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

Cronbach L. J., (1957). The two disciplines of scientific psychology. *American Psychologist* 12, 671-684.

Culham J .C., Cavanagh P., Kanwisher N. G., (2001). Attention response functions: Characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron 32*, 737-745.

Culham J. C., Brandt S A, Cavanagh P, Kanwisher N G, Dale A M, Tootell R B H (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology 80*, 2657-2670.

Delvenne J. F. (2005). The capacity of visual short-term memory within and between hemifields. *Cognition 96*, B79-B88.

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

Duncan J., (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review 87*, 272-300.

Duncan J., (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General 113*, 501-517.

Duncan J., Bundesen C., Olson A., Humphreys G., Chavda S., Shibuya H., (1999). Systematic analysis of deficits in visual attention. *Journal of Experimental Psychology: General 128*, 450-478.

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

Engle R. W. (2002) Working memory capacity as executive attention. *Current Directions in Psychological Science 11*, 19-23.

Egly R., Driver J., Rafal R. D., (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General 123*, 161-177.

Eimer M., (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and clinical Neurophysiology 99*, 225-234.

Eriksen C. W., St. James J. D., (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics 40*, 225-240.

Eriksen C. W., Yeh Y. Y., (1985) Allocation of attention in the visual field *Journal of Experimental Psychology: Human Perception and Performance 11*, 583-597.

Fehd H. M., Seiffert A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition 108*, 201-209.

Fencsik D. E., Urrea J., Place S. S., Wolfe J. M., Horowitz T. S., (2006). Velocity cues improve visual search and multiple object tracking. *Visual Cognition 14(1)*, 92-95.

Feria, C. S. (2008). The distribution of attention within objects in multiple-object scenes: Prioritization by spatial probabilities and a center bias. *Perception & Psychophysics, 70(7)*, 1185-1196.

Flombaum, J. I., Scholl, B. J., & Pylyshyn, Z. W. (2008). Attentional resources in tracking through occlusion: The high-beams effect. *Cognition, 107*, 904-931.

Fougnie D, Marois R, (2006). Distinct Capacity Limits for Attention and Working Memory: Evidence From Attentive Tracking and Visual Working Memory Paradigms. *Psychological Science 17, 526-534.*

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

Gibson B. S., Li L., Skow E., (2000). Searching for one versus two identical targets: When visual search has a memory. *Psychological Science 11*, 324-327.

Hahn S., Kramer A., (1998). Further evidence in favor of division of the spotlight among non-contiguous locations. *Visual Cognition 5*, 217-256.

Hallett P. E., (1986). Eye movements, in *Handbook of perception and human performance* Eds K R Boff, L Kaufman and J P Thomas (New York: New York) pp 10.1-10.112.

Handy T. C., Mangun G. R., (2000). Attention and spatial selection: Electrophysiological evidence for modulation by perceptual load. *Perception & Psychophysics 62*, 175-186.

Harter M. R., Miller S. L., Price N. J., LaLonde M. E., Keyes A. L. (1989). Neural processes involved in directing attention. *Journal of Cognitive Neuroscience 1*, 223-237.

Heinze H. J., Luck S. J., Mangun G. R., Hillyard S. A. (1990). Visual event-related potentials index focused attention within bilateral stimulus arrays. I. Evidence for early selection. *Electroencephalography and Clinical Neurophysiology, 75*, 511-527.

Heinze, H. J., Mangun, G. R., Burchert, W., Hinrichs, H., Scholz, M., Münte, T. F., et al. (1994a). Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature, 372*, 543-546.

Heinze, H. J., Luck, S. J., Münte, T. F., Gös, A., Mangun, G. R., & Hillyard, S. A. (1994b). Attention to adjacent and separate positions in space: An electrophysiological analysis. *Perception & Psychophysics, 56*, 42-52.

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society: Biological Sciences, 353*, 1257-1270.

Hillyard S. A., Galambos R. (1970). Eye movement artifact in the CNV. *Electroencephalography and Clinical Neurophysiology 28*, 173-182.

Hopf J.-M., Luck S. J., Girelli M., Hagner T., Mangun G. R., Scheich H., Heinze H. J. (2000). Neural sources of focused attention in visual search. *Cerebral Cortex 10*, 1233-1241.

Hopf J.-M., Boelmans K., Schoenfeld A. M., Heinze H. J., Luck S. J. (2002). How does attention attenuate target-distractor interference in vision? Evidence from magnetoencephalographic recordings. *Cognitive Brain Research 15*, 17-29.

Hopf J. M., Luck S. J., Boelmans K., Schoenfeld A. M., Boehler C. N., Rieger J., Heinze H. (2006). The neural site of attention mtaches the spatial scale of perception. *Journal of Neuroscience 26*, 3532-3540.

Horowitz T. S., Klieger S. B., Fencsik D. E., Yang K. K., Alvarez G. A., Wolfe J. M., (2007). Tracking unique objects. *Perception & Psychophysics 69*, 172-184.

Horowitz T. S., Wolfe .J M., (1998). Visual search has no memory. *Nature 394*, 575-577.

Howe P. D., Horowitz T. S., Morocz I. A., Wolfe J., Livingston M. S. (2009). Using fMRI to distinguish components of the multiple object tracking task. *Journal of Vision 9*, 1-11.

Huk A. C., Dougherty R. F., Heeger D. J. (2002) Retinotopy and functional subdivision of human areas MT and MST. *Journal of Neuroscience 22*, 7195-7205.

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

Itti L., Koch C., (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research 40*, 10-12.

Itti L., Koch C., (2001). Computational modelling of visual attention. *Nature reviews neuroscience 2*, 194-203.

James W., 1890 *The Principles of Psychology* (New York: Holt).

Jiang Y., Olson I. R., Chun M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory & Cognition 2*, 683-702.

Jovicich J., Peters R. J., Koch C., Braun J., Chang L., Ernst T., (2001). Brain areas specific for attentional load in a motion-tracking task. *Journal of Cognitive Neuroscience 13*, 1048-1058.

Kane M. J., Engle R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review 9*, 637-671.

Kane, M. J. & Engle, R. W. (2003). Working memory capacity and the control of attention: the contribution of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General 132*, 47-70.

Kahneman D., Treisman A., (1984). Changing views of attention and automaticity, in *Varieties of Attention.* Ed R P a R Davies (London: London) pp 29-61.

Kahneman D., Treisman A., Gibbs B. J., (1992). The reviewing of object files: Object–specific integration of information. *Cognitive Psychology 24*, 175–219.

Kastner S., Ungerleider L. G. (2000) Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience 23*, 315-341.

Kazanovich Y., Borisyuk R., (2006). An Oscillatory Neural Model of Multiple Object Tracking. *Neural Computation 18*, 1413-1440.

Klein R., (1988). Inhibitory tagging system facilitates visual search. *Nature* 334, 430-431.

Kraft A., Muller N. G., Hagendorf H., Schira M. M., Dick S., Fendrich R. M., Brandt S. A., (2005). Interactions between task difficulty and hemispheric distribution of attended locations: Implications for the splitting attention debate. *Cognitive Brain Research 24*, 19-32.

Kramer A., Weber T. A., Watson S. E., (1997). Object-based attentional selection--Grouped-arrays or spatially invariant representations?: Comment on Vecera and Farah. (1994) *Journal of Experimental Psychology: General 126*, 3-13.

Kramer A. F., Hahn S., (1995). Splitting the beam: Distribution of attention over noncontiguous regions of the visual field. *Psychological Science 6*, 381-386.

LaBerge D., Brown V., (1989). Theory of attentional operations in shape identification. *Psychological Review 96*, 101-124.

Lavie N., (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance 21*, 451-468.

Lavie N., Tsal Y., (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics 56*, 183-197.

Linden D. E. J., Bittner R. A., Muckli L., Waltz J. A., Kriegeskorte N., Goebel R., Singer W., Munk M. H. J. (2003). Cortical capacity constraints for visual working memory: dissociation of fMRI load effects in a fronto-parietal network. *Neuroimage 20*, 1518-1530.

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 and Performance 31,* 235-247.

Luck, S. J. (1995). Multiple mechanisms of visual-spatial attention: Recent evidence from human electrophysiology. *Behavioural Brain Research, 71*, 113-123.

Luck S. J., Hillyard S. A., Mangun G. R., Gazzaniga M. S., (1989). Independent Hemispheric Attentional Systems Mediate Visual-Search in Split-Brain Patients *Nature 342*, 543-545.

Luck S. J., Hillyard S. A., (1994a). Electrophysiological correlates of feature analysis during visual search *Psychophysiology 31*, 291-308.

Luck S. J., Hillyard S. A., (1994b). Spatial filtering during visual search: Evidence from human electrophysiology *Journal of Experimental Psychology: Human Perception and Performance 20*, 1000-1014.

Luck, S. J., Hillyard, S.A., Mouloua, M. Woldorff, M.G., Clark, V.P., Hawkins, H.L. (1994). Effects of spatial cuing on luminance detectability: Psychophysical and electrophysiological evidence for early selection. *Journal of Experimental Psychology: Human Perception & Performance*, *20*, 887-904.

Luck S J, Hillyard S A, (1995). The role of attention in feature detection and conjunction discrimination: An electrophysiological analysis *International Journal of Neuroscience 80*, 281-297.

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

Luck S. J., Chelazzi L., Hillyard S. A., Desimone R., (1997a). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex *Journal of Neurophysiology 77*, 24-42.

Luck S. J., Girelli M., McDermott M. T., Ford M. A., (1997b). Bridging the gap between monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention *Cognitive Psychology 33*, 64-87.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 1057-1074.

Martinez, A., Teder-Sälejärvi, W., Vazquez, M., Molholm, S., Foxe, J. J., Javitt, D. C., et al. (2006). Objects are highlighted by spatial attention. *Journal of Cognitive Neuroscience, 18*, 298-310.

Malinowski P., Fuchs M., Muller M. M., (2007). Sustained division of spatial attention to multiple locations within one hemifield *Neuroscience Letters 414*, 65-70.

Mazza V., Turatto M., Umilta C, Eimer M. (2007). Attentional selection and identification of visual objects are reflected by distinct electrophysiological responses. *Experimental Brain Research 181*, 531-536.

McCarthy G., Wood C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology 62*, 203-208.

McCollough A.W., Machizawa M. G., Vogel E. K. (2007). Electrophysiological measures of maintaining representations in visual working memory. *Cortex 43*, 77-94.

McMains S. A., Somers D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron 42*, 677-686.

McMains S. A., Somers D. C. (2005). Processing efficiency of divided spatial attention mechanisms in human visual cortex. *Journal of Neuroscience 25*, 9444-9448.

McNab, F. & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience, 11*(1), 103-107.

Miller G. A., (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review 63*, 81-97.

Moran J., Desimone R., (1985). Selective attention gates visual processing in the extrastriate cortex. *Science 229*, 782-784.

Müller M. M., Hübner R., (2002). Can the spotlight of attention be shaped like a donut? Evidence from steady-state visual evoked potentials. *Psychological Science 13*, 119-124.

Müller M. M., Malinowski P., Gruber T, Hillyard S. A., (2003). Sustained division of the attentional spotlight. *Nature 424*, 309-312.

Müller M. M., Teder-Sälejärvi W., Hillyard S. A., (1998). The time course of cortical facilitation during cued shifts of spatial attention. *Nature Neuroscience 1*, 631-634.

Nobre A. C., Sebestyen G. N., Miniussi C. (2000). The dynamics of shifting visuospatial attention revealed by event-related potentials. *Neuropsychologia 38*, 964-974.

Ogawa H., Takeda Y., Yagi A., (2002). Inhibitory tagging on randomly moving objects. *Psychological Science 13*, 125-129.

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

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.

Oksama L., Hyönä J., in press Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology.*

Pillow J., Rubin N., (2002). Perceptual completion across the vertical meridian and the role of early visual cortex. *Neuron 33*, 803-813.

Posner M. I., Dehaene S. (1994). Attentional networks. *Trends in Neuroscience, 17(2),* 75-79.

Posner M. I., Snyder C. R. R., Davidson B. J., (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General 109*, 160-174.

Pylyshyn Z., (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition 32(1)*, 65-97.

Pylyshyn Z., Burkell J., Fisher B., Sears C., Schmidt W., Trick L., (1994a). Multiple parallel access in visual attention. *Canadian Journal of Experimental Psychology 48*, 260-283.

Pylyshyn Z., Burkell J., Fisher B., Sears C., Schmidt W., Trick L., (1994b). Multiple Parallel Access in Visual-Attention. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale 48*, 260-283.

Pylyshyn Z. W., (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition 11*, 801-822.

Pylyshyn Z. W., (2006). Some puzzling findings in multiple object tracking (MOT): II. Inhibition of moving nontargets. *Visual Cognition 14*, 175-198.

Pylyshyn Z. W., Annan V., (2006). Dynamics of target selection in Multiple Object Tracking (MOT). *Spatial Vision 19*, 485-504.

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

Raffone A., Wolters G., (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience 13*, 766-785.

Scholl B. J., in press What have we learned about attention from multiple object tracking (and vice versa)? in *Computation, cognition, and Pylyshyn* Eds D Dedrick and L Trick (Cambridge, MA: Cambridge, MA) .

Scholl B. J., Pylyshyn Z. W., Feldman J., (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition 80(1-2)*, 159-177.

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

Seidemann E., Newsome W. T. (1999). Effect of spatial attention on the responses of area MT neurons. *Journal of Neurophysiology 81*, 1783-1794.

Sereno A., Kosslyn S. M., (1991). Discrimation within and between hemifields - a new constraint on theories of attention. *Neuropsychologia 29*, 659-675.

Shaw M. L., Shaw P., (1977). Optimal allocation of cognitive resources to spatial locations. *Journal of Experimental Psychology: Human Perception and Performance 3*, 201-211.

Shim W. M., Alvarez G. A., Jiang Y., in press Spatial separation between targets constrains maintenance of attention on multiple targets. *Psychonomic Bulletin & Review.*

Simpson G. V., Dale C. L., Luks T. L., Miller W. L., Ritter W., Foxe J. F. (2006). Rapid targeting followed by sustained deployment of visual spatial attention. *Neuroreport 17*, 1595-1599.

Sohn Y.-S., Liederman J., Reinitz M. T., (1996). Division of inputs between hemispheres eliminates illusory conjunctions. *Neuropsychologia 34*, 1057-1068.

Suganuma M., Yokosawa K., (2006). Grouping and trajectory storage in multiple object tracking: Impairments due to common item motions. *Perception 35*, 483-495.

Todd J. J., Marois R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature 428*, 751-754.

Todd J. J., Marois R. (2005). Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. *Cognitive Affective & Behavioral Neuroscience 5*, 144-155.

Tootell R. B. H., Reppas J. B., Dale A. M., Look R. B., Sereno M. I., Malach R., Brady T. J., Rosen B. R. (1995). Visual-Motion Aftereffect in Human Cortical Area Mt Revealed by Functional Magnetic-Resonance-Imaging. *Nature 375*, 139-141.

Tootell R. B. H., Mendola J. D., Hadjikhani N. K., Ledden P. J., Liu A. K., Reppas J. B., Sereno M. I., Dale A. M. (1997). Functional analysis of V3A and related areas in human visual cortex. *Journal of Neuroscience 17*, 7060-7078.

Townsend J. T., (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science 1*, 46-54.

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

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

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

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

Trick L. M., Pylyshyn Z. W., (1994). Why Are Small and Large Numbers Enumerated Differently - a Limited-Capacity Preattentive Stage in Vision. *Psychological Review 101*, 80-102.

Tsotsos J. K., (1995). Toward a computational model of visual attention, in *Early Vision and Beyond* Eds T V Papathomas, C Chubb, A Gorea and E Kowler (Cambridge, MA: Cambridge, MA) pp 207-218.

Unsworth N., Schrock J. C., Engle R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology-Learning Memory and Cognition 30*, 1302-1321.

vanMarle K., Scholl B. J., (2003). Attentive tracking of objects versus substances. *Psychological Science 14*, 498-*504.*

Van Velzen J., Eimer M. (2003). Early posterior ERP components do not reflect the control of attentional shifts toward expected peripheral events. *Psychophysiology 40*, 827-831.

Vecera S. P., Farah M. J., (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General 123*, 146-160.

Vogel E. K., Machizawa M. G., (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature 428*, 748-751.

Vogel E. K., McCollough A. W., Machizawa M. G., (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature 438*, 500-503.

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

Watson S., Kramer A., (1999). Object-based visual selective attention and perceptual organization. *Perception & Psychophysics 61*, 31-49.

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

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

Wolfe J. M., Horowitz T. S., Michod K. O., (2007). Is visual attention necessary for robust picture memory? *Vision Research 47*, 955-964.

Wolfe J. M., Pokorny C. W., (1990). Inhibitory tagging in visual search: A failure to replicate. *Perception & Psychophysics 48*, 357-362.

Woodman G., Min-Suk K., Rossi A. F., Schall J. D., (2007). Nonhuman primate event-related potentials indexing covert shirts of attention. *Proc Natl Acad Sci U S A 104*, 15111-15116.

Woodman G. F., Luck S. J., (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature* 400, 867-869.

Woodman G. F, Luck S. J., (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology-Human Perception and Performance 29*, 121-138.

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

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

Xu Y. (2002) Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception & Performance 28*, 458-468.

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

Yantis S., (1992). Multi-Element visual tracking - attention and perceptual organization. *Cognitive Psychology 24*, 295-340.

Yantis S., Hillstrom A. P., (1994). Stimulus-driven attentional capture: Evidence from equiluminant visual objects. *Journal of Experimental Psychology: Human Perception and Performance 20*, 95-107.

Yantis S., Johnson D. N., (1990). Mechanisms of attentional priority *Journal of Experimental Psychology: Human Perception and Performance 16*, 812-825.

Yantis S., Johnston J. C., (1990). On the locus of visual selection: Evidence from focused attention tasks. *Journal of Experimental Psychology: Human Perception and Performance 16*, 135-149.

Yantis S., Serences J., (2003). Cortical mechanisms of space-based and object-based attentional control. *Current Opinion in Neurobiology 13*, 187-193.