2 What Have We Learned about Attention from Multiple-Object Tracking (and Vice Versa)?

Brian J. Scholl

1 Introduction

If you weren't paying attention, you could be forgiven for thinking that this chapter was part of a collection assembled in honor of several people named Zenon Pylyshyn: the philosopher of psychology who has helped define the relation between mind and world; the computer scientist who has characterized the power of computation in the study of cognition; the cognitive psychologist whose imagery research is in every introductory textbook; and the vision scientist whose ideas and experimental paradigms form a foundation for work in visual cognition. (When I first learned of "Zenon Pylyshyn" in college, I figured that this couldn't really be someone's name, and given the breadth and importance of his contributions I figured that "he" must be some sort of research collective—a Nicolas Bourbaki of cognitive science. I was lucky to have been able to study later with this excellent research collective in graduate school, though I discovered it was housed in one head.)

This chapter is about the last of the Zenons noted above: the vision scientist. In the study of visual cognition, his lasting influence has stemmed in part from the way that he has bucked one of the most dangerous trends in experimental research: whereas most of us too easily fall into the trap of constructing theoretical questions to fit our experimental paradigms, Zenon has consistently managed the reverse. And there is perhaps no better example of this than his development of the multiple-object tracking (henceforth MOT) paradigm. This chapter focuses on the nature of MOT, with three interrelated goals: (1) to explore what makes MOT unique—and uniquely useful—as a tool for studying visual cognition; (2) to characterize the relationship between attention and MOT; and (3) to highlight some of the most important things we've learned about attention from the study of MOT—and vice versa.

2 Multiple-Object Tracking

Perhaps the most active area in visual cognition research in the last few decades has been the study of attention. Attention seems to involve a perceptual resource that can both intentionally and automatically select—and be effortfully sustained on—particular stimuli or activities. The core aspects of attention comprise three phenomena (Pashler 1998): (1) the fact that we can process some incoming stimuli more so than others (selectivity), (2) an apparent limitation on the ability to carry out simultaneous processing (capacity-limitation), and (3) the fact that sustained processing of visual stimuli seems to involve a sense of exertion (effort).

There is no paradigm that more viscerally illustrates these three components of attention than MOT (Pylyshyn and Storm 1988). One of the appeals of MOT is that at root it is a very simple task. In a typical experiment (see figure 2.1), observers initially see a number of identical objects. A subset of these are then flashed to indicate their status as targets, after which all of the (again identical) objects begin moving independently and unpredictably about the display. When they stop moving, observers must indicate which of the objects are the original targets.

2.1 What Makes MOT Special?

This procedure contrasts with most other paradigms that have been used to study attention in several ways. First, MOT requires continuous sus-

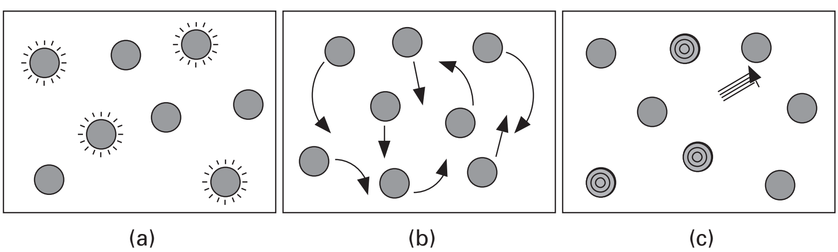


Figure 2.1
A schematic depiction of multiple object tracking. (a) Four items are initially flashed to indicate their status as targets. (b) All items then begin moving independently and unpredictably around the display. (c) At the end of the motion phase, the subject must move the cursor about the screen to highlight the four targets—here the subject has just highlighted three of the targets, and is moving the mouse to the fourth. Animations of many different variants of this task—including those of all the figures in this chapter—can be viewed at or downloaded from http://www.yale.edu/perception/.

tained attention over time rather than brief attentional shifts (as in spatial cueing studies). Second, MOT involves attention to multiple objects rather than focal attention to only a single object at a time (as in most attentional capture studies). Third, MOT is an inherently active task, rather than requiring mere passive vigilance (e.g., when waiting for a target to appear). Fourth, the magnitude of the attentional demands in MOT can be directly manipulated in terms of the underlying tracking load, rather than via indirect temporal manipulations (e.g., as used in the brief masked displays of many divided-attention experiments). Moreover, it is worth noting that each of these features is characteristic of real-world visual cognition: day-to-day experience is filled with situations—driving, hunting, sports, or even just trying to cross a street—that call for sustained attention to multiple objects over time and motion. As such, MOT has proven to be one of the most useful tools in the study of attention. (As a bonus, MOT typically yields relatively large and robust effects, making it ideal for studies that need to distinguish several different levels of performance, beyond simply demonstrating that various attentional effects do or do not exist.)

Perhaps the most central result in the study of MOT is simply that it is possible in the first place. As Pylyshyn and his colleagues have noted, this was not a foregone conclusion, given that classical theories of attention tended to assume a single unitary "spotlight" of selection. Since targets and distractors are spatially interleaved in MOT, though, the only natural way for a unitary spotlight to succeed would be if it cycled repeatedly from target to target, storing and updating their "last known addresses." This possibility seems implausible on its face, given the phenomenology of MOT: It certainly does not seem as if one's attention is constantly cycling around to different targets (though of course it is possible to attend to each of the objects independently or to consider them as a single global deforming shape—e.g., as a deforming polygon with targets at the corners; Yantis 1992). This is not an entirely empty point, perhaps, given the tight relation between attention and awareness (see Most et al. 2005): In most situations, you are at least somewhat aware of how and where you are attending. At the same time, however, phenomenology is often a poor guide to the underlying nature of the mind, and so that alone cannot definitively rule out a "single roving spotlight" explanation.

The initial report of MOT, however, effectively ruled out single-spotlight explanations via additional computational modeling results (Pylyshyn and Storm 1988). This model focused on how well a particular single-spotlight model could do when faced with actual MOT trajectories, when the spotlight was constrained to move at physiologically plausible speeds. Even given very generous assumptions about such speeds, the central result of this modeling project was that single-spotlight performance could never match actual human tracking abilities for those same trajectories. Moreover, this is true even when the spotlight is made as intelligent as we can think to make it—for example, employing subtle heuristics that involve extrapolating objects' trajectories over multiple temporal scales, and prioritizing objects in locally dense regions of the display from moment to moment (Chan et al., in preparation). The reasonable conclusion is that the underlying architecture of MOT must involve parallel selection and tracking—perhaps including up to four separate loci of attention, which might then directly explain the fact that tracking suffers beyond this number of targets (see Hulleman 2005).

2.2 MOT as a Phenomenon and a Paradigm

Since its introduction, MOT has been used in many different studies of visual cognition. Some of this work has focused on MOT as a phenomenon in its own right, exploring its constraints and underlying processes. For example, research has characterized how the ability to track multiple objects is influenced by the number of targets (Oksama and Hyönä 2004), their speeds (Liu et al. 2005), their relative depths (Viswanathan and Mingolla 2002), the reference frame in which they move (Liu et al. 2005), individual differences (Oksama and Hyönä 2004), and various higher-level strategies that may be employed by observers (Yantis 1992). Perhaps the most surprising result to come out of this larger research project is the discovery that MOT seems not to involve much extrapolation of objects' trajectories: with only a few exceptions (Franconeri, Pylyshyn, and Scholl 2006), observers are better at MOT when objects that have disappeared reappear at their last known addresses, rather than where they "should" be had their motions continued (Fencsik et al. 2007; Franconeri, Pylyshyn, and Scholl 2006; Keane and Pylyshyn 2006).

MOT has also been frequently used as a tool with which to study other aspects of visual cognition. This work may not necessarily depend on any details about how MOT does or does not work, but simply employs it to manipulate attention in the study of other topics, such as working memory (Fougnie and Marois 2006; Postle, D'Esposito, and Corkin 2005), task switching (Alvarez et al. 2005), spatial resolution (Intriligator and Cavanagh 2001), occlusion (Flombaum, Scholl, and Pylyshyn 2008; Scholl and Pylyshyn 1999), dual-task interference (Allen et al. 2004, 2006; Fougnie and Marois 2006; Trick, Guindon, and Vallis 2006), or even self-regulation (Oaten and Cheng 2006). More generally, MOT has been used as a tool to study the operation of attention in many different populations, including young children (O'Hearn, Landau, and Hoffman 2005), older adults (Trick, Guindon, and Vallis 2005; Sekuler, McLaughlin, and Yotsumoto 2008), special populations (O'Hearn, Landau, and Hoffman 2005), and visual experts such as radar operators (Allen et al. 2004) and videogame players (Green and Bavelier 2006).

The goal of this chapter is to explore MOT as both a phenomenon and a paradigm, focusing on how it interacts with visual attention.

3 The Relationship between MOT and Attention

A common assumption is that MOT is an illustration of the dynamics of attention; indeed, it is sometimes even referred to as simply "attentive tracking" (e.g., Fougnie and Marois 2006; vanMarle and Scholl 2003) or "multifocal attention" (e.g., Cavanagh and Alvarez 2005). However, the relationship between MOT and attention in Pylyshyn's own work is more subtle.

3.1 Visual Indexing

As noted in the introduction to this chapter, Pylyshyn initially created (discovered?) MOT for a specific theoretical purpose. In order to detect even simple geometrical properties among the elements of a visual scene (e.g., being collinear, or being "inside"), he argues, the visual system must be able to simultaneously reference—or "index"—multiple objects in parallel, and to maintain that referential contact over time. This indexing is even necessary to shift attention to an object, since you can't shift attention *to* anything unless you are already referencing it. Pylyshyn noted that this visual indexing theory (e.g., Pylyshyn 1989, 1994, 2001, 2003, 2007) predicted that something like MOT should be possible, and so he created the paradigm in order to test this prediction.

In Pylyshyn's theory, visual indexes (or "FINSTs," for FINgers of INSTantiation, by analogy to pointing fingers) are independently assigned to various items in the visual field on the basis of bottom-up salience cues, and the indexes serve as a means of access to those items for the higher-level processes that allocate focal attention. In this regard, they function like pointers in a computer data structure: They reference certain items in the visual field (identifying them as distinct objects), without themselves encoding any properties of those objects. Indexes are thought to be assigned to objects in the visual field regardless of their spatial contiguity (in contrast with spotlight models), but with the restriction that the architecture of the visual system provides only a limited number of indexes (roughly four). Furthermore, the indexes are sticky: If an indexed item in the visual field moves, the index moves with it, maintaining the referential connection.

3.2 Indexing and Attention

A key assumption of the indexing theory has been that (at least part of) the assignment and maintenance of indexes—that is, the selection of targets and the actual tracking in MOT—is preattentive, automatic, and data driven. This is a key assumption because it underlies the entire reason for indexing in the first place. This aspect of the proposal serves to link visual processing up with the world, providing an exit from the regress in which various representational systems are explained in terms of other representational systems. If a significant portion of the indexing process is truly data driven, then indexing might serve as a sort of interface between the world and the mind, and could underlie higher-level types of object-based processing. In the words of Fodor (this volume), indexing "is where the intentional gets its grip on the physical; it's where psychology starts to get 'naturalized' "(**XX** [last p.]). But indexing can't serve this function unless it operates at least in part at a lower level than attention.

The strongest form of this assumption would be that MOT is entirely preattentive, but this view is clearly wrong. For example, without other assumptions this view is inconsistent with the basic finding that MOT decays with increased tracking durations, and that it is subject to large individual differences that correlate with other aspects of attention (Oksama and Hyönä 2004). However, this no-attention view is (and always has been) a straw man. The initial presentation of MOT (Pylyshyn and Storm 1988) did explicitly suggest that some of the actual tracking was preattentive: the "stage . . . that maintains the identity of a visual feature as it moves about in the visual field" can be "shown to have more than one independent locus and may thus actually be a 'preattentive' stage" (180). However, even this initial report noted that attention was likely to be involved in MOT in other ways—that indexing "is a preattentive operation, although the selection of some subset of these automatically indexed places for . . . tracking may involve deliberate cognitive intervention" (181).

Later discussions of visual indexing have helped to clarify this view (Pylyshyn 1994, 2001, 2003, 2007). These discussions have maintained the view that the actual tracking is in part an automatic and preattentive function (such that tracking is "primitive" and a part of "early vision"), but they have noted that MOT may nevertheless be effortful and attentionally demanding, since indexes may have to "be periodically refreshed" to prevent decay (Pylyshyn 1994, 369), since the task "requires effort inasmuch as it involves warding off competing events" (Pylyshyn et al. 1994, 266) and since observers must also employ an "error recovery stage" to rescue "lost" objects during motion (Sears and Pylyshyn 2000). In sum, "more is going on in tracking tasks than the mere invocation of an automatic tracking mechanism" (Pylyshyn et al. 1994, 266). This is also true for the initial assignment of the indexes in MOT: while the theory has always maintained—as it must, given its purpose—that indexes can be assigned in an automatic and data-driven manner, this is not exclusive, and indexes can also be assigned deliberately via focused attention (Pylyshyn and Annan 2006). In some ways, of course, this has to be true, since the MOT task is a task, and like all tasks it involves central executive resources involving goal maintenance, response selection, and performance monitoring. This can also be easily demonstrated experimentally, since for example MOT interferes with even very general tasks involving auditory tone-monitoring (e.g., Alvarez et al. 2005).

3.3 What If There's No More to MOT Than Attention? How Could We Tell? And Where Is the Burden of Proof?

In noting the ways in which attention may influence MOT, Pylyshyn has commented that "it is clear that more is going on in MOT experiments than just tracking based on data-driven index maintenance" (Pylyshyn 2001, 149). Here I would like to turn this question around: given that the role of attention in MOT is so salient (even phenomenologically), is it clear that there is any more going on in MOT experiments than the application of attention itself? Put more bluntly, is there reason to think that there is any data-driven index maintenance in MOT? (Note that Pylyshyn has employed compelling conceptual arguments to suggest that there must be some data-driven system of visual demonstrative reference in the mind in order to "get vision off the ground," but of course that doesn't mean that it plays a role in this particular task, despite its provenance.) Here I propose that although MOT may have taught us several important things about attention, there may be nothing to MOT beyond attention.

This is a difficult view to defend, simply because it is not obvious how one could falsify the possibility that data-driven index maintenance is involved at some stage. After all, any example of an attentional effect on MOT can be easily (and perhaps too easily) deflected to some other aspect of the global "task"—for example, to an "error recovery stage" or to response selection—without any preexisting constraints on when and how such stages should and should not operate. Moreover, since MOT is "interruptible"—you can do other things for up to at least several hundred milliseconds while you ignore tracking (Alvarez et al. 2005)—any attentional effects during MOT could also always be argued to reflect additional processing that simply occurred "in between" periods of data-driven tracking. In short, to borrow a phrase from a recent study of individual differences in this task (Oksama and Hyönä 2004), in order to evaluate whether MOT involves anything other than attention, one would need a measure of "pure tracking"—but such a measure has never been developed.

Where this leaves us depends on where one thinks the "burden of proof" lies. Pylyshyn has always been clear on this issue: because the indexing theory is a bold attempt to "ground" cognition in a type of brute demonstrative reference (see especially Pylyshyn 2001, 2007), it is worth taking seriously. The view that there is some "pure tracking," in other words, is "the more interesting hypothesis to pursue, pending evidence to the contrary" (Pylyshyn 2001, 149). However, though it may be true that this is a good reason for "pursuing" the hypothesis, I question whether this is a good reason for (even provisionally) accepting the hypothesis. We already know that attention exists from countless studies, that it can be "split" under several circumstances (e.g., Cassidy, Sheremata, and Somers 2007; Castiello and Umiltà 1992; Driver and Baylis 1989; Kramer and Hahn 1995; McMains and Somers 2004), and that it can move (e.g., Cavanagh 1992; Driver and Baylis 1989; Verstraten, Cavanagh, and Labianca 2000). Meanwhile, the visual indexing view proposes an entirely new mechanism of mind—one without a large body of independent supporting evidence, and without any independent evidence for involvement in MOT. So, I suggest, we should prefer the attentional theory of MOT simply on the grounds of parsimony, without some positive evidence for the involvement of a novel "extra" mechanism. Of course, on this view it may still be important to pursue the possibility that visual indexing exists and is involved in MOT, but we should not start from that position without such evidence.

3.4 Is There Evidence against Indexing in MOT? Tracking Individuals versus Sets

As noted above, it is not clear how the hypothesis that indexing is involved in MOT could be directly tested and potentially refuted. As such, in this section I will argue against the involvement of indexing in MOT (and thus argue indirectly for the view that MOT is realized only by attentional tracking) in a different way, by emphasizing an aspect of MOT that seems inconsistent with the purpose of indexing. And, in what is perhaps an unorthodox move, I will make this argument based on one of Pylyshyn's own recent discoveries about MOT.

One of the key assumptions about MOT since its initial discovery has been that each target object is being tracked as a distinct individual: during tracking one is keeping track of *this* target, *that* target, and *that* target as each moves about the display. Recently, however, Pylyshyn (2004) noted an apparent challenge to this view. This challenge can be readily appreciated by any observer in the following way: During the initial target phase, internally name each of the targets. (If you try this using online MOT movies, you can also simply pause the movie during this phase.) For example, if you must track four of eight objects, think of the four targets as A, B, C, and D. Then, at the end of the tracking interval (when your task would normally be to indicate the four targets), give yourself the additional following task: Identify which is which. What you will find is that this is extremely difficult—and is certainly much more challenging than the basic MOT task. Indeed, when you've accurately tracked the four targets, it can be exceedingly difficult even to identify one of them in this way (e.g., which one is B, or which one started out in the upper right quadrant). Pylyshyn (2004) experimentally confirmed the extreme difficulty of keeping track of "which is which" during MOT, and showed that this difficulty is not due to any general dual-task interference (since there is no such deficit when the "labels" on static objects must be remembered, even through a separate tracking interval with additional objects).

This result is exactly what you would expect if targets are maintained during MOT simply by split foci of object-based attention. Under this view, there is nothing that makes one focus of attention different from another: They simply enhance processing on (and as a result, help us keep track of) each of the targets, as a set, not as individuals. As such, the attentional tracking view provides a ready mechanism for keeping targets separate from distractors, but not for keeping any of the targets distinct from each other.

In contrast, I suggest that this result is potentially a much greater challenge for the visual indexing view than Pylyshyn realizes. The reason is simply that this inability eliminates that one part of MOT that most directly supports the purpose of visual indexing in the first place: the ability to keep referring to an individual over time and motion such that its properties can be probed, or attention can be shifted to it. This is clearly not possible if you never know which target you are indexing: Any reliance on visual indexes as a foundation for attentional shifts, for example, would lead you to frequently shift attention to the wrong target. This problem can perhaps be most easily appreciated by harkening back to the initial analogy of visual indexes with pointers in computer data structures: Such pointers are of no use (or worse) if different pointers can frequently end up swapping their referents! Similarly, this result undercuts the analogy with pointing fingers—the idea that "the access that the finger [or visual index] contact gives makes it inherently possible to track a particular token, that is, to keep referring to what is . . . the same object" (Pylyshyn 1989, 68). The inability to do just this in Pylyshyn's experiments is essentially equivalent to tracking two objects by continually pointing to one with each index finger, but then later having no idea which object you were initially pointing to with your left index finger!

In Pylyshyn's article, the inability to track individuals per se is ultimately explained away by appeal to the idea that during tracking some targets are mistakenly "swapped" with other targets—and that target-target swaps are more frequent than target-distractor swaps. Such data are reported in a final experiment, showing that errors when attempting to track individuals are more likely to be errors of mistakenly "ID-ing" other targets: For example, when asked which object was target B, you'll mistakenly select target C more often than you'll select one of the distractors. I suggest that this interpretation is not convincing, however, for three reasons. First, it does not really help to salvage a link between MOT and indexing, since even under this interpretation the frequent target-target swaps would still frustrate any automatic target maintenance via indexing. Indexing, in other words, would still not be especially *useful* for MOT. Second, note that these experiments do not actually provide data that directly support this view; rather, they are merely consistent with it. For again, these results are exactly what you would expect if target maintenance is due solely to attention maintained on the targets as a set. Under this scenario, what Pylyshyn calls "target-target swaps" are nothing of the sort: There is nothing to swap, because there is nothing distinguishable about individual targets in the first place. In other words, the response that is being interpreted as a target-target swap is really just a guess: Observers know which items are the targets, but they have no idea which is which, and so during forced-choice responses they frequently ID the wrong target.

The third argument against Pylyshyn's interpretation, I suggest, is that it clearly doesn't apply in all of the cases where it would have to apply. Even when there is no special danger of targets being "swapped" during tracking—say, because they never come near each other—you still have essentially no idea which is which! This can be readily appreciated by viewing any MOT display in which two of the targets never approach each other. Here you can readily discern at the end of the motion that they are both targets, but you will have no idea which is which.

On balance, then, I suggest that what Pylyshyn's (2004) experiments show is exactly what they intuitively seem to show: We can keep track of the targets in MOT, but not which one is which. This undercuts any reason to suggest that data-driven index maintenance is playing any role in MOT, though, since the only way to modify the functioning of indexes to match these results would be to strip them of the one property they must have in order to fulfill the purpose for which they are theorized to exist in the first place. But again, all of this seems easily explained—and perhaps even necessarily predicted—by the view that MOT is simply realized by split object-based attention to the MOT targets as a set.

3.5 MOT = Tracking in the Present

One reason that Pylyshyn (2004) thinks that there must still be some bona fide tracking of individuals going on in MOT, despite the results discussed above, is that he thinks this is conceptually necessary. He calls this the *discrete reference principle*, and suggests that "a critical part of determining whether some object is a target is being able to trace its individuality . . . back over time to the start of each trial. . . . [T]he only way to determine that a particular individual object belonged to the target set in the previous instant is by knowing which particular individual in the target set it had been" (804, 805).

This logic seems mistaken. In order to identify an object as a target, you need only know that it was a target an instant ago—and everything that came before that moment can be "flushed" from the system without any cost. This is, in fact, what I think occurs during MOT: We are continually tracking *in the present*, without necessarily storing and using some sort of spatiotemporal trace back to the start of the trial. (This is not to say that such implicit memories are not possible, just that we don't use them during tracking.) During the very first frame of motion, you may indeed have a representation that demonstratively IDs each target, but the very next moment that information is gone, and all you know is that it was *a* target a moment ago, and so it must still be *a* target now.

I think this view—that tracking does not require a spatiotemporal trace back to the start of a trial—can be appreciated empirically as well as logically. One way to highlight this is to explore the ways in which tracking can be interrupted and resumed. Dual-task studies of MOT and visual search, for example, have convincingly shown that observers can switch back and forth between these two tasks in sequence, picking up the tracking from where it left off (Alvarez et al. 2005). This seems mysterious according to the indexing view, however, since presumably the indexes would also be required to help implement the search task: Given that search proceeds via the movement of attention, each shift of attention to a potential search target would by hypothesis have to be preceded by the assignment of an index to that object. But given the limited number of available indexes, this means that the indexes would have to be removed from the MOT targets during these "search interruptions," with no data-driven means to later reassign them to the targets. Nevertheless, tracking is not impaired. Why? Because you don't need to trace each target back to its origin in order to succeed in tracking through interruptions: All you need to know is where the targets are as a set in order to recover them, without any need to know which is which. This is also what happens, I suggest, from moment to moment during MOT even without any extrinsic interruptions: We track only in the present, knowing that the tracked objects are the targets, but without any necessary memory trace of how or where they initially acquired that status. (Indeed, note that two people could even "hand off" the tracking tack back and forth to each other, if the display paused at the right moments: the first person could simply describe to the second person where the four targets are, so that the second person could continue the tracking when the motion restarts. In this case, there would obviously be no possibility of maintaining an explicit tag back to the start of the trial, since the second person might not even have been present at the start of the trial!)

4 What Have We Learned about Attention from MOT?

For the remainder of this chapter I will assume—based on the arguments presented above—that MOT just is attentional tracking of multiple objects. In fact, the only initial difference between MOT and focal attentional tracking—though this is an important difference indeed!—is that attention is (necessarily) split during MOT. Other paradigms have also been used to demonstrate the ability to split attention (e.g., Cassidy, Sheremata, and Somers 2007; Castiello and Umiltà 1992; Driver and Baylis 1989; Kramer and Hahn 1995; McMains and Somers 2004) and for attention to track motion (Cavanagh 1992; Driver and Baylis 1989; Verstraten, Cavanagh, and Labianca 2000), but no paradigm has ever illustrated either of these features of attention more powerfully than MOT—or shown how they can be combined.

The goal of this penultimate section is to emphasize that this "attentional tracking" view is in no way deflationary. MOT may not interact with theories of visual indexing in this view, but it has nevertheless allowed us to make several important discoveries about the nature of attention—including several that would not likely have been possible without MOT. This section briefly reviews five such examples from our laboratory.

4.1 Attention Is (Sometimes Necessarily) Object Based

One key question about any cognitive or perceptual process concerns the units over which it operates. As noted earlier, most traditional theories of attention either assumed or explicitly argued that attention was fundamentally spatial, as in metaphors based on spotlights or zoom lenses (for a review, see Cave and Bichot 1999). Such spatial models inherently ignored the structure of the attended information: The process of selection was based on an extrinsic filter, and as a result you could attend to an object, multiple objects, only parts of objects, or even nothing at all—whatever fell within the spotlight. More recent models of attention, in contrast, have stressed the complex interplay between attention and the structure of the attended information (see Ben-Shahar, Scholl, and Zucker 2007). For example, many studies of object-based attention have demonstrated that the underlying units of attention are often discrete visual objects: Rather than spreading uniformly through a spatially defined region, attention flows more readily through individual objects—or alternately, attention is constrained by their boundaries (for a review, see Scholl 2001a).

The possibility of MOT in the first place demonstrates, as do many other paradigms, that attention can be object based in at least one sense, since the targets and distractors are frequently spatially interleaved. But MOT is still consistent with the possibility that attention is simply split into multiple spatial spotlights. Additional experiments using MOT, however, confirm that in some cases attention is necessarily directed only to discrete objects. For example, observers in one experiment still attempted to track multiple independently and unpredictably moving items, but the nature of these items was altered so that target-distractor pairs were perceived as single objects—with a target at one end and a distractor at the other end (Scholl, Pylyshyn, and Feldman 2001). Such a pair might be drawn as a simple line segment connecting the two points, as in figure 2.2b. Crucially, each end of a pair still moved completely independently. Tracking was greatly impaired in such conditions, despite the use of identical sets of trajectories and target selections: Observers could track individual objects, but not individual ends of uniform objects. This result is readily explained

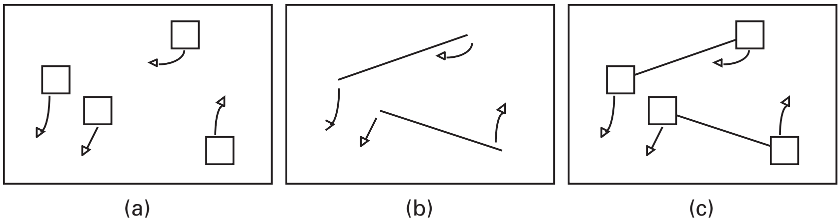


Figure 2.2
Sample "target merging" displays from Scholl et al. 2001. Each display shows four items, each of which always moves independently from all other items. (Actual displays had eight items total.) (a) A control condition, where observers must track punctate objects and perform as in most MOT tasks. (b) Items are merged into pairs, with each pair always consisting of a target and a distractor. Observers are greatly impaired when trying to track one end of each line, though they move through the same trajectories as in (a). (c) When curvature discontinuities are added to the ends of the lines by redrawing the boxes, tracking is better than with the lines alone, but worse than with the boxes alone.

in terms of object-based attention: Selection spreads uniformly throughout the lines, causing observers to lose track of which end was the target.

This demonstration of object-based attention has two advantages over similar demonstrations using paradigms of divided attention and spatial cueing (among others). First, these results demonstrate that object-based attention is in some cases a necessary "mode" of attention that cannot be avoided even when observers have specific task goals to the contrary. In contrast, object-based attention in most other paradigms is heavily influenced by task goals and various other details (e.g., the specific types of cues used and their probabilistic structure). Second, these results indicate that object-based attention can in some cases have a phenomenological component: When trying to track the undifferentiated ends of the lines in this paradigm, you can *feel* object-based attention in action.

Further manipulations of the precise ways in which the targets were connected in such displays indicated how MOT can be used to explore subtler aspects of object-based attention. For example, when observers had to track ends of "dumbbells" as in figure 2.2c, performance was worse than with boxes alone (figure 2.2a), but better than with lines alone (figure 2.2b). This indicates that object-based attention is not an all-or-nothing phenomenon (see also Marino and Scholl 2005), but can be independently affected by multiple cues including connectedness and curvature discontinuities.

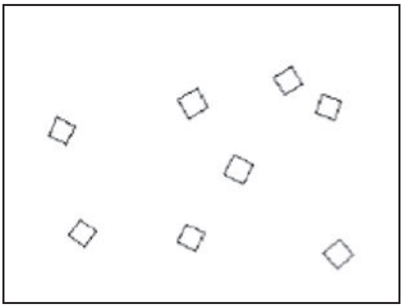
4.2 Dynamic Object-Based Attention Requires Cohesive Objects

Objects are most commonly contrasted with spatial areas (as in section 4.1) or visual surface features such as color and shape (see section 4.5). But another contrast that is common from the study of objects in developmental psychology is that of objects versus nonsolid substances. In the study of infant cognition, for example, one of the most powerful principles of "core knowledge" is that of cohesion: An object must maintain a single bounded contour over time (see, e.g., Spelke 1990, 1994). Indeed, this principle may be uniquely important in that it helps define what counts as an object in the first place. If you want to know what an object is, just "grab some and pull"; the stuff that comes with your hand is the object, and the stuff that doesn't (and thereby fails to maintain a single unified boundary with the stuff that moved with your hand) is not. This has led some theorists to claim that cohesion is perhaps the single most important principle of what it means to be an object (e.g., Bloom 2000; Pinker 1997). And, correspondingly, infants' object-tracking abilities are greatly impaired by simple cohesion violations (Cheries et al. 2008; Huntley-Fenner, Carey, and Solimando 2002).

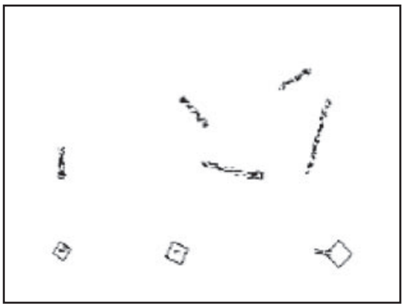
Using MOT, we were able to demonstrate that object-based attention in adult visual cognition is also constrained by cohesion. For example, observers can be asked to track spatially extended objects that move repeatedly in a particular type of noncohesive motion (figure 2.3): Each object began as a small square, but then split into many smaller units and moved in a nonrigid manner—essentially "pouring" from one location to another, as would a nonsolid substance. This manipulation greatly impaired tracking, despite the fact that the "objects" still followed the same trajectories as in typical MOT control conditions (vanMarle and Scholl 2003). We argue that this was due to the fact that each object's location could no longer be characterized by a single point, so that there was no unambiguous location for attention to select on this shrinking and growing extended object.

4.3 Beyond Object-Based Attention: Nonuniform Attention to Uniform Objects

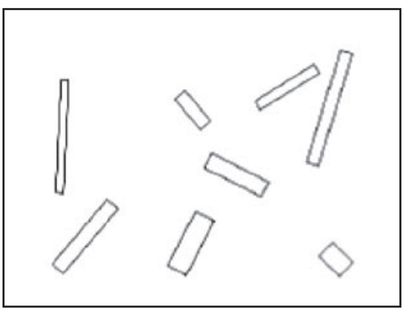
The distinction between object-based and space-based attention need not always be a dichotomy: These views can interact, such that attention can be both spatially oriented and object based, in different ways but at the same time. This is the conclusion drawn from another recent study of MOT that used spatially extended objects. The first study of mandatory object-based attention using MOT, described in section 4.1 (Scholl, Pylyshyn, and Feldman 2001) assumed that attention was spreading equally throughout



(a) : Object condition



(b) : Substance condition



(c) : Morphing condition

Figure 2.3

Sample midtrial screenshots for studies of cohesion and spatial extent in MOT (vanMarle and Scholl 2003). (a) With punctate objects, tracking is accurate. (b) The "objects" move through the same trajectories but split into multiple units during their motion, as if they were liquids being "poured" from one location to another—a manipulation that greatly disrupts tracking. (c) Tracking is also disrupted when each square simply "stretches" its leading edge to its new location (becoming a long thin rectangle), then shrinks its trailing edge, as if it were a caterpillar. Tracking is also greatly disrupted here, perhaps because there is no unambiguous point on the object for attention to select.

the spatially extended lines. But it turns out that this is not the case: Though the lines are uniform, the distribution of attention within them is not.

In these experiments (Alvarez and Scholl 2005), observers had to track three of six long lines that moved haphazardly around a display. The lengths of the lines were randomly increased and decreased as the objects moved, since each of the lines' endpoints moved independently. To allow for an assessment of the distribution of attention within these objects, observers performed a simultaneous probe-detection task in which they were required to press a button whenever they detected the appearance of a probe (a small gray circle). Probes could appear at an object's center or near one of its ends, as depicted in figure 2.4. If attention was uniformly distributed over an object during the MOT task, we might expect that probe-detection rates would be similar for both center and end probes. However, this was not the case. Center probes were detected far more accurately than end probes, suggesting that more attentional resources were concentrated on the centers of the lines than near their ends. This effect was termed *attentional concentration*. Furthermore, the attentional concentration effect was modulated by the lengths of the objects being probed: As a line's length increased, center probes were detected increasingly well and end probes were detected increasingly poorly. In other words, the size of the concentration effect was largest for long lines and smallest for short lines, suggesting that the distribution of attention within an object becomes increasingly concentrated on its center as its length increases. This effect was termed *attentional amplification*, to emphasize that the attentional concentration effect was exaggerated or amplified by increased object length. These effects were both extremely robust (with differences in probe-detection accuracy on the order of 25%–50%), and they cannot be explained by differential patterns of eye fixations (Doran, Hoffman, and Scholl, in press). Both of these effects are illustrated schematically in figure 2.5.

These results begin to show how object-based and space-based attention interact, and they complement the other MOT results described above by narrowing in on the constraints that determine how and whether objects can be attentionally tracked. Both attentional concentration and amplification may reflect the difficulty of tracking spatially extended objects in the first place. Whereas such tracking is impossible for spatially extended objects that grow and shrink at especially fast rates (see the study of cohesion in section 4.2), it is possible when the lines' endpoints simply move independently, as in these studies. Because there is no single explicit

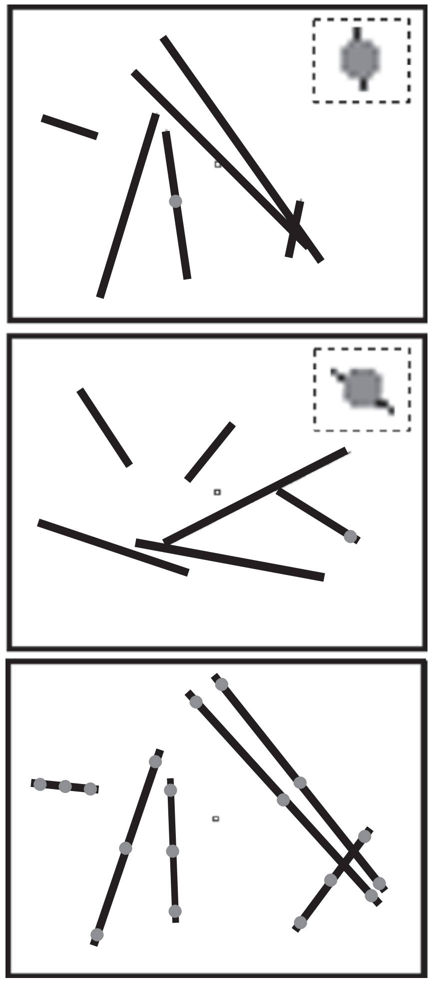


Figure 2.4
The concurrent MOT and probe-detection tasks used to discover the effects of attentional concentration and attentional amplification (Alvarez and Scholl 2005). Observers were required to keep track of three out of six moving lines while concurrently monitoring for the appearance of gray dot probes. (The box near the center of the displays is a fixation marker.) (a) A center probe trial in which a gray dot appears at the center of a line during the tracking task. The inset shows the local contrast of the center probe. (b) An endpoint probe trial in which a gray dot appears near the end of a line during the tracking task. (Note that the local contrast here in the inset is identical to that for center probes.) (c) A single frame of a trial highlighting all possible probe positions within that frame. (Only one probe was presented at a time in the actual experiment.)

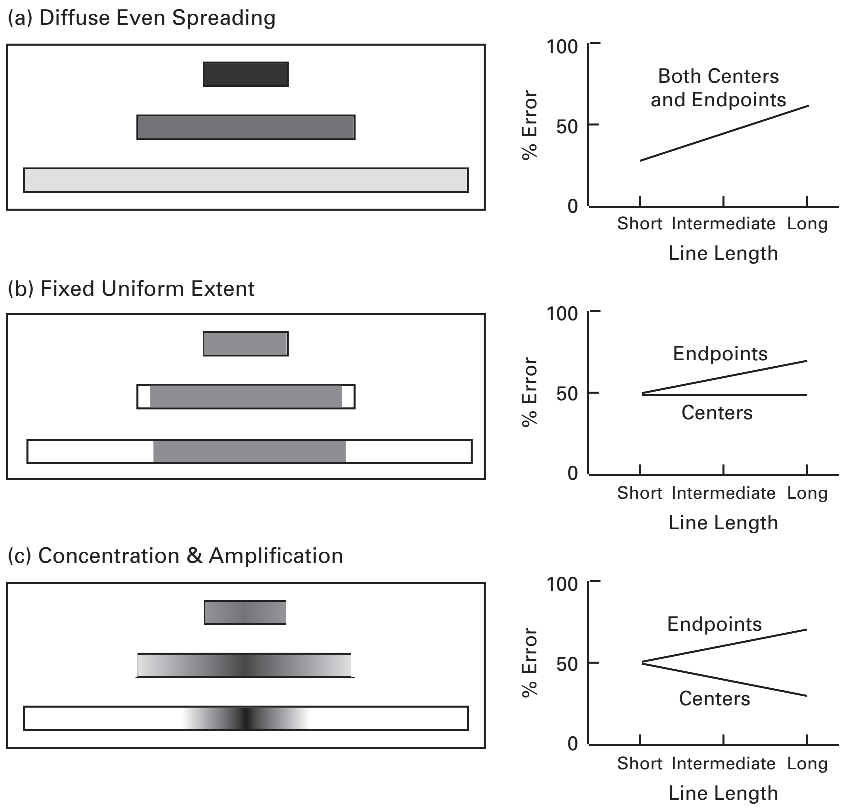


Figure 2.5
Three possible patterns of attentional distribution that could arise from the studies of MOT and probe detection from figure 2.4 (Alvarez and Scholl 2005). Here we depict three possible patterns of attention across the centers and endpoints of both long and short lines. In each case, the color of the line at each point represents the amount of attention (and the likelihood of probe detection), with darker areas indicating more attention, and lighter areas indicating less attention. (a) The performance predicted by a model in which attention always spreads uniformly through entire objects, but becomes more diffuse with increasing spatial extent. (b) The performance predicted by a model in which attention spreads uniformly through as much of a line as is allowed by available capacity. In short lines this yields uniform attention over the whole object, whereas in long lines this yields uniform attention over only a central portion, with little or no attention at the endpoints. (c) A schematic depiction of the actual results, illustrating both concentration and amplification: attention is concentrated at centers compared to endpoints, with centers receiving relatively more attention as line length increases, and endpoints receiving relatively less attention as line length increases.

punctate location for attention to select, a prioritized location may have to be effectively "constructed" via an attentional discontinuity (as in the concentration effect), and the need for such a discontinuity may map onto the degree to which there fails to be a single salient point-location for such objects, which would increase the prevalence of this effect (i.e., attentional amplification) as the lines grow longer.

4.4 Attention Is Influenced by Spatiotemporal Stability

Because MOT is an inherently dynamic paradigm, it allows us to ask questions about attention that would not be possible with paradigms employing only static displays. For example, in one recent study we asked about how attention is influenced by spatiotemporal stability (Alvarez, White, and Scholl, in preparation). In our earlier work described in the preceding section (Alvarez and Scholl 2005), we showed that when tracking spatially extended objects, attention is often concentrated at their centers. The centers of such objects may be important in part because they prove to be the most stable points across various types of motion. To track a person, for example, you would do well to track a point along his torso rather than his hands or feet (which may undergo many spurious local motions). Thus, under conditions of high load, as in MOT, attention might have a tendency to concentrate near the most stable point within an object, as a heuristic to help keep track of it. (This idea may help to explain why the attentional concentration effect exists, but it cannot explain away the effect: It persists even when the subject is comparing probes at endpoints and centers that are matched for velocity; Alvarez and Scholl 2005.)

We recently explored directly whether spatiotemporal stability influences attention by combining the tracking of long lines with probe detection as in our earlier studies, but now using lines that moved in different ways, making some points more stable than others (Alvarez, White, and Scholl in preparation). In these conditions, the "attentional concentration" effect still dominates: Attention is concentrated at lines' centers even when those points are the least stable of all. This was demonstrated by having observers track "bows" as in figure 2.6—long curves whose centers were constantly oscillating in a direction orthogonal to the endpoints' orientation. Probe detection revealed that attention was concentrated at the curves' centers (more so than their endpoints), despite the fact that the centers were always moving faster than the endpoints.

However, an effect of spatiotemporal stability can be observed when stability is not competing with attentional concentration. To demonstrate this, we had observers track "walkers"—long rigid lines where only one

Figure 2.6
Illustration of the "bows" used by Alvarez, White, and Scholl (in preparation). Observers tracked long curves whose centers were constantly oscillating in a direction orthogonal to the endpoints' orientation (as indicated by the arrows, which were not present in the actual displays). Probe detection revealed that attention was concentrated at the curves' centers (compared to near their endpoints), despite the fact that the centers were always moving faster than the endpoints. As described the text, a different stimulus—"walkers"—yielded a different result, wherein there was an advantage in probe detection for more stable positions along tracked objects.

endpoint moves at a time—with the static and moving endpoints frequently swapping. (Static frames of this experiment thus looked just like those in figure 2.4, though now only one endpoint was moving at a time.) Probe detection revealed that attention concentrated at the lines' centers, but also prioritized the lines' stable (unmoving) endpoints over their moving ends. (The same effect obtained for slow vs. fast moving endpoints.) This is, to our knowledge, the first demonstration that spatiotemporal stability influences attention. This phenomenon presumably operates frequently in the real world, but would not be apparent in most experimental paradigms, since (unlike MOT) they are not able to test the distribution of attention on objects that move in such ways over relatively long periods of time.

4.5 Spatiotemporal Priority and Multiple Types of Attention

In section 3, I argued that MOT reflects attentional tracking rather than any special kind of indexing mechanism. However, the way that attention operates during MOT may still be interestingly different than in some other tasks. In particular, there may be different types of attention that support different kinds of visual processing. This was the conclusion drawn from a recent dual-task study that explored the nature of visual memory for natural scenes. Like most aspects of visual processing, the perception of scenes seems quick and effortless, as does the resulting memory for scenes: We can retain accurate memories for thousands of scenes based on only quick presentations (see, e.g., Standing 1973). This may seem to illustrate a type of automaticity, but in such situations observers are typically attending to the scenes that must be encoded, and without attention we often fail to see (much less remember) anything at all (Mack and Rock 1998; Most et al. 2005).

So, does scene memory require attention? This question can be studied via dual-task experiments, exploring the fidelity of both short- and long-terms scene memory when the presentation of the initial scenes occurs while observers are engaged in an attentionally demanding competing task. The results of such studies, however, turn out to depend on the specific types of tasks that are used to engage attention. When attention is engaged by a visual search task during initial scene presentation, for example, the resulting scene memory suffers (Wolfe, Horowitz, and Michod 2007)—and indeed it suffers beyond the baseline impairment produced by combining scene presentation with a generic central executive task such as auditory tone monitoring. When scenes must be encoded during MOT, however, a different picture emerges (no pun intended). In a recent study, we had observers complete a standard MOT task while several scenes were presented (see figure 2.7), but the resulting impairments of scene memory

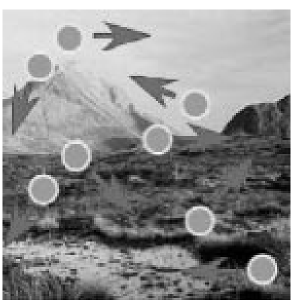


Figure 2.7
A screenshot from an experiment wherein natural scenes were presented in the background of a MOT task (Jungé et al. under review). (Gray arrows indicate motion of the discs, and were not present in the actual displays.) Unlike other competing attention tasks such as visual search, MOT did not greatly impair the resulting scene memory. See the text for details.

did not exceed those produced by a baseline central executive task (Jungé et al., unpublished).

Why would scene memory be especially disrupted by one attention-demanding task (visual search) but not another (MOT)? We think this is because there are different forms of attention. In particular, many studies indicate that distinct attentional processes may be involved in *identification* (i.e., the processing of what an object is, on the basis of surface features) versus *individuation* over time (i.e., determining how and where objects move, on the basis of spatiotemporal information). Perhaps most famously, these sorts of processes seem to be localized in anatomically distinct cortical streams (e.g., Livingstone and Hubel 1988), with the ventral pathway corresponding to identification, and the dorsal pathway corresponding to individuation. In addition, a variety of behavioral evidence supports this distinction. The surface features of objects (e.g., their colors and shapes), while obviously critical for many visual processes including object recognition, seem to be largely discounted by many other processes (for a review, see Flombaum, Scholl, and Santos, in press). For example, surface features play little or no role in determining apparent motion correspondence (Burt and Sperling 1981), identity over time in the tunnel effect (Flombaum et al., 2004; Flombaum and Scholl 2006; Michotte, Thinès, and Crabbé 1964/1991), or object-specific priming (Mitroff and Alvarez 2007).

This distinction can help to explain the scene memory results. In particular, we suggest that the two relevant types of attention can be characterized in terms of the distinction between identification and individuation. Visual search (as employed in Wolfe, Horowitz, and Michod 2007) seems chiefly concerned with identifying objects on the basis of what they look like. In contrast, MOT (as employed in Jungé et al., unpublished) seems principally concerned with keeping a set of objects distinct from others over time on the basis of how and where they move (regardless of what they look like.) Thus, though both search and MOT can be highly attentionally demanding, they may do so via demands on partially independent attentional subsystems. In particular, visual search may interfere dramatically with scene encoding because both processes rely heavily on the same underlying ventral identification-based form of attention. In contrast, MOT fails to interfere with scene encoding more than central executive tasks because MOT relies primarily on a different underlying type of visual attention, one that is dorsal and individuation based. This distinction may also help to explain why MOT and search interfere with each other so little (Alvarez et al. 2005): They may both be highly attention-demanding, yet they may draw on fundamentally different forms of attention. Similarly, this may help to explain why observers are relatively poor at encoding surface features of objects in MOT—including those of tracked targets (see, e.g., Bahrami 2003; Ko and Seiffert 2006; Scholl, Pylyshyn, and Franconeri 1999). In sum, MOT may contrast with most other paradigms used to study attention not only in its requirements for attention to multiple objects and for attention to moving objects, but also in the type of attention it invokes.

5 Conclusions

Research on MOT—particularly as a tool with which to study and manipulate attention—is thriving. Indeed, in the last two decades since Pylyshyn's initial report of this phenomenon, the year with the most publications using MOT was 2008 (this year, as of this writing), and the runner-up was 2006. (A frequently updated bibliography of all work employing MOT can be found online at http://www.yale.edu/perception/MOT-Papers/.) The ideas and results discussed in this chapter suggest two reasons for this. First, the special nature of MOT matches key aspects of real-world visual experience: Whereas many or even most paradigms of attention involve unitary attentional shifts to single objects in static displays, real-world perception—and MOT—involves sustained attention to multiple moving objects. Second, these very features of MOT have allowed us to ask and answer questions about attention that we would not otherwise be able to address.

Acknowledgments

For helpful conversation and/or comments on earlier drafts, I thank George Alvarez, Jon Flombaum, and the members of the Yale Perception and Cognition Laboratory. None of our own work as described in this essay would have been possible without the encouragement and mentorship of Zenon Pylyshyn.

References

Allen, R., P. McGeorge, D. G. Pearson, and A. B. Milne (2004). Attention and expertise in multiple target tracking. *Applied Cognitive Psychology* 18: 337–347.

Allen, R., P. McGeorge, D. G. Pearson, and A. B. Milne (2006). Multiple-target tracking: A role for working memory? *Quarterly Journal of Experimental Psychology* 59: 1101–1116.

Alvarez G. A., T. S. Horowitz, H. C. Arsenio, J. S. Dimase, and J. M. Wolfe (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., and B. J. Scholl (2005). How does attention select and track spatially extended objects? New effects of attentional concentration and amplification. *Journal of Experimental Psychology: General* 134: 461–476.

Alvarez, G. A., A. White, and B. J. Scholl (in preparation). Attention and spatiotemporal stability. Manuscript in preparation.

Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition* 10: 949–963.

Ben-Shahar, O., B. J. Scholl, and S. W. Zucker (2007). Attention, segregation, and textons: Bridging the gap between object-based attention and texton-based segregation. *Vision Research* 47: 845–860.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge, Mass.: MIT Press.

Burt, P., and G. Sperling (1981). Time, distance, and feature trade-offs in visual apparent motion. *Psychological Review* 88: 171–195.

Cassidy, B. S., S. Sheremata, and D. C. Somers (2007). Spatially specific training effects in multiple spotlight attention [Abstract]. *Journal of Vision* 7(9): 700, 700a.

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

Cavanagh, P. (1992). Attention-based motion perception. *Science* 257: 1563–1565.

Cavanagh, P., and G. A. Alvarez (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences* 9: 349–354.

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

Chan, D. T., B. J. Scholl, B. Scassellati, and H. Qian (in preparation). Computational models of heuristic strategies in multiple object tracking. Manuscript in preparation.

Cheries, E. W., S. R. Mitroff, K. Wynn, and B. J. Scholl (2008). Cohesion as a principle of object persistence in infancy. *Developmental Science* 11: 427–432.

Doran, M. M., J. E. Hoffman, and B. J. Scholl (in press). The role of eye fixations in concentration and amplification effects during multiple object tracking. *Visual Cognition*.

Driver, J., and B. Baylis (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance* 15: 448–456.

Fencsik, D. E., S. B. Klieger, and T. S. Horowitz (2007). The role of location and motion information in the tracking and recovery of moving objects. *Perception and Psychophysics* 69: 567–577.

Flombaum, J. I., and B. J. Scholl (2006). A temporal same-object advantage in the tunnel effect: Facilitated change detection for persisting objects. *Journal of Experimental Psychology: Human Perception and Performance* 32(4): 840–853.

Flombaum, J. I., S. M. Kundey, L. R. Santos, and B. J. Scholl (2004). Dynamic object individuation in rhesus macaques: A study of the tunnel effect. *Psychological Science* 15(12): 795–800.

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

Flombaum, J. I., B. J. Scholl, and L. R. Santos (in press). Spatiotemporal priority as a fundamental principle of object persistence. In *The Origins of Object Cognition*, ed. B. Hood and L. Santos. Oxford: Oxford University Press.

Fougnie, D., and R. Marois (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. L., Z. W. Pylyshyn, and B. J. Scholl (2006). Spatiotemporal cues for tracking multiple objects through occlusion. *Visual Cognition* 14: 100–103.

Green, C. S., and D. Bavelier (2006). Enumeration versus object tracking: Insights from video game players. *Cognition* 101: 217–245.

Hulleman, J. (2005). The mathematics of multiple object tracking: From proportions correct to number of objects tracked. *Vision Research* 45: 2298–2309.

Huntley-Fenner, G., S. Carey, and A. Solimando (2002). Objects are individuals but stuff doesn't count: Perceived rigidity and cohesiveness influence infants' representations of small groups of distinct entities. *Cognition* 85: 203–221.

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

Jungé, J. A., J. S. DiMase, B. J. Scholl, M. M. Chun, T. S. Horowitz, and J. M. Wolfe (unpublished). Attentional demands of encoding scenes into memory: Evidence from interference with multiple object tracking.

Keane, B. P., and Z. W. Pylyshyn (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level non-predictive function. *Cognitive Psychology* 52: 346–368.

Ko, P., and A. E. Seiffert (2006). Visual memory for colors of tracked objects [Abstract]. *Journal of Vision* 6(6): 1080, 1080a.

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

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

Livingstone, M. S., and D. H. Hubel (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science* 6: 740–749.

Mack, A., and I. Rock (1998). *Inattentional Blindness*. Cambridge, Mass.: MIT Press.

Marino, A. C., and B. J. Scholl (2005). The role of closure in defining the "objects" of object-based attention. *Perception and Psychophysics* 67: 1140–1149.

McMains, S., and D. Somers (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron* 42: 677–686.

Michotte, A., G. Thinès, and G. Crabbé (1964/1991). Les complements amodaux des structures perceptives. In S*tudia Psychologica*. Louvain: Publications Universitaires. Reprinted and translated as: Michotte, A., G. Thinès, and G. Crabbé (1991). Amodal completion of perceptual structures. In *Michotte's Experimental Phenomenology of Perception*, ed. G. Thines, A. Costall, and G. Butterworth, 140–167. Hillsdale, N.J.: Lawrence Erlbaum.

Mitroff, S. R., and G. A. Alvarez (2007). Space and time, not surface features, underlie object persistence. *Psychonomic Bulletin and Review* 14: 1199–1204.

Most, S. B., B. J. Scholl, E. Clifford, and D. J. Simons (2005). What you see is what you set: Sustained inattentional blindness and the capture of awareness. *Psychological Review* 112: 217–242.

Oaten, M., and K. Cheng (2006). Longitudinal gains in self-regulation from regular physical exercise. *British Journal of Health Psychology* 11: 717–733.

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

Oksama, L., and J. Hyönä (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.

Pashler, H. (1998). *The Psychology of Attention*. Cambridge, Mass.: MIT Press.

Pinker, S. (1997). *How the Mind Works*. New York: Norton.

Postle, B. R., M. D'Esposito, and S. Corkin (2005). Effects of verbal and nonverbal interference on spatial and object visual working memory. *Memory and Cognition* 33: 203–212.

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

Pylyshyn, Z. W. (1994). Some primitive mechanisms of spatial attention. *Cognition* 50: 363–384.

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

Pylyshyn, Z. W. (2003). *Seeing and Visualizing: It's Not What You Think*. Cambridge, Mass.: MIT Press.

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

Pylyshyn, Z. W. (2007). *Things and Places: How the Mind Connects with the World*. Cambridge, Mass.: MIT Press.

Pylyshyn, Z. W., and V. Annan (2006). Dynamics of target selection in multiple object tracking (MOT). *Spatial Vision* 19: 485–504.

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

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

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition* 80(1/2): 1–46.

Scholl, B. J., and Z. W. Pylyshyn (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology* 38: 259–290.

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

Scholl, B. J., Z. W. Pylyshyn, and S. Franconeri (1999). When are spatiotemporal and featural properties encoded as a result of attentional allocation? [Abstract]. *Investigative Ophthalmology and Visual Science* 40: S797.

Sekuler, R., C. McLaughlin, and Y. Yotsumoto (2008). Age-related changes in attentional tracking of multiple moving objects. *Perception* 37: 867–876.

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

Spelke, E. (1990). Principles of object perception. *Cognitive Science* 14: 29–56.

Spelke, E. (1994). Initial knowledge: Six suggestions. *Cognition* 50: 431–445.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology* 25: 207–222.

Trick, L. M., J. Guindon, and L. Vallis (2006). Sequential tapping interferes selectively with multiple-object tracking: Do finger-tapping and tracking share a common resource? *Quarterly Journal of Experimental Psychology* 59: 1188–1195.

Trick, L., T. Perl, and N. Sethi (2005). Age-related differences in multiple object tracking. *Journal of Gerontology* 60B: P102–P105.

vanMarle, K., and B. J. Scholl (2003). Attentive tracking of objects vs. substances. *Psychological Science* 14(5): 498–504.

Verstraten, F., P. Cavanagh, and A. T. Labianca (2000). Limits of attentive tracking reveal temporal properties of attention. *Vision Research* 40: 3651–3664.

Viswanathan, L., and E. Mingolla (2002). Dynamics of attention in depth: Evidence from multi-element tracking. *Perception* 31: 1415–1437.

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

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology* 24: 295–340.