Research Plan

(a) Specific Aims

This research adapts an original method we have developed over the years, called Multiple Object Tracking (MOT), to investigate the nature of visual attention. This technique, and other closely related methods, allows one to study the automatic and arguably preattentive mechanisms involved in assigning and maintaining selective attention. In MOT we briefly indicate (usually by flashing) which 4 or 5 out of 8 identical items are to be treated as "targets." The subject then tracks these target objects over a 15 second trial and selects them at the end of the trial using a computer mouse. In earlier work we showed that people can easily track of 4 or 5 randomly moving target objects in this way. The method is simple and robust and in the past 4 years has allowed us to uncover many counter-intuitive findings that challenge traditional views of the nature of visual attention and of the relation between visual representations and the objects in the world that they represent. It also provides support for a general theory of a preconceptual mechanism in vision, called the visual indexing (or FINST) hypothesis which has been applied both to understand how attention is distributed to objects in the world and also to help clarify some long-standing puzzles concerning how vision connects with individual objects in the world.

The current proposal is to extend this work, using the MOT technique as well as new methods developed to study object-based attention, to explore such questions as how people manage to keep track of 4 or 5 objects under the conditions of MOT and why they cannot track more than this number. The answer to the latter may help shed light on one of the crucial bottlenecks in visual information processing, as well as help us to understand how early vision and cognition interact to allow us to "situate" our perception in the world. Vision must provide not only a conceptualization of the visual world, but must also anchor our perception to allocentric 3D space. It is only by achieving this special anchoring that we are able to compute a stable allocentric representation of 3D space despite ever-changing stimulus information, and it is also how we manage to coordinate our perceptions with our motor system in order to act upon perceived elements in space. Visual Index Theory, which serves as a motivation for many of our studies, is presented as a provisional account for how part of this problem may be solved (a topic discussed at length in Pylyshyn, 2000; Pylyshyn, 2001b, in press-b).

(b) Background and significance

Although this research is primarily motivated by the basic research goal of understanding visual attention and its component processes, the fact that it focuses on the problem of selecting and keeping track of multiple visual elements at the same time has wider implications. To that end we propose to continue to address questions such as the following:

- How (by what mechanism) can people keep track of some subset of objects that are all identical and moving randomly (as in MOT)? And why can they only keep track of 4 or 5 under most conditions?
- How much spatial flexibility does attention distribution have: can it be directed at unoccupied places in the visual field or only to occupied places? Can it be directed to locations or to properties, or only to individual objects? Can its scope be widened (as in the zoom lens view) or split it into separate foci or attentional beams? Can it be spread and molded to fit a large but unitary shape?
- Can some objects in the visual field be selected in a punctate manner while other interspersed objects are ignored? If so, are the ignored objects excluded from all processing – for example are they inhibited?
- Can what is called "focal attention" be decomposed into separate components, specifically individuation, selection (or indexing), access, and encoding (or recognition)?
- Might visual indexes (the core of our Visual Index, or FINST theory) be responsible for our seeing a stable world despite a constantly changing retinal image?

The Multiple Object Tracking paradigm has been our main methodology because it has allowed us to study certain involuntary and unconscious properties of split attention. Understanding the visual mechanism that makes it possible to select and keep track of moving objects is of considerable theoretical and practical importance. The question of what constitutes the units upon which attention is directed is one of central questions in visual cognition today, and the proposal that vision is directed towards objects, as opposed to locations or properties has captures the interest of the vision science community.

The method we call Multiple Object Tracking or MOT, constitutes a pure case of object-based attention, because it demonstrates that visual attention is directed primarily, if not exclusively, to *objects* as opposed to *places*. In addition, the skill involved in MOT is a component of many skilled performances involving coordinating actions in a moving environment, including team sports, driving, air traffic control, and any form of locomotion through an environment containing obstacles. Thus a better understanding of the nature and limits of MOT is an important research goal. There have also been a number of attempts to understand how the brain accomplishes this task (Culham et al., 1998);Robertson, 1997 #1023] since it seems to be a very general principle by which perceptual information is organized. From a theoretical perspective the MOT paradigm is extremely rich, offering many unanticipated findings and suggesting a mechanism (the Visual Index or FINST) by which people can pick out individual visual objects prior to encoding any of their properties. This brings the study of visual attention into contact with a number of lines of research in both human vision and computational vision, where the active or embedded nature of vision is of concern (Ballard, Hayhoe, Pook, & Rao, 1997; Pylyshyn, 2000; Suchman, 1987; VanLehn & Ball, 1991). It also brings it into contact with some philosophical issues, such as whether reference can be made directly to an individual entity without first encoding the unique properties that distinguish that particular individual – in other words can be have deictic or indexical reference in vision (Glenberg & Robertson, 2000; Pylyshyn, 2001b; Slezak, 1999), and what role diagrams may play in reasoning (Allwein & Barwise, 1996; Pylyshyn, in press-b).

(c) Progress Report

1. Summary of progress in the period 01/01/00 – 01/01/03

1.1. Some findings with MOT

The MOT task has become a method of choice for studying both high attention-load processing and for studying certain automatic attentional processes (there were 11 papers or posters at the Vision Sciences 2002 meeting that reported results using MOT).

The MOT paradigm was introduced in (Pylyshyn & Storm, 1988), where it was shown that people could track up to 4 or 5 visual targets (small squares) that moved randomly among an equal number of identical moving nontargets, and that they could do so under conditions that prevented them from using a serial strategy of scanning attention to each element in sequence and updating its stored location. In the past 4 years members of our lab examined a large number of properties of visual attention using MOT. For example, we showed that moving targets could be successfully tracked in MOT even when they disappeared briefly but totally behind an occluding surface, as long as their mode of disappearance provided local cues that the target's disappearance was due to its moving behind an occluding surface (i.e., with deletion occurring at a fixed leading edge and accretion occurring at a fixed trailing edge of the moving object). When the disappearance and reappearance did not exhibit these cues (even when they had the same temporal pattern of decay and growth of luminance) performance was attenuated (Scholl & Pylyshyn, 1999). We also showed that certain clearly visible property clusters could not be tracked if they were not perceived as an object, but rather were seen as part of another object (Scholl, Pylyshyn, & Feldman, 2001), More recently (vanMarle & Scholl, submitted) showed that observers could also not track objects that moved from place to place by a motion indicating it was a substance rather than a rigid objects – e.g., by pouring or generally not occupying a well-defined position in the course of its motion. We also showed that changes in the shape or color of tracked objects could not be reported at better than chance level immediately after they occurred (Scholl, Pylyshyn, & Franconeri, 1999; Scholl, Pylyshyn, & Franconeri, submitted).

More recently we showed that even when target objects were successfully tracked, their individual identities (provided by an initial label associated with each target) were poorly retained, despite the fact that keeping track of which objects are targets *requires* that individual object identities be traced to the start of the trial when the targets were identified as such. Further analysis showed that this puzzling result may be a consequence of targets being more readily confused with other targets than with nontargets (Pylyshyn, submitted). More recent studies suggest that this, in turn, may be due to the inhibition of nontargets during tracking (Pylyshyn & Leonard, Submitted). Inhibition of nonattended objects has been reported by many researchers (Maljkovic & Nakayama, 1994; Mounts, 2000; Theeuwes & Godijn, 2002). Recently (Watson & Humphreys, 1997) introduced a probe-dot paradigm for demonstrating such inhibition at particular locations (but see the differing views of Donk & Theeuwes, 2001; Efron & Yund, 1999; Tipper, 2001) and we have adopted this probe-dot technique in our experiments with MOT.

In addition to these studies we carried out a large number of experiments to try to ascertain how observers are able to keep track of 4 or 5 randomly-moving targets in a field of identical distractors. In the original paper in which we introduced the paradigm, we showed that a strategy of serially visiting each target and encoding its location on a stored list, then sequentially visiting each location on the list, finding the closest object and updating its encoded location, and so on, cannot account for the performance that observers routinely achieve (Pylyshyn & Storm, 1988), which led us to hypothesize the existence of resource-limited mechanism called a visual index or FINST that would automatically stay attached to individual objects and would allow it to be accessed (i.e., examined further) the way a pointer in a data structure does in a computer (Pylyshyn, 1989). Many different experiments were performed to explore this hypothesis, including not only those using MOT but also studies of visual search through a subset if items [where the subset was selected by onsets (Burkell & Pylyshyn, 1997) or through MOT (Cohen & Pylyshyn, 2002)], subitizing (Trick & Pylyshyn, 1994b), and attention-based line motion illusion [the results are summarized in (Pylyshyn, 1994, 1998) and a way of looking at this process which views indexing as deictic reference is discussed in (Pylyshyn, 2000, 2001a, 2001b, in press-b)]. In each case the results showed that people could automatically select a subset of items, marked by sudden onset or other "popout" property, and then serially process them in a variety of ways, including subitizing them or tracking them. Moreover, the fact that the objects were moving made no difference so long as they were being tracked, as we showed by demonstrating that the kind of subset-selection that (Burkell & Pylyshyn, 1997) found also appeared for the subset of items that were being tracked in MOT (Cohen & Pylyshyn, 2002) (i.e., the selected items could be searched as though the nontargets were not there, so that the fact that the superset represented a conjunction search did not slow down the subset search).

Although we have used many techniques to study attention and preattention, the MOT paradigm has turned out to be a particularly fruitful way to study object-based attention, deictic strategies of attention, and visual-motor coordination (as developed, for example, by Ballard et al., 1997), as well as of properties of divided attention and attentional resolution (Intriligator & Cavanagh, 2001). We have, over the years, demonstrated a variety of properties of attention using MOT. For example, detection of change is better when the change occurs on a target that is being tracked than on a nontarget, and this superiority appears to be punctate – it does not spread to locations between targets (Sears & Pylyshyn, 2000). A systematic analysis of the factors associated with tracking errors (e.g., speed of motion, number of nontargets, density of nontargets, diameter of objects, area of display, duration of trial, minimum distance between targets and nontargets) shows that errors are primarily associated with inter-object distance, with only a secondary effect of speed and trial duration (Blaser, Dennis & Pylyshyn, in preparation). An examination of the minimum and time-weighted mean distance between the target-nontarget pairs in those trials where a nontarget was erroneously reported as a target, showed a significant relation between errors and distance (Pylyshyn, submitted). This is in agreement with a suggestion by (Intriligator & Cavanagh, 2001), who showed that when the visual angle of the display is reduced (by moving the display further away) MOT performance suffers.

We have also examined the question of which information-processing resources are being drawn on in MOT. While it is widely recognized that MOT is a highly attention-demanding task, it is not clear what resources or stages are being taxed, whether it is the tracking itself, as opposed to other aspects of the MOT task (such as making a response, or even ignoring other events in the experiment room). In examining this question we have confirmed that tracking performance is somewhat (though not greatly) reduced when a secondary color-change monitoring task is carried out in parallel (Leonard, Pylyshyn, Dennis, & Cohen, 2002). However more recent studies have shown that the monitoring task does not produce any decrement in performance if a response is not required during the tracking trial, but only at the end of each trial (Leonard & Pylyshyn, submitted). This suggests that MOT may not draw on the same attentional resources as required for monitoring and recall. In fact (Alvarez, Horowitz, & Wolfe, 2000) have shown that tracking does not draw on the same resources as visual search, which hints at the interesting possibility that the mechanism of tracking is itself preattentive (as initially proposed by the FINST theory).

Another question that arises with regard to the nature of MOT includes the question of whether tracking involves predicting the trajectory of target objects (or the most likely next time and location of targets), or whether tracking is a "blind" process that automatically keeps track of the individual target to which the index is bound. We have carried out a number of studies of this question and are currently collaborating with my former student, Brian Scholl (now at Yale), on other related studies. So far we have found no evidence suggesting that tracking involves prediction. For example we have shown that if objects disappear for as long as 450ms, tracking is unimpaired when the objects reappear exactly where they had disappeared and significantly impaired when they reappear where they would have been had they followed their trajectory (in these studies the trajectories were kept linear, except for "bouncing" off the edge of the screen, and the speeds were constant, though the speed and initial direction was randomly chosen for each object). Performance when objects reappeared exactly where they had disappeared showed no decrement over the baseline condition where they did not disappear at all. Moreover, performance when the objects reappeared at their extrapolated locations was worse than when the entire display was rigidly translated by the same distance (Keane & Pylyshyn, submitted). In another pilot study we showed that when targets disappeared and then reappeared, performance was not affected unless they reappeared more than 4 diameters away (around 6° of visual angle), and was not affected at all if a nontarget disappeared, whether or not it ever reappeared (Kelly, 2000). Maintaining predicted time seems to be even less helpful in aiding tracking. In a situation where targets go behind an occluder, Scholl (personal communication) found that the best performance always occurred when they reappeared with the shortest delay, as opposed to the predicted delay based on their speed and the width of the occluder. Thus it appears that tracking does not benefit from any prediction of where and when targets will reappear after they have disappeared.

We have also investigated the process by which objects are selected for tracking – this is the process that we refer to as indexing (or "index grabbing") and that visual indexing theory claims is primarily data driven. We showed that objects *can* indeed be voluntarily indexed, but this only occurs when the targets can be sequentially visited (as suggested by Pylyshyn, 2001b). For example, we showed that items that are automatically indexed by being flashed on and off, can be selected for tracking in less than 300 ms. But when the task is to track the non-flashed items, at least 900 ms are required before they can be selected (Annan & Pylyshyn, 2002), presumably because, unlike the flashed objects, nonflashed ones had to be visited serially in order to be indexed. We have recently shown that the longer time it takes to select a nonflashed object is unlikely to be due to the extra time it takes to disengage attention from the automatically-indexed objects, since the same pattern is found when no targets are first selected automatically, e.g., when objects of one color or shape rather than another are selected (Annan & Pylyshyn, submitted).

1.2. Visual Indexes and the stability of perceived allocentric space.

One of the perennial puzzles in vision science is how we can perceive a stable world (with which we interact fluently) in the face of several saccadic eye movements each second. We have begun a complex series of studies in which we asked whether visual indexes might play a role in spatial stability by maintaining pointers to spatially fixed objects during eye movements. Since the locations of no more than about 4 objects are retained over a saccadic movement (Irwin, 1993), visual indexes might be able to solve this problem if it could be shown that they stay attached to 4 salient objects in a scene during a saccade, or to put it another way, if visual indexes could solve the correspondence problem for 4 objects between fixations. To examine this question, I collaborated with a post-doctoral fellow visiting my laboratory, Chrisopher Currie, to carry out three separate experiments at George McConkie's laboratory at the University of Illinois (where Currie received his PhD). The experiments, shown in Figure 1-1) were similar to the subset selection study of (Burkell & Pylyshyn, 1997). A set of "placeholder" X's (randomly arranged) was presented for 1 s. Then an additional 3, 4 or 5 "late onset" X's appeared for 300 ms (always making 8 X's in total). At this point a saccade was induced (by one of several different methods – see below). Once the monitored saccade was completed, all the X's changed into search items (by dropping one of their bars and changing color so as to provide a set of search items consisting of left- or right-oblique red or green bars). The observers' task was to search for a given target only among the elements marked by the late onset markers. (Burkell & Pylyshyn, 1997) showed that without eye movements, subjects could pick out the late onset subset and essentially treat the corresponding subset elements as if they were the only search items in view (so that the superset could be ignored). Since the superset was always a conjunction search set, whereas the subset could be either a feature or a conjunction search set, ignoring the superset conferred a search-time advantage, particularly when the subset was a feature set. The present question is whether a subset advantage could be observed despite the fact that the eye had moved between the time when the subset was identified (by the late onset cues) and when the search could begin (after the saccade, when Xs changed to colored search items).

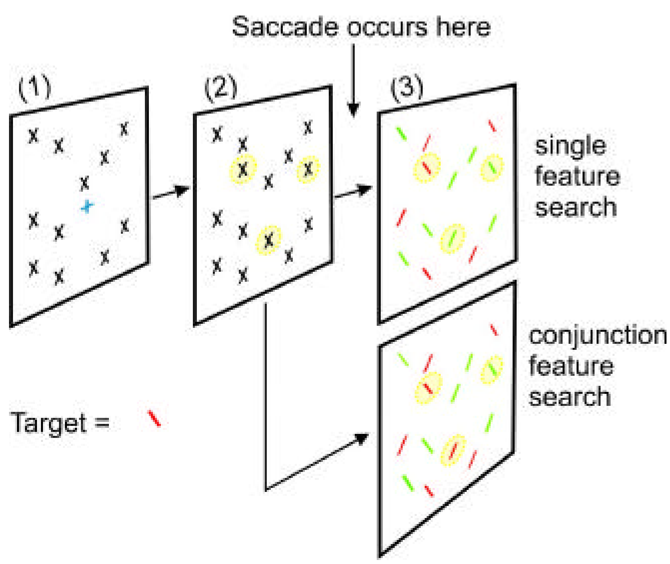


Figure 1-1. The figure above shows the setup used in the saccade study, as adapted from (Burkell & Pylyshyn, 1997). In (1) SEVERAL placeholders are presented for one second, then in (2) an additional 3-4 placeholders appear (shown in yellow for expository purposes). Immediately after, subjects make a saccade from the fixation cross to any of the late-onset objects. After completion of the saccade is detected, all placeholders turn into search items by dropping one bar and changing color. The search, within the late-onset subset, can be either a feature or a conjunction search. We found that Ss are able to confine their search to only the subset.

We tried three methods of inducing a saccade. (a) The first one required subjects to look at a small figure located immediately to the left or the right of the display shown as (1). Then after the late onsets came on at (2) in the above figure, they were to saccade to the fixation point. (b)The second method provided two fixation points in the display. Subjects looked at the first fixation point in (1) and after the late onsets came on, were required to saccade to the second fixation point (the first fixation cross disappeared to speed up the launch of the saccade). (c) The third method was described above: It required subjects to look at the first fixation cross and were instructed that when the late onsets came on, they were to move their eyes to any of the late onset objects. The data for these three experiments are still being analyzed and interpreted, but the following preliminary results can be summarized. We found that subset superiority occurred in could only clearly be demonstrated in (c) when observers were free to decide where to make the eye movement (with the general instruction to glance "in the direction of the late onset X's"). Method (a) did produce suggestive evidence for the subset selection, but the variability was very high and the effect did not reach significance. Method (b), using two fixation points, yielded no evidence of a subset carryover with saccades. These results suggest that when a specific search-irrelevant fixation target is provided, indexing is disrupted. This in turn provides some support for the saccade-target hypothesis (Currie, McConkie, & Carlson-Radvansky, 2000) since it appears that the saccade target has a special status and cannot be ignored in the subsequent task. It also provides support for the visual indexing mechanism as a basis for solving the object-correspondence problem across saccades, since when subjects are allowed to move their eyes to the task-relevant objects, up to 4 of them can be successfully selected and retained across the saccade (Currie, 2002). This is similar to the conclusion reached by (Irwin, 1993) using a quite different method.

1.3. Generalization of the notion of object-based attention

The MOT paradigm suggested a way to generalize the study of object-based attention. As it is generally used, MOT involves tracking objects moving through physical space (both 2D and 3D space have been used Blaser, Pylyshyn, & Domini, 1999). In these cases it is hard to distinguish whether it is a particular individual object that is being tracked or whatever occupies that particular sequence of locations – in other words, whether one is using location as a basis for tracking the objects. We conducted a series of studies that explicitly investigated this puzzle using "objects" that were not defined in terms of their location in space, but in terms of their values along nonspatial property dimensions. We used two "objects" consisting of Gabor patches that varied continuously in spatial frequency, color saturation, and orientation (shown in the figure in section 3.1 of the Design and Methods part of this proposal). When these patches were presented in rapid alternation (117 frames/sec), observers saw two overlaid transparent patches that varied smoothly and independently along the three dimensions. Observers were asked to keep track of the two changing objects and report which one was descended from a particular specified initial object (which observers could easily do). They were also asked the direction of any sudden jumps in one or more of their properties. In one of the conditions, the pair of properties that underwent a sudden change belonged to one "object," and in the other they belonged to two different "objects". These judgments showed the classical single-object superiority effect – a pair of judgments concerning one object was made faster than a pair of judgments that concerned one property of each of two objects.

The importance of this finding has been recognized, not only by those of us interested in tracking, but also be neuroscientists concerned with how information is packaged for storage in the brain (As Braun, 2000recognized in his accompanying Nature news article).

1.4. Summary

These studies provide evidence concerning the nature and role of Visual Indexes in vision and also evidence concerning which aspects of attention might be carried out in parallel and which required serial application. The visual indexing mechanism also has important implications for understanding how we represent space – and indeed how the spatial properties of mental images may be derived from the spatial properties of sensed space providing spatially fixed objects can be indexed (Pylyshyn, 2002a; Pylyshyn, in press-a; Pylyshyn, in press-b).

Note:
VSS abstracts for 2002 can be viewed at: http://ruccs.rutgers.edu/faculty/vss-abstracts-all.htm
VSS Abstracts for 2003 can be viewed at: http://ruccs.rutgers.edu/faculty/vss-abstracts-2003.htm

Most papers sited here can be viewed from pointers at: http://ruccs.rutgers.edu/faculty/pylyshyn.html

(d) Research Design and Methods

Studies using the multiple object tracking task have proven to be a rich source of evidence concerning visual attention. In the next phase we will continue using this methodology, and extend the studies to using related methods aimed at understanding the attentional selection process (or what we refer to as index assignment). We will look more closely at the question of why we cannot track more than 5 objects in MOT to determine whether it is an intrinsic limitation of our visual/attentional architecture or a result of other factors in MOT. We will also further develop the notion of an "abstract" object by studying objects defined across modalities.

1 Factors affecting the number of targets that can be tracked

1.1. Object spacing and "attentional resolution"

The evidence we have so far suggests that objects are confused with one another primarily when they come close to one another, although this confusion is moderated in the case of target-nontarget pairs by what appears to be the inhibition of nontargets. Our earlier evidence suggests that when objects are kept apart (though the use of a "barrier" around each object that deflects other objects, or by using an simulated inverse square "repulsion force" to keep them apart), tracking performance improves somewhat. The degree of improvement has not been systematically studied and we intend to do so in the present series *Studies 1.1.* The surprising finding, on which we have concentrated so far, is that if we allow objects to move freely and occlude one another, tracking performance remains high, providing there are local occlusion cues (T-junctions) as to which object is in front when a pair cross over one another. But we also know that independent of visual acuity, objects too close together cannot be individuated and attended as separate objects. (Intriligator & Cavanagh, 2001; Intriligator, 1997) has mapped out the minimum "attention resolution" distance as a function of retinal locus (resolution decreases approximately linearly with eccentricity). The resulting attentional resolution map is estimated to allow for around 60 possible attentionally distinguishable locations in the visual field. In *Studies 1.1* we will carry out MOT studies in which we vary the size of the "virtual barrier" that defines a circular exclusion zone around objects. Several types of spacing manipulations will be studied:

(a) Vary the diameter of the exclusion circle from zero to three times the object diameter, or about 4º of visual angle (according to Intriligator & Cavanagh, 2001, this should allow pairs of targets to be attentionally resolved for eccentricities of up to about 17º radially, and somewhat more tangentially, and higher for the upper visual field than the lower).

(b) Vary the diameter of the exclusion circle only around the targets. Since the only property of the nontargets that appears to make a difference to tracking is their proximity to targets, varying the size of the exclusion circle around targets will simply keep the targets away from nontargets which would make this condition essentially comparable to (a).

(c) Adjust the diameter of the exclusion zone as a function of eccentricity of objects, so that objects are always spaced further away from one another than the minimum attention resolution reported by Intriligator. We will repeat previous MOT studies, maintaining roughly the minimal spacing specified in the Intriligator map, while increasing the number of targets (adding an equal number of nontargets). At some number of targets either the tracking performance will drop below criterion (e.g. 70%) or the spacing constraint will result in clearly aberrant movement, since enforced minimal spacing constrains the quasi-random choice of speed and direction of motion.

(d) Repeat the above study varying the maximum visual angle of the display up to the far periphery. This will get us into the studies we propose in section 3.3, where the properties of the far periphery will be directly investigated with respect to attention resolution. Although peripheral vision has been studied, we now know that attention does not operate the same way as either pattern of motion acuity. Moreover, there is reason to expect that tracking in the periphery may be surprisingly high (see 3.3).

<u>Instrumentation Note 1.1</u>: In order to allow increased spacing between objects without the projective distortion produced by a large screen, we propose to purchase an *<u>Elumens VisonStation</u>*® (see budget) which allows geometrically corrected projection of a display onto a hemisphere in which equal pixel separation corresponds to equal visual angles and the display is said to be capable of projecting to a visual angle of up to 180º x 135º. We will also need to use this instrument (shown here as seen from the back) when we increase the visual angle of the tracking display in connection with the peripheral tracking studies (*Studies 3.3)* and the sound-vision objects in proposed *Studies 3.2*.

1.2. Visual discriminability of neighboring pairs

In a recently reported pilot study (Dennis & Pylyshyn, 2002) we showed that notwithstanding the poor recall of object colors and shapes in MOT (reported in Scholl et al., submitted), when each object had a distinct color, targets were easier to track. This conclusion was reached by studying the case where every object was a different color from every other object at each instant in time, yet did not remain a fixed color, thus preventing the use of the memorized color to accomplish tracking. Eight colors equidistant on a color circle were selected and applied to each object. In the course of the trial, the selected points on the color circle were rotated rigidly so as to maintain a distinct set of colors which nonetheless changed continuously. When we compared the tracking performance using this asynchronously changing color display with that in which the objects' colors changed in exactly the same manner but were synchronized, so that at every instant the color of all objects was the same, we found the asynchronous color change case to yield superior tracking performance. (We attempted to use object diameter change in the same way, but discontinued the study when we discovered that the apparent speed of movement appears to change with the size of the object – an illusion we intend to study more closely). We intend to pursue this pairwise discriminability effect using different properties and different rates of change because it has implications for understanding the reason for the 4-5 object limitation in MOT, and especially for the "error recovery" explanation for the effects of different factors on tracking performance (described in the original research proposal, and in Pylyshyn, 1989). We will continue the work we started in our pilot studies and will extend it to compare tracking performance when objects remain a fixed shape and color, with tracking performance when objects vary (morph) at different rates, with or without property discontinuities, and either synchronously, asynchronously, or randomly. Because we generate trajectories off-line, we will also be able to keep objects the same color/shape until just before they come within a critical distance from one another, to more directly determine whether it is pairwise discriminability that is responsible for the improved performance we reported in (Dennis & Pylyshyn, 2002). These experiments will form *Studies 1.2*.

1.3. Predictability of object motion: Is tracking predictive?

(Yantis, 1992) showed that when the motion of objects was constrained (in his case the motion of targets was constrained so that targets always formed a convex hull) tracking performance improved. An interesting question is whether *any* increased predictability improves performance, or only particular kinds of constrained motion. In early modeling of MOT (Pylyshyn & Eagleson, 1994) we incorporated a predictive component to the tracking algorithm, but made sure it was based on information local to each target (e.g., we suggested a Kahlman linear adaptive filter, which statistically predicts the future movement of each target by estimating statistical parameters from immediately past motion). But as mentioned in the progress report, we have recently found evidence that tracking is relatively blind and does not reply on prediction of where the objects will move next. For example, while tracking objects that disappear behind an occluder it appears that people are not predisposed to favor particular times and locations of emergence of the occluded target (within limits – observers do notice if the target reappears 4 diameters away) and indeed, they favor overly short reappearance times over realistic extrapolations (despite the fact that people are very good at estimating time-to-contact). We also showed recently that when all objects disappear for up to 450 ms, performance is degraded when they reappear where they would be predicted to reappear (i.e., extrapolating their trajectories to the other side of the visible occluding surface) but performance is <u>not</u> significantly worse if they reappear exactly where they were when they disappeared.

The question of whether there is any predictive component to the tracking algorithm is of considerable theoretical interest. It bears on the question of the sort of mechanism or neural process that might be involved in tracking. Because so much about MOT is counterintuitive we cannot anticipate how the process might work in the absence of experimental evidence (we have been frequently wrong before). Consequently we will examine various issues related to what the tracker does when objects disappear and later reappear. We will examine one individual target in each trial and determine which particular object is selected as its continuation under various manipulations ("right" in this case means "consistent with prediction based on a simple extrapolation"). The cases are illustrated in the corresponding figure. [This project was started while Steve Franconeri and Brian Scholl were students in my lab, so it will be continued in collaboration with them.]

(a) The target goes behind one occluder and emerges on the other side but translated some distance perpendicular to the disoccluding edge.

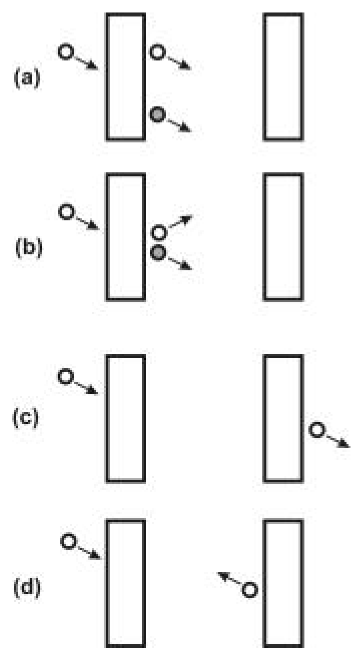
(b) With the same display the target goes behind the occluder and emerges in the expected place, but in an altered direction.

(c) The target goes behind the occluder but emerges at the appropriate time from the far side of the second occluder, going in the right direction and at the right vertical displacement.

(d) The target goes behind the occluder and emerges from the wrong edge of the second occluder (going in the wrong *x* direction).

(e) Targets will have different randomly assigned shapes and colors [as in (a) and (b), where the disks shown in gray are in a different color from the white ones]. The critical target will go behind the occluder and two targets will emerge, equally displaced from the correct location. One will be the same color/shape as the disappearing target, and the other will be different.

(f) In only one trial (somewhere near the end of the experiment), a target will go behind an occluder and will not emerge at all. Subjects (who are told that the software is being debugged – as it always is!) will be asked in the debriefing session whether anything unusual occurred.



1.4. Inhibition of nontargets and the analysis of the role that attention/inhibition play in MOT
We showed that target-target pairs tend to be confused more often than target-nontarget pairs and that this is likely due to the inhibition of nontargets (Pylyshyn, submitted; Pylyshyn & Leonard, Submitted). Inhibition was demonstrated using a probe dot method devised by (Watson & Humphreys, 1997). Inhibition of rejected objects is a general phenomenon and several varieties of attentional inhibition have been studied, the best known being Inhibition of Return (IOR). The finding

of inhibition of nontargets raises special problems for visual indexing theory. What is inhibited seems to be the nontargets, rather than everything that is not a target – i.e. rather than the space outside the targets (although inhibition of the region around an abrupt onset has also been shown by Mounts, 2000). Others have also found that inhibition is object-centered: IOR appears to have a significant object-centered component, as shown by (Tipper, Driver, & Weaver, 1991) and my former students (Sears, 1996; Wright & Richard, 1996) have shown that such inhibition can occur on multiple objects. But if punctate inhibition can occur on multiple nontargets, the question arises; by what mechanism can the inhibition be associated with (and travel with) moving nontargets? The only mechanism that we have proposed for this is visual indexes, but they are already being fully used to track the targets!

This puzzle suggest that we need to more carefully study such questions as whether there are limits to how many nontargets can be inhibited and what criteria must be met in order for an object to be inhibited. There has been some disagreement as to whether inhibition can be applied strategically in a top-down fashion, as (Watson & Humphreys, 1997) maintain, or whether objects are automatically marked for inhibition based on the onset of the complement target set, as argued by (Donk & Theeuwes, 2001; Theeuwes, Kramer, & Atchley, 1998). Theeuwes et al. found that as many as 15 items can be inhibited (though, unlike our nontargets, their objects were static), which is far beyond the visual index capacity and suggests that some other mechanism must be responsible. We propose to map out the options in the case of moving objects (i.e., nontargets) in *Studies 1.4* as follows.

(a) We will examine inhibition (using the dot probe method) at various distances from targets and nontargets in order to see whether the center-surround enhancement and inhibition of probe detection reported by (Mounts, 2000) is evident in the case of either targets or nontargets.

(b) We will examine inhibition on nontargets as a function of the number of nontargets, increasing the number from 4 to 10.

(c) We will examine inhibition as a function of how difficult the tracking task is by varying the distance between targets and nontargets. If, as Watson & Humphreys believe, inhibition is applied flexibly as the task and the goal requires, cases where the attentional discrimination of targets from nontargets is more difficult should lead to an increase in inhibition.

(d) We will plot the time course of inhibition in relation to such events in the tracking trial as the start of the trial, the time when additional flashes occur (as described in 2.2 below), and the time after targets stop moving. Watson & Humphreys found that under certain conditions, transient events "clear" the inhibition markers.

This study of inhibition will also provide evidence concerning the nature of the tracking mechanism. A natural interpretation of the capacity to track multiple objects in MOT is that the focal attentional beam is split into 4 or 5 separate beams, each of which is similar to the beam of focal attention studied by psychologists over the past 40 years, but weaker than a single beam. Yet MOT exhibits many properties that are different from those of an attentional beam (apart from being "weaker"). For example, as reported in the progress report section of this proposal, properties of the items that are being tracked do not appear to be encoded, objects can be tracked when they disappear behind an occluding surface (and this appears to be done without predicting where they will reappear), it seems unlikely that a set of objects is tracked by encoding and updating the targets' locations, the ID of individual objects is readily lost, the facilitation of access to property change does not spread to adjacent locations, the nontargets are not just excluded, but appear to be actually inhibited. In general MOT does not behave as though the observer had four eyes or four attentional loci for following the objects By studying the distribution of attention and inhibition around targets and nontargets in detail we may be able to show whether attention and/or inhibition spreads the way an attentional beam is known to spread (Barriopedro & Botella, 1998) or whether inhibition follows the center-surround pattern reported by (Mounts, 2000), and by examining the temporal pattern of growth and decay of inhibition and facilitation we can see whether it follows the same pattern in relation to the targets as reported for focal attention (as well as for inhibition, including IOR) whether it exhibits the same temporal pattern of attentional blink as shown for focal attention (Kristjansson & Nakayama, 2002;

Theeuwes & Godijn, 2002) and summarized in the Feature Gate model (Cave, 1999) or the attentional dynamics modeled by the "episodic" theory of (Sperling & Weichselgarter, 1995).

2 Factors affecting assignment (and deassignment) of indexes (or selecting and deselecting sets of objects)

2.1 Are indexes assigned automatically (in a data-driven manner) or voluntarily and serially?
In recent studies (Annan & Pylyshyn, 2002; Annan & Pylyshyn, submitted) we showed that indexes can be assigned voluntarily so long as there is enough time to do so (these studies are not crucially dependent on FINST theory, so for present purposes one can read "assign indexes" as "select objects"). This was interpreted as suggesting that voluntary index assignment required that attention be scanned serially to individual targets and indexes be "dropped off," as proposed in (Pylyshyn, 2001b, in press-b). While the finding that more time is required for non-automatic assignment is suggestive, it does not entail serial scanning. It is also possible that withdrawing attention from objects that were automatically indexed incurs a "dwell time" increment in RT (Duncan, Ward, & Shapiro, 1994). We propose to extend the studies we have already reported by showing that the time to index objects that do not automatically draw attention increases with increasing number of targets and with increasing inter-target spacing (other phenomena we attributed to the use of visual indexes, such as subitizing and selecting subsets for visual search, show independence of RT and spatial dispersion or spatial precuing). *Studies 2.1* will replicate the design we used in the above studies but will systematically vary number of targets and their mean spacing. The number of targets will always be kept under the maximum number that can be tracked in MOT (i.e., 3-5) and their dispersion will be kept above the Intriligator limits for attentional resolution (spacing will be more than 3º of visual angle for a screen size of 25º of visual angle). If targets are indexed serially we would expect the selection performance for non-automatically indexed objects (e.g., for objects with properties that do not lead to popouts in visual search or to preattentive texture segregation) to improve significantly as the display time increases (compared with automatically indexed properties) and that this effect would be greater for larger numbers of targets (since more targets lead to longer times for a serial process). We measure indexing performance in terms of the number of objects successfully tracked for short trial durations, since we assume that for short tracking trials the only performance limitation is due to indexing or selection. Tracking performance is also used instead of a more direct measure of selection (e.g. indicate which objects blinked) in order to eliminate the use of location recall alone since it is know that location-persistence is strong in vision (Krekelberg, 2001).

2.2 What causes indexes to be deassigned or lost?
One of the factors that makes MOT appear to be difficult and attention-demanding is that whatever signals cause indexes to be assigned to objects might also result in their being erroneously *reassigned* to task-irrelevant event under certain conditions. We have just begun a systematic investigation of this question. We have shown (Kushnier & Pylyshyn, submitted) that flashing objects on and off, which is one of the best cues for assigning indexes, can also lead to errors when the flashes occur during the tracking trial. Around the middle of a 15 sec target tracking trial, with 4 targets and 4 nontargets, we flashed 4 of the 8 objects; two targets that were being tracked and 2 nontargets. In analyzing the data we selected only those trials (about 37% of all trials) in which exactly one tracking error was made, because these trials are ones in which we could use our record of object trajectories to unambiguously determine which target had been lost and replaced by which nontarget (objects have an internal number that allows us to keep track of which ones were targets, which ones had been flashed and which ones were chosen in the response). We found that flashed targets were no more likely to be lost (and result on an error) that nonflashed targets, but flashed nontargets were significantly more likely to capture a tracked target and to generate substitution errors than nonflashed nontargets. We interpreted these results to suggest that once objects were indexed for tracking, flashing them did not tend to cause the index to be dropped, but flashing a nontargets tended to draw indexes away from tracked targets. The same did not hold for gradually introduced targets or nontargets. In a separate pilot study we showed that observers could easily add new targets (up to a maximum of 4 or 5) that were introduced

gradually either at fixation or from the side of the display (these studies were originally carried out to see whether the number of targets that were tracked could be increased by adding them in gradually; we found that the number of targets that could be tracked was not increased in this way, but neither was tracking disrupted by this gradual introduction of new targets).

In replicating and extending these pilot studies we plan to systematically explore the factors that result in the loss of tracked objects. The proposed studies include introducing a secondary task that requires general attentional resources, a task that involves the automatic seizing of attention (or indexes) by transient visual signals, and a monitoring task that involves visually attending to targets, nontargets or task-irrelevant locations in the display. Studies 2.2 will compare the impact that each type of secondary task has on MOT performance. Among the variables we will test are (a) changes in luminance/color/shape of targets (complementing studies 1.2) (b) whether tasks that do not require spatially moving attention away from targets (e.g. monitoring events on targets themselves) interfere with MOT performance, (c) whether onset events close to targets or to nontargets are more disruptive than ones farther away. The more general question of what resources MOT draws on will be examined by the proposed experiments described in the next section.

2.3 Does MOT use the same limited resources as other attentive visual tasks?

MOT is a difficult and attention-demanding task. Yet, according to visual index theory, indexed objects reta in their assigned indexes automatically. Why then does the task seem so difficult? Clearly carrying out the MOT task draws on *some* limited resource. We showed (Kushnier & Pylyshyn, submitted) that onsets occurring in the visual field during tracking draw indexes away from tracked targets resulting in tracking errors. Perhaps other events in the experimental situation also do. Thus it may be that inhibiting irrelevant events from seizing the limited number of indexes is what demands attention, rather than keeping track of targets. Yet, as reported in the progress report section of this proposal, we have shown that a fairly difficult visual monitoring task does not produce any decrement in performance, *providing a response is not required during the tracking trial*, but only at the end of each trial (Leonard & Pylyshyn, submitted). This suggests that MOT may not draw on the same attentional resources as required for monitoring and recall. (Alvarez et al., 2000) also showed that MOT does not appear to draw on the same resources as visual search, since both can be performed without mutual decrement. While it is too early to draw strong conclusions, these results hint at the interesting possibility that the basic mechanism of index maintenance may itself be preattentive (as initially proposed by the FINST theory).

We propose to continue to pursue this possibility in the next phase of the research. First we will vary the nature and difficulty of the secondary task. One task that has been used to monitor attentional load is an RSVP task presented at fixation (e.g., one version of RSVP requires that subjects monitor a rapid – 15-20 items/sec – sequences of letter presentations for the occurrence of a particular letter – see examples in Shapiro, 2001). We plan to run studies of MOT in which an RSVP monitoring task will be done concurrently, with the response being reported at the end of the trial. We will instruct subjects that the MOT task is the primary task, although we will record the RSVP performance as well. To prevent a tradeoff of resources between the MOT task and the secondary RSVP task we will ensure that the secondary task always meets a high criterion of performance.

A pilot study showed that in with two alternative forced choice color monitoring task, performance on the secondary task was worse at fovea, perhaps due to inattentional blindness. We will investigate more difficult localized tasks (monitoring which direction a briefly presented arrow points) when the task is presented on a target or on a nontarget, and how this affects tracking performance. It is possible that if attention does not have to be spatially moved from the tracked targets, MOT performance will not be degraded by a secondary monitoring task.

We also plan to use an indirect attention-load measure based on the duration of motion after-effect (MAE) from a peripheral flow field presented during the MOT task and assessed at the end of the trial. (Chaudhuri, 1990) showed that the duration of MAE decreases when an attentionally demanding task is performed in the presences of an irrelevant expanding field of dots at the periphery. More recently

(Rees, Frith, & Lavie, 1997) also showed that the degree of MAE was modulated by the attentional load of an irrelevant task, as measured both behaviorally and by degree of motion-related activity in cortical area V5. In addition, the MAE measure would also be expected to decrease with the attentional load of the MOT task because higher attention demand tends to focus attention more narrowly at the fovea, so the motion flow stimulus, being peripheral, should have a weaker effect when the task demand of the main task is higher. On the bases of these two considerations we will carry out experiments in which subjects engage in a typical MOT task while a field of expanding dots is presented in the periphery (at around 10º – 12º from fixation). At the end of the trial (and after the target selection responses) a full field of stationary dots will be presented. We will use the method used by Rees et al. and by Chaudhuri to assess MAE strength and duration and will compare the MAE measure for MOT with 3, 4 and 5 objects as well as with baseline conditions where observers have to pick out the targets when the objects all remain stationery and also when all objects disappear for the trial duration and then reappear. In all conditions the target objects would be selected using a mouse and then a full field of random dots would appear and observers would have to indicate when these dots were perceived as stationery, exactly as in the (Chaudhuri, 1990) studies.

2.4 Indexing and location encoding: Are objects selected through prior encoding of their location?
In Visual Index Theory, the claim is made that the first stage in the attentional selection of objects involves (a) *individuation* and (b) *indexing* (or deictic referencing). This means that before properties of objects can be encoded, the objects must first be individuated and a pointer assigned to them; only subsequently can their shape or color or location be encoded. This pointer is assumed to be like a pointer in a computer data-structure: It does not itself reply on a code for the physical location of the object that is indexed, but simply provides a functional reference to that object. Another way to word this assumption is that the arguments to visual predicates, such as **Red**(x) or **Inside** (x,y) or **Collinear**(x,y,z), can only be evaluated if all their arguments are bound to visual objects, where the mechanism for this binding is called a visual index. This assumption flies in the face of the widely-held view that objects can only be selected by first encoding their locations. For example, Featre Integration theory (Treisman & Gelade, 1980) assumes that in detecting a conjunction of properties the viewer first detects one conjunct and then uses its location, as represented in the feature map, to locate and detect the second conjunct (but see Goldsmith, 1998, for a different view). This particular claim in Visual Index Theory has far reaching implications since it states that when an object is first encountered, the visual system creates a special sort of reference to it that does not identify it by its unique properties, but only by its individuality. Such a reference is sometimes called a deictic (or demonstrative) reference (see Ballard et al., 1997; for a more detailed discussion of the need for deictic references see Pylyshyn, 2000; Pylyshyn, 2001b). For present purposes the essential claim is that objects are not selected by their location or by any other encoded property, but by their individual identity *as the same token object.* MOT provides one of the more direct illustrations of this claim since objects are tracked and identified independent of their location or any other property (and we argued that the parameters of the task in most studies prohibit the objects from being tracked by encoding and updating their locations, Pylyshyn & Storm, 1988).

We have recently begun to develop other direct demonstrations that individuating and indexing objects precedes encoding their location. One of these techniques involves using subitizing – the rapid enumeration of sets of objects when the cardinality of the set is about 4 or less. We had earlier argued that subitizing may involve the rapid enumeration of active indexes without the need to scan the display to locate the objects themselves (which is done automatically and in parallel by the indexing mechanism). We operationalize the existence of subitizing as the presence of an initial linear segment of the function relating reaction time to number *(n)*, which is then followed by a significant quadratic or higher order component. Using that criterion we find that subitizing only occurs if there is automatic individuation (based on simple "popout" features), that the slope (or enumeration speed) is insensitive to the spatial locations of objects or to precuing their locations (these results have been described in detail in several publications, and are summarized in Trick & Pylyshyn, 1994a). If subitizing requires only individuating objects and counting indexes, and if indexing does not require the prior encoding of location, then it might be possible to subitize objects without being able to recall their location. In pilot studies in our laboratory we showed that when observers do subitize sets of objects they were nevertheless poor at indicating which of the locations on the screen these objects had occupied (Dennis & Annan, submitted). This is consistent with the principle of object-based attention and with the finding that visual short-term memory is measured in terms of objects rather than locations (Lee & Chun, 2001; Luck & Vogel, 1997).

The particular technique we use relies on an observation we made during our pilot research. We found that objects of a particular color could be selected very rapidly (as in "select the *reds* among the *greens*") compared to the time it takes to select objects by other properties such as shape or orientation (in fact we found that red/green colored objects could be selected in the time of one screen fill, or 8.5 ms for the 117 Hz display, plus phosphor persistence time). This may be due to low-level display properties such as saturation and luminance differences among the colors (and we are currently exploring this question). It is, however, consistent with findings that one can efficiently use color cues to narrow a visual search (Laarni, 2001). The pilot experiments we reported in (Dennis & Annan, submitted) involve presenting red and green objects (small squares) with the instruction to count those of one particular color – say, the red ones. After the brief (300 ms) colored display, all objects turned to gray. Subjects then reported how many red objects there were (by pressing one of three keys, numbered *"3", "4",or " 5"*) and then immediately after this response they used a mouse to select the 3, 4, or 5 objects they had just counted from the set of 10 objects on the screen. The preliminary finding is that the probability of counting errors was much lower than the probability of location errors. Of course the task of locating objects is very different from the task of counting them, and there may well be cross-task interference and order effects since counting preceded locating, so this result by itself does not yet provide strong evidence for the independence of individuation and location encoding. What we plan to do next (*Studies 2.4*) is to vary the time available for processing the display in this dual task. If, as we surmise, subitizing involves rapid enumeration of active indexes, without having to encode the location of indexed objects, shortening supra-threshold display times should depress accuracy in reporting the location of subitized objects since without significantly reducing subitizing performance since, presumably, encoding location requires that objects be visited serially with focal attention. Above the subitizing range (i.e., when there are more than 4 objects) both counting and locating performance should be diminished as exposure time is shortened. If both the subitizing and location tasks prove to be easy we may need to mask the display, since the time during which the display is available is critical and various low-level effects, such as positional persistence (Krekelberg, 2001), iconic memory (Becker, Pashler, & Anstis, 2000) and even phosphor persistence (see the warning in Groner, Groner, Mueller, Bischof, & et al., 1993), can prolong the effective display duration. This proposal complements *Studies 2.1* described above; both sets of studies ask whether indexing is fast and parallel, although only the present one compares time for indexing with time required for encoding of location.

The subitizing experiments will begin by using display times of 51 ms, 205 ms and 821 ms and will present 3, 4, 5, and 7 stimuli to which the subject will respond with one of four fingers of their non-dominant hand resting on keyboard keys (keys A, S, D, F for right handed subjects), while their dominant had grasps the mouse so it can be used without glancing away from the screen. After selecting the objects the subject believes is correct, he or she clicks on the button on the screen marked "continue" and receives visual feedback of which selections were incorrect. As in our earlier experiments, we will record the reaction time and number of objects correctly located, and the statistic of interest is the interaction between number and exposure. We will be carried out this analysis for all the data and also for the data divided into the subitizing and nonsubitizing range, where the cutoff will be computed separately for each subject.

We can also indirectly test the hypothesis that indexing provides the basis for enumeration but not for encoding the locations of objects by eroding the ability of objects to be indexed by color. (Maljkovic & Nakayama, 1994) showed that objects of different colors can be located much faster if the colors of targets and nontargets are blocked, as we planned to do the studies above. But they also found evidence that it is more difficult to sele ct items by color when the color of targets is changed from trial to trial and also when the color of nontargets changes from trial to trial. They attributed this to the effect of priming and inhibition: the trials blocked by target color benefited from priming by the previous trial compared with mixed target trials and the trials blocked by nontarget color benefited from inhibition of nontargets of that color. By the same token, mixing the colors of nontargets and nontargets reduced performance and made the search look like a conjunction search. This is a technique that we will also employ, though for a slightly different purpose. By increasing the difficulty of the task of locating the items to be enumerated (be preventing them from being automatically individuated) we may benefit the location task more than the counting task, because in that case the items being enumerated will have to be searched out serially and thus the location encoding may have a better opportunity to operate in tandem with the enumeration task. If this happens then we should also see signs that the enumeration task was no longer a subitizing task – it should lose its signature two-process bilinear form while the difference between accuracy in counting and in locating should decrease.

3 Object-based attention in spatially-coincident and multi-modal "objects"

3.1 Tracking spatially coincident patterns

Our study of tracking of superimposed Gabor patches has shown that observers can track objects that move through a "feature space" even when they remain fixed in physical space – see figures below (Blaser & Pylyshyn, 1999; Blaser, Pylyshyn, & Holcombe, 2000a; Blaser, Pylyshyn, & Holcombe, 2000b). Yet the demonstration leaves many open questions: Do the "objects" behave the way objects in typical MOT studies behave – can they briefly disappear without damaging tracking performance, do observers "predict" the next state of the Gabors and use that information in tracking, can these objects be automatically vs. voluntarily selected, is tracking better when the patches are kept away from each other in the "property space"? These are the very same questions we have studied in the case of regular spatial MOT, and need to be addressed for the more general case of object-based attention. These studies are continuing in collaboration with Erik Blaser, now at U Mass., Boston.

In addition to the superimposed Gabor patches, we have been exploring the task of tracking constantly-deforming overlapping amoeba-like closed curves (see illustration below). It has always been of interest in the study of object-based attention whether an entire curve could be treated as an object. While it seems unlikely that one can attentionally individuate and index an entire curve, there are several experiments that suggest that we may get close to that. For example, implicit learning, inattentional blindness and negative priming studies have all suggested that novel closed curves may be processed as a whole (DeSchepper & Treisman, 1996; Rock & Gutman, 1981; Treisman & DeSchepper, 1995). We have set up pilot studies in which up to 4 smoothly deforming curves were superimposed. (Each curve is independently computed by choosing some fixed number *n* of points – 30 in the case of the figure shown – placing one on each of a randomly arranged set of *n* radials drawn from a single fixation point, locating each point between specified limits $\mathbf{r}_1$ and $\mathbf{r}_2$, and then applying a smooth spline-fitting algorithm to the set of points to construct each curve. In the animation, the points are moved smoothly between the $\mathbf{r}_1$ and $\mathbf{r}_2$ limits using a one-dimensional random-motion algorithm similar to the one used in MOT with regard to rate of change and "bouncing" off the limits). In order to better distinguish the curves they are drawn in different colors which were continually changing but maintaining their pairwise distinctiveness, as in the studies described in 1.2. Care was taken that individual curves could not be tracked by attending only to a small region of the display. The curves proved to be difficult to track – there were particular configurations (e.g., when two curves happened to share a tangential segment) that led to confusion of curves and loss of tracking. But it was generally possible to keep track of at least one individual curve among four curves, and perhaps if we adjust the speed (and maybe even the algorithm used) for deformations, and the smoothness and distinctiveness of the curves, it might be possible to track two curves among four. We will examine the conditions under which such curved objects can be tracked and also whether they show single-object advantage effects (e.g., by asking for judgments of sudden translations and color changes).

We will also examine the possibility that what is tracked is the part-decompositions of a curve, since there is evidence that such decomposition is computed early in the visual process (Xu & Singh, 2002). We will do this by cutting out pieces of the curves at their curvature minima and determining whether the subcurves were being individually tracked. We can do that by looking at the *curve-parts* that subjects selected with the mouse at the end of each trial (measuring co-occurrence probabilities of parts of the same curve vs parts of different curves). Of course since these curves are continually morphing what constitutes the "same part" is not well-defined, unless we constrain the curve-generation process. In our algorithm we can ensure that the part structure remains constant, even while the overall shape of a curve changes. We can do that by locating the curvature minima at the beginning of a trial and ensuring that the curve maintains the same number of parts even as the parts themselves continually transform (a way of looking at this is to think of each curve as consisting of ***n*** curves broken at the curvature minima, and allowing each of these parts to morph freely within the bounds of the algorithm we use and subject to the mutual constraint that the meeting of the parts is maintained – so they always "fit" into one another even as the meeting points are allowed to transform. We have found that we can do this by applying a smoother morphing algorithm to the segment where the parts meet). These studies of superimposed patterns constitute *Studies 3.1*.

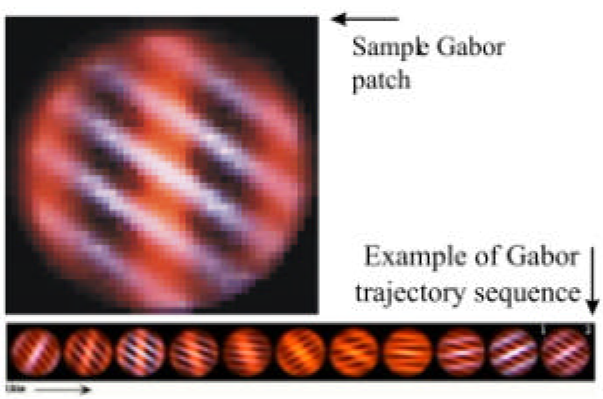


Illustration of overlapping Gabor patches display

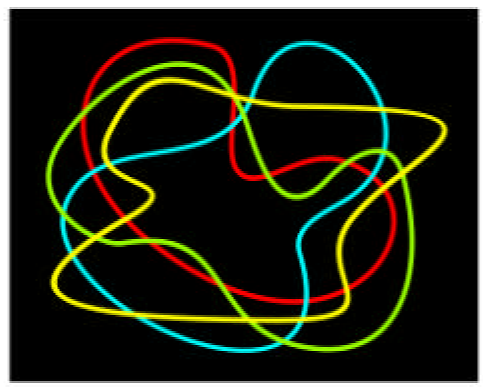


Illustration of the "Amoebas" display

3.2 Multimodal Objects

There is considerable evidence that attention shift effects can occur across modalities: sound can cue the allocation of visual attention (Spence & Driver, 1996, 1997) and auditory signals are hard to ignore at the locus of visual attention (Spence, Ranson, & Driver, 2000). Other studies include investigating whether sound-vision pairs can be created with practice when the two do not move in synchrony. Although being in the same location is important to the perception of sound-appearance objecthood, the influence of sound on visual attention can occur with or without their spatial coincidence and with or without the sound being informative as to where a visual signal would occur. We plan to reexamine some these questions in the context of the object-based attention. We will study merged visual-acoustical objects that move with or without spatial synchrony/coincidence in the two modalities. As in the Blaser et al. study, the sound and/or visual features of these merged objects will change by a sudden step and observers will be asked to indicate the nature (direction) of the change.

Instrumentation Note 3.2: We have acquired (through funding from Rutgers) an Ausim® audio location simulator (http://www.ausim3d.com/products/ausim3d.html) which takes up to 16 sound sources, together with 3D location coordinates for each, and uses a head-related transform to generate stereo output simulating the signal that would be heard if the input sounds were coming from the specified locations. We are aware that audio localization presents special problems. In our initial exploration with the Ausim we found the localization to be imprecise. The degree of imprecision depends on the match between and observer's head and the HRTs that are provided, on the type of sound signal used (percussive sounds are better than continuous tones) and on source movements (moving sources are easier to locate than static ones). In order to synchronize the location of sounds and visual objects it will be necessary to arrange for the objects to move through a larger visual angle than would be possible on a computer monitor screen. We have tried using a projecting monitor, and while it works it has shortcomings because of the distortions (shape, size and chromatic) that occur near the edge of the screen when an observer is seated close to the screen. We are therefore proposing to use the Ausim in conjunction with the Elumens VisionStation, mentioned above.

This sort of facility is appropriate not only for studying multimodal sound-vision objects, but also for studying human spatial abilities. The extraordinary spatial recall (see, for example, the findings reported by Krekelberg, 2001), together with the ability to use indexes to bind imagined objects to perceived objects, where the latter need not be visual but may be in various modalities, has become a cornerstone for my account of the spatiality of mental imagery (Pylyshyn, 2002a, 2002b; Pylyshyn, in press-a; Pylyshyn, in press-b). While the theoretical ideas concerning imagery are not directly relevant to the present proposal, the use of indexes in modalities other than vision is one of the directions I plan to take the present research.

3.3 Tracking objects in the far periphery

It had been rumored that athletes have a superior ability to track a large number of objects in MOT (although it seems that the rumor was not entirely justified Rivest, 2002). Certainly a very good basketball or hockey player appears to have a knack for keeping track of members of his or her team as well as of the opposing team. While most such prodigious skills have turned out to be domain-specific and based on being able to predict and anticipate movements of players, it is still true that even modest skill in such sports requires the ability to track players over a wide visual angle, including when they move into and past the far periphery and even out of the actual visual field (i.e., more than 90º away). Tracking moving objects that are no longer in sight requires at least the ability to recall where they went out of sight, or even better, to predict their movement and to await their reappearance in the appropriate region. It may also depend on the fact that visual sensitivity in the far periphery (even at 90º from fovea in the temporal visual field) is surprisingly high (Busey, 1999; Tynan & Sekuler, 1982; Van de Grind, Koenderink, & Van Doorn, 1987) and may affect midfield attention orientation, even when the objects in the periphery cannot be consciously reported (Lambert, Naikar, McLachlan, & Aitken, 1999). In addition to such visual skills, it is known that people can recall locations in space very accurately (Attneave & Farrar, 1977), even when the observers move about without vision (Farrell & Thomson, 1999; Rieser, 1989).

We will examine the extent to which observers can keep track of objects that sometimes move into the far periphery, and also completely off the visual field. We already know that observers can track objects that disappear briefly behind an occluding surface, and that they are able to pick up and continue tracking objects that disappear from view and reappear some 500 ms later in the same place as they were when they disappeared (Keane & Pylyshyn, submitted), so perhaps they can keep track of objects that move off their field of view. To investigate these questions we will replicate our past MOT experiments using the hemispherical display provided by the Elumens VisionStation (see *Instrumentation Note 1.1*). We will use the algorithm that maintains exclusion circles that are progressively larger with eccentricity (as described in ***c*** of section 1.1) and examine the effect on tracking performance of increasing the maximum visual angle of the MOT display. We expect that while performance may degrade somewhat as the visual angle is increased into the far periphery, performance will still remain high (over 70%) as long as the interobject distance is kept above the limit of attentional resolution measured by Intriligator. If this prediction is confirmed it could provide a new methodology for investigating the role of attention in peripheral vision, and thus its role in locomotion and sports.

Instrumentation Note 3.3: The above line of research raises technical and instrumentation issues. Ideally they require a facility that approximates as much as possible a room in which visual objects can move freely, in which the location of sound sources can be controlled, and in which the location of limbs and pointers can be monitored. Some of this equipment is available in our lab – including eye tracking equipment, the Ausim sound localizer, and flock-of-birds direction monitor – as well as the Elumens VisionStation we propose to buy. But another solution to the more difficult technical problems in the long term may require carrying out some of the experiments in other laboratories that possess more elaborate equipment. Dana Ballard and Mary Hahoe have invited us to use the NIH National Resource facilities at the Rochester Center for Vision Science. If we do decide to collaborate with these investigators we will obtain the required IRB approvals from both Rutgers and the University of Rochester, as we did in the previous grant period with Professor McConkie's laboratory at the University of Illinois.

4 Development of visual attention and its control

It is known that visual and other attention skills undergo interesting developmental changes during the preschool and early school years. For example, 6 and 8 year olds appear to be unable to take advantage of the greater predictability provided by a peripheral cue in a speeded classification task (Enns & Brodeur, 1989) as adults do. In fact younger children appear to be slower in their ability to control attention. For example young children appear to be unable to control the breadth of their attention in order to "zoom in" to a narrower range of foci (Maurer & Lewis, 1998) and are less able to use cues to preselect objects in rapid visual search (e.g., grade 2 children are poor at using color to narrow search Miller, 1978). In general younger subjects appear to be less able to filter out unwanted stimuli than older children (Pasto & Burack, 1997) and are not as good at moving their attention in a covert manner, compared with college age adults (Pearson & Lane, 1990). On the other hand, at least one study showed that children aged 5 to 12 can be as efficient as adults in selecting objects when they are required to make a spatially compatible response (Tipper & McLaren, 1990). (Enns, 1993) has argued forcibly that there is much that we can learn about how attention functions by studying its developmental progression, which can help us to draw distinctions among different types, aspects and stages of attention Thus in the present context, where we are exploring the question of whether there are preattentive (and precognitive) stages in what most people group together into the category of "attention" – such as the attention-like process of visual indexing that operates in MOT – it behooves us to look at how such processes develop. For example, it appears that subitizing, which we believe involves only the process of individuation and indexing, develops in both speed and accuracy between the ages of 2 and 5 years (Starkey, Spelke & Gelman, 1990). Yet we know very little about how rapid and largely involuntary attentional processes works in children. I have argued elsewhere that tracking is the most primitive perceptual process that precedes the concept of *individual*, or of "same object." Indeed (Leslie, Xu, Tremolet, & Scholl, 1998) has appealed to the concept of "object index" to explain infants' ability to keep track of the cardinality of small sets of objects (see Gelman, 1977 for more on numerical abilities in children). Clearly tracking (or "keeping track") is an important precursor to many important cognitive skills that develop early in life. Tracking also appears to be highly developed in older school age children, as witnessed by their proficiency and dedication to computer games.

We plan to begin our studies of the development of attention and tracking skills by conducting MOT studies on children from the age of about 5 to college age, as well as older adults. It appears from observations of our own children that even those younger that 5 are able to track 3 or 4 targets in MOT. Our initial research will consist in measuring baseline tracking performance in 6 year olds. The basic MOT task will be modified in several ways:

(i) Children will first receive some additional experience in using the mouse pointer
(ii) In addition to the mouse, we will experiment with using a touch-sensitive screen for children's responses. There have been reports of success in using such input devices with children (Fitzhugh & Katsuki, 1971), and indeed the work by (Tipper & McLaren, 1990) suggests that children can be quite proficient at demonstrating selective attention when they can interact with the display using a spatially compatible response.
(iii) Children under 12 will participate in the MOT task for only 20 minutes at a time.

(iv) The number of targets and the speed with which they move will be adjusted for each age level so that subjects will have reasonable success in tracking at least 3 targets.

(g) References Cited (* denote reports from the Pylyshyn lab, ** Appended)

Allwein, G., & Barwise, J. (Eds.). (1996). *Logical reasoning with diagrams*. New York: Oxford University Press.

Alvarez, G. A., Horowitz, J. M., & Wolfe, J. M. (2000). Multielement tracking and visual search use independent resources (Abstract). *Investigative Ophthalmology & Visual Science, 41*(4).

*Annan, V., & Pylyshyn, Z. W. (2002). *Can indexes be voluntarily assigned in Multiple Object Tracking?* Paper presented at the Vision Sciences 2002, Sarasota, FL.

*Annan, V., & Pylyshyn, Z. W. (submitted). *Voluntary Indexing requires serial visitation*. Abstract submitted for presentation at Vision Sciences 2003, Sarasota, FL.

Attneave, F., & Farrar, P. (1977). The visual world behind the head. *American Journal of Psychology, 90*(4), 549-563.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*(4), 723-767.

Barriopedro, M. I., & Botella, J. (1998). New evidence for the zoom lens model using the RSVP technique. *Perception & Psychophysics, 60*(8), 1406-1414.

Becker, M. W., Pashler, H., & Anstis, S. M. (2000). The role of iconic memory in change-detection tasks. *Perception, 29*(3), 273-286.

*Blaser, E., & Pylyshyn, Z. W. (1999). Measuring the independence of attention to multiple features (abstract). *Perception, 28*, p56.

*Blaser, E., Pylyshyn, Z. W., & Domini, F. (1999). Measuring attention during 3D multielement tracking. *Investigative Ophthalmology and Visual Science, 40*(4), 552 (Abstract).

*Blaser, E., Pylyshyn, Z. W., & Holcombe, A. (2000a). With two objects at one location, attention is "object-based" (abstract). *Investigative Ophthalmology and Vision Research*.

**Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000b). Tracking an object through feature-space. *Nature, 408*(Nov 9), 196-199.

Braun, J. (2000). Intimate attention (Science News comment on Blaser, Pylyshyn & Holcombe, 2000). *Nature, 408*(9), 154-155.

*Burkell, J., & Pylyshyn, Z. W. (1997). Searching through subsets: A test of the visual indexing hypothesis. *Spatial Vision, 11*(2), 225-258.

Busey, T. A. (1999). Localization and identification tasks rely on different temporal frequencies. *Vision Research, 39*(3), 513-532.

Cave, K. R. (1999). The FeatureGate model of visual selection. *Psychological Research, 62*(2-3), 182-194.

Chaudhuri, A. (1990). Modulation of the motion aftereffect by selective attention. *Nature, 344*(6261), 60-62.

*Cohen, E. H., & Pylyshyn, Z. W. (2002). *Searching through subsets of moving items.* Paper presented at the Vision Sciences 2002, Sarasota, FL.

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. *J Neurophysiology, 80*(5), 2657-2670.

*Currie, C. (2002). *Maintenance of FINSTs across eye movements: An ongoing research program*, http://ruccs.rutgers.edu/~ccurrie/TitlePage.html

Currie, C. B., McConkie, G. W., & Carlson-Radvansky, L. A. (2000). The role of the saccade target object in the perception of a visually stable world. *Perception and Psychophysics, 62*, 673-683.

*Dennis, J. L. M., & Annan, V. (submitted). *Can visual objects be accessed in rapid counting without their positions being encoded?* Abstract submitted for presentation at Vision Sciences 2003, Sarasota, FL.

*Dennis, J. L. M., & Pylyshyn, Z. W. (2002). *Does Multiple Object Tracking improve when objects are more discriminable?* Paper presented at the Vision Sciences 2002, Sarasota, FL.

DeSchepper, B., & Treisman, A. (1996). Visual memory for novel shapes: Implicit coding without attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*(1), 27-47.
Donk, M., & Theeuwes, J. (2001). Visual marking beside the mark: Prioritizing selection by abrupt onsets. *Perception & Psychophysics, 63*(5), 891-900.
Duncan, J., Ward, R., & Shapiro, K. L. (1994). Direct measurement of attentional dwell time in human vision. *Nature, 369*(6478), 313-315.
Efron, R., & Yund, E. W. (1999). Attentional inhibition or paracontrast? *Brain & Cognition, 41*(2), 111-149.
Enns, J. T. (1993). What can be learned about attention from studying its development? *Canadian Psychology, 34*(3), 271-281.
Enns, J. T., & Brodeur, D. A. (1989). A developmental study of covert orienting to peripheral visual cues. *Journal of Experimental Child Psychology, 48*(2), 171-189.
Farrell, M. J., & Thomson, J. A. (1999). On-line updating of spatial information during locomotion without vision. *Journal of Motor Behavior, 31*(1), 39-53.
Fitzhugh, R. J., & Katsuki, D. (1971). The touch-sensitive screen as a flexible response device in CAI and behavioral research. *Behavior Research Methods & Instrumentation, 3*(3), 159-164.
Gelman, R. (1977). How young children reason about small numbers. In N. Castellan, D. B. Pisoni and G. Potts (Eds.), *Cognitive Theory. Vol. 2*. Hillsdale, NJ: Erlbaum.
Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language, 43*(3), 379-401.
Goldsmith, N. (1998). What's in a location? Comparing object-based and space-based models of feature integration in visual search. *Journal of Experimental Psychology: General, 127*(2), 189-219.
Groner, R., Groner, M. T., Mueller, P. U., Bischof, W. F., & et al. (1993). On the confounding effects of phosphor persistence in oscilloscopic displays. *Vision Research, 33*(7), 913-917.
Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of attention. *Cognitive Psychology, 4*(3), 171-216.
Intriligator, J. M. (1997). *The spatial resolution of attention.* Unpublished Ph.D., Harvard, Cambridge, MA.
Irwin, D. E. (1993). Perceiving an integrated visual world. In D. E. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV*. Cambridge, MA: MIT Press.
*Keane, B., & Pylyshyn, Z. W. (submitted). *Does tracking disappearing objects in MOT involve predicting the locus of reappearance?* Abstract submitted for presentation at Vision Sciences, Sarasota, FL.
*Kelly, D. (2000). *In the Blink of a Flicker: Testing the Error Recovery Hypothesis using Multiple Object Tracking* (Cognitive Science Certificate Project). New Brunswick, NJ: Rutgers Center for Cognitive Science.
Krekelberg, B. (2001). The persistence of position. *Vision Research, 41*(4), 529-539.
Kristjansson, A., & Nakayama, K. (2002). The attentional blink in time and space. *Vision Research, 42*(17), 2039-2050.
*Kushnier, A., & Pylyshyn, Z. W. (submitted). *Do flashing objects grab visual indexes away in multiple object tracking?* Abstract submitted for presentation at Vision Sciences 2003, Sarasota, FL.
Laarni, J. (2001). Colour-cueing in visual search. *Scandinavian Journal of Psychology, 42*(1), 71-78.
Lambert, A., Naikar, N., McLachlan, K., & Aitken, V. (1999). A new component of visual orienting: Implicit effects of peripheral information and subthreshold cues on covert attention. *Journal of Experimental Psychology: Human Perception & Performance, 25*(2), 321-340.
Lee, D., & Chun, M. M. (2001). What are the units of visual short-term memory, objects or spatial locations? *Perception & Psychophysics, 63*(2), 253-257.
*Leonard, C., & Pylyshyn, Z. W. (submitted). Measuring the attentional demand of multiple object tracking (MOT), *Vision Sciences 2003*. Sarasota.

*Leonard, C., Pylyshyn, Z. W., Dennis, J. L., & Cohen, E. H. (2002). *The effect of a secondary monitoring task on Multiple Object Tracking.* Paper presented at the Vision Sciences 2002, Sarasota, FL.

Leslie, A. M., Xu, F., Tremolet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: Developing `what' and `where' systems. *Trends in Cognitive Sciences, 2*(1), 10-18.

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

Maljkovic, V., & Nakayama, K. (1994). Priming of popout: I. Role of features. *Memory and Cognition, 22*(6), 657-672.

Maurer, D., & Lewis, T. L. (1998). Overt orienting toward peripheral stimuli: Normal development and underlying mechanisms. In *Richards, John Edward (Ed). (1998). Cognitive neuroscience of attention: A developmental perspective. (pp. 51-102). x, 452pp.*

Miller, L. K. (1978). Development of selective attention during visual search. *Developmental Psychology, 14*(4), 439-440.

Mounts, J. R. W. (2000). Attentional capture by abrupt onsets and feature singletons produces inhibitory surrounds. *Perception & Psychophysics, 62*(7), 1485-1493.

Pasto, L., & Burack, J. A. (1997). A developmental study of visual attention: Issues of filtering efficiency and focus. *Cognitive Development, 12*(4), 427-439.

Pearson, D. A., & Lane, D. M. (1990). Visual attention movements: A developmental study. *Child Development, 61*(6), 1779-1795.

*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. (1998). Visual indexes in spatial vision and imagery. In R. D. Wright (Ed.), *Visual Attention* (pp. 215-231). New York: Oxford University Press.

**Pylyshyn, Z. W. (2000). Situating vision in the world. *Trends in Cognitive Sciences, 4*(5), 197-207.

*Pylyshyn, Z. W. (2001a). Connecting vision and the world: Tracking the missing link. In J. Branquinho (Ed.), *The Foundations of Cognitive Science* (pp. 183-195). Oxford, UK: Clarendon Press.

**Pylyshyn, Z. W. (2001b). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*(1/2), 127-158.

*Pylyshyn, Z. W. (2002a). Mental Imagery: In search of a theory. *Behavioral and Brain Sciences, 25*(2), 157-237.

*Pylyshyn, Z. W. (2002b). Where is fancy bred … in the world or in the head? Reply to Commentators. *Behavioral & Brain Sciences, 25*(2), xxx-xxx.

*Pylyshyn, Z. W. (in press-a). Return of the Mental Image: Are there pictures in the brain? *Trends in Cognitive Sciences, x*(x), xxx.

*Pylyshyn, Z. W. (in press-b). *Seeing and visualizing: It's not what you think*. Cambridge, MA: MIT Press/Bradford Books.

**Pylyshyn, Z. W. (submitted). Tracking without keeping track: some puzzling findings concerning multiple object tracking. (To be published after minor revisions)

*Pylyshyn, Z. W., & Eagleson, R. A. (1994). Developing a network model of multiple visual indexing (abstract). *Investigative Ophthalmology and Visual Science, 35*(4), 2007-2007.

*Pylyshyn, Z. W., & Leonard, C. (Submitted). *Inhibition of Nontargets During Multiple-Object Tracking.* Paper presented at the Vision Sciences 2003, Sarasota.

*Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 1-19.

Rees, G., Frith, C. D., & Lavie, N. (1997). Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science., 278*(5343), 1616-1619.

Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 1157-1165.

Rivest, J. (2002). *Tracking performance in professional athletes*. Personal Communication, (Sarasota, May, 2002)

Rock, I., & Gutman, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 275-285.

*Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology, 38*(2), 259-290.

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

*Scholl, B. J., Pylyshyn, Z. W., & Franconeri, S. L. (1999). When are featural and spatiotemporal properties encoded as a result of attentional allocation? *Investigative Ophthalmology & Visual Science, 40*(4), 4195.

*Scholl, B. J., Pylyshyn, Z. W., & Franconeri, S. L. (submitted). The relationship between property-encoding and object-based attention: Evidence from multiple-object tracking.

*Sears, C. (1996). *Visual Indexing and Inhibition of Return of Visual Attention.* Unpublished PhD, University of Western Ontario, London, Canada.

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

Shapiro, K. (Ed.). (2001). *The limits of attention: Temporal constraints in human information processing*. New York: Oxford University Press.

Slezak, P. (1999). Situated cognition: Empirical issue, paradigm shift or conceptual confusion? In J. Wiles & T. Dartnall (Eds.), *Perspectives on Cognitive Science, Vol 2* (pp. 69-98). Stamford, CT: Ablex Publishers.

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Perception and Psychophysics, 22*, 1005-1030.

Spence, C., & Driver, J. (1997). Audiovision links in in exogenous covert spatial attention. *Perception and Psychophysics, 59*, 1-22.

Spence, C., Ranson, J., & Driver, J. (2000). Crossmodal selective attention: on the difficulty of ignoring sounds at the locus of visual attention. *Perception and Psychophysics, 62*, 410-424.

Sperling, G., & Weichselgarter, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review, 102*(3), 503-532.

Starkey, P., Spelke, E.,and Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, **36**, 97-127.

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge, England UK: Cambridge University Press.

Theeuwes, J., & Godijn, R. (2002). Irrelevant singletons capture attention: Evidence from inhibition or return. *Perception & Psychophysics, 64*(5), 764-770.

Theeuwes, J., Kramer, A. F., & Atchley, P. (1998). Visual marking of old objects. *Psychonomic Bulletin and Review, 5*(1), 130-134.

Tipper, S., Driver, J., & Weaver, B. (1991). Object-centered inhibition of return of visual attention. *Quarterly Journal of Experimental Psychology, 43A*, 289-298.

Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views. *Quarterly Journal of Experimental Psychology B, 2*, 321-343.

Tipper, S. P., & McLaren, J. (1990). Evidence for efficient visual selectivity in children. In J. T. Enns (Ed.), *The development of attention* (pp. 197-210). New York: North Holland.

Treisman, A., & DeSchepper, B. (1995). Object Tokens, Attention, and Visual Memory. In T. Inui & J. McClelland (Eds.), *Attention and Performance XVI: Information Integration in Perception and Communication* (pp. 15-46). Cambridge, MA: MIT Press.

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

*Trick, L. M., & Pylyshyn, Z. W. (1994a). Cuing and counting: Does the position of the attentional focus affect enumeration? *Visual Cognition, 1*(1), 67-100.

*Trick, L. M., & Pylyshyn, Z. W. (1994b). Why are small and large numbers enumerated differently? A limited capacity preattentive stage in vision. *Psychological Review, 101*(1), 80-102.

Tynan, P. D., & Sekuler, R. (1982). Motion processing in peripheral vision: Reaction time and perceived velocity. *Vision Research, 22*(1), 61-68.

Van de Grind, W. A., Koenderink, J. J., & Van Doorn, A. J. (1987). Influence of contrast on foveal and peripheral detection of coherent motion in moving random-dot patterns. *Journal of the Optical Society of America, A*, 1643-1652.

VanLehn, K., & Ball, W. (1991). Goal reconstruction: How TETON blends situated action and planned action. In K. VanLehn (Ed.), *Architectures for Intelligence* (pp. 147-188). Hillsdale, NJ: Lawrence Erlbaum Associates.

vanMarle, K., & Scholl, B. J. (submitted). Attentive tracking of objects vs. substances.

Watson, D. G., & Humphreys, G. W. (1997). Visual marking: prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychological Review, 104*(1), 90-122.

Wright, R. D., & Richard, C. M. (1996). Inhibition-of-return at multiple locations in visual space. *Canadian Journal of Experimental Psychology, 50*(3), 324-327.

Xu, Y., & Singh, M. (2002). Early computation of part structure: Evidence from visual search. *Perception and Psychophysics, 64*, 1039-1054.

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