

Multiple Object Tracking and Attentional Processing

Christopher R. Sears
University of Calgary

Zenon W. Pylyshyn
Rutgers University

How are attentional priorities set when multiple stimuli compete for access to the limited-capacity visual attention system? According to Pylyshyn (1989) and Yantis and Johnson (1990), a small number of visual objects can be preattentively indexed or tagged and thereby accessed more rapidly by a subsequent attentional process (e.g., the traditional "spotlight of attention"). In the present study, we used the multiple object tracking methodology of Pylyshyn and Storm (1988) to investigate the relation between what we call "visual indexing" and attentional processing. Participants visually tracked a subset of a set of identical, independently randomly moving objects in a display (the "targets"), and made a speeded identification response when they noticed a target or a nontarget (distractor) object undergo a subtle form transformation. We found that target form changes were identified more rapidly than nontarget form changes, and that the speed of responding to target form changes was unaffected by the number of nontargets in the display when the form-changing targets were successfully tracked. We also found that this enhanced processing only applied to the targets themselves and not to nearby nontarget distractors, showing that the allocation of a broadened region of visual attention (as in the "zoom-lens" model of attentional allocation) could not account for these findings. These results confirm that visual indexing bestows a processing priority to a number of objects in the visual field.

It is widely accepted that visual attention can be shifted from one location to another independently of eye movements, and that the processing of stimuli appearing at attended locations is enhanced. The methodological paradigm that produced much of the evidence for this is the attentional cueing procedure (Posner, Snyder, & Davidson, 1980). In a cueing experiment, visual attention is shifted to a predetermined location either endogenously, in which the shift is under the volition of the observer, or exogenously, in which the shift is involuntarily elicited by a highly salient cue. A substantial number of studies have demonstrated that the speed and accuracy of processing at cued locations is superior to that at uncued locations. For example, Downing (1988) found that perceptual sensitivity was enhanced at the location of a cue, and concluded that focused attention serves to facilitate visual information processing.

Previous research suggests that, whether controlled endogenously or exogenously, there can be only one focus of attention at any one time. Posner, Snyder, and Davidson (1980) provided evidence that visual attention is allocated to single contiguous regions of the visual field, enhancing the processing of stimuli falling within the single contiguous "spotlight". Eriksen and St. James (1986) subsequently showed that this enhanced processing falls off monotonically as one moves out from the locus of visual attention, and that the resolution of the spotlight varies inversely with the size of the region encompassed (the "zoom-lens" model). Many investigators have concluded that the spotlight is the primary processing bottleneck of the attentional system, as only stimuli falling within this region undergo extensive perceptual

analysis (e.g., Eriksen & Hoffman, 1974; Yantis & Johnston, 1990), and only one such region can be attended to at any one time (Eriksen & Yeh, 1985). There is also evidence that focal attention may be directed at objects in the field of view rather than at spatial regions (Baylis and Driver, 1993) and that the objects selected in this way continue to be identified as the same objects as they move about (Kahneman and Treisman, 1992; Pylyshyn, 1989).

Although the evidence for the unitary nature of focal attention is somewhat contentious (e.g. Castiello & Umiltà, 1990; Joula, Bouwhuis, Cooper, & Warner, 1991; Pashler, 1998), there do seem to be severe limitations on the allocation of focal attention. Thus, when multiple stimuli compete for attention there must be some mechanism that prioritizes the stimuli in some way so that stimulus selection can proceed in an efficient manner. Consequently, an understanding of the means by which attentional priorities are assigned is of some importance.

Yantis and Johnson (1990) sought to determine how attentional priorities are set under conditions in which attention is allocated in a stimulus-driven manner. They had participants search for a target letter in static multielement displays composed of multiple abrupt onset and no-onset items. Abrupt onset and no-onset items are distinguished by their attentional saliency: No-onset items are presented by removing camouflaging line segments from an existing object in the display until the target is revealed, while onset items abruptly appear in previously empty locations of the display. In previous studies, Yantis and his colleagues (Jonides & Yantis, 1988; Yantis & Jones, 1991; Yantis & Jonides, 1984) found that single abrupt

onsets automatically capture attention, and that abrupt onset items are processed before other no-onset items in a display. Yantis and Johnson (1990) found that when multiple abrupt onsets are present, a limited number of them (approximately four) can be processed before other no-onset items. They proposed that attentional priorities are set by means of attentional “tags” that are bound to the representations of highly salient objects (such as abrupt onsets), and that a limited number of abrupt onsets can be attentionally tagged and then given priority access to focused attention.

Like Yantis’ priority tag model, Pylyshyn’s (1989) FINST model of visual indexing provides a means for setting attentional priorities among multiple stimuli. In the FINST model, a limited number of objects can be preattentively and simultaneously indexed independently of their retinal locations or identities (FINST is an acronym for FIngers of INSTantiation, a reference to the idea that these indexes point to objects and provide a way to bind objects to internal symbolic arguments). Pylyshyn (1989) suggested that a primary function of visual indexing is to individuate a small number of objects so that they may be directly accessed and subjected to focused attentional processing. Indexing provides direct access to the objects, so that once an object is indexed it is not necessary to use attentional scanning to find that object (Pylyshyn et al., 1994). In contrast, in order to selectively attend to a non-indexed object, its position must first be ascertained through attentional scanning (e.g., Treisman & Gelade, 1980; Tsal & Lavie, 1993). Visual indexing thus provides a means of setting attentional priorities when multiple stimuli compete for attention, as indexed objects can be accessed and attended before other objects in the visual field.

In the FINST model, visual indexes are assigned primarily in a stimulus-driven manner, so that salient feature characteristics or changes are automatically indexed. Typical stimulus events that would be indexed roughly correspond to stimuli that automatically attract focal attention, such as peripheral cues (Posner & Cohen, 1984), abrupt visual events such as onset stimuli (Yantis & Jonides, 1984), and “pop-out” visual features (e.g., a red circle embedded in a display of black circles; Treisman & Gelade, 1980).

Burkell and Pylyshyn (1997) have provided more direct evidence of the ability of several simultaneous onsets to attract indexes which could then be used to access and examine the indexed items. They used Treisman-type visual search tasks to study the selection of search subsets by onset cues, and they appealed to the fact that visual

search behavior depends on the nature of the set through which one has to search. In the original visual search studies, Treisman and Gelade (1980) showed that search targets that differed from each nontarget in the search set by a single feature (called the “single-feature” search condition) were easily recognized and the time it took to recognize them was very nearly independent of the number of nontargets in the search set. In contrast, displays in which each nontarget shared some feature with the target, so that only a combination of two features together defined the target (called the “conjunction-feature” search condition), resulted in generally slower searches as well as reaction times that increased substantially with increasing numbers of nontargets in the display. Burkell and Pylyshyn used displays which were of the “conjunction search” type. However, they selected subsets through which subjects were required to search, and the *subsets* could constitute either a single-feature or a conjunction-feature search.

The subsets were selected as follows. All members of the search set were precued by “X” place holders, but a subset of variable size that was to contain the target was cued by placeholders that occurred about one second later than the other placeholders, and about 100 ms before the search began. The precued subset itself could constitute either a single-feature or a conjunction-feature search task (i.e., the target could be distinguished from members of the subset by one feature or only by a conjunction of two features). Burkell and Pylyshyn found that a subset precued by these late onset cues could, for the purposes of speeded visual search, be separated from the rest of the display and treated as though they were the only items present. Burkell and Pylyshyn showed, for example, that only the single-feature versus conjunction-feature property of the *subset* was relevant to the pattern of search reaction times (the overall display always provided a conjunction-search condition because it contained items with all combinations of properties). Moreover, they found that increasing the relative distance between search targets did not increase the search time, showing that the indexed targets could be accessed without having to find them first by scanning the display.

Visual indexes point to features or objects, not to the locations that these stimuli occupy. Like the object files of Kahneman, Treisman, and Gibbs (1992), visual indexes are object-centered and continue to reference objects despite changes in their location. According to the visual indexing model, a visual index automatically individuates and tracks moving objects. Because there are a small

number (around 4 or 5) of such indexes, observers can track around 4 or 5 independently-moving distinct visual objects in parallel. In an empirical test of this hypothesis, Pylyshyn and Storm (1988) had participants visually track a prespecified subset of a larger number of identical, randomly moving objects in a display. The members of the subset to be tracked (the targets) were identified by briefly flashing them several times, prior to the onset of movement. According to the model, targets designated in this fashion are automatically indexed. During the tracking task the targets were indistinguishable from the other distractor objects, which made the historical continuity of each target's motion the only clue to its identity. Participants tracked the target objects for 5 to 10 seconds, after which either a target or a distractor was probed by superimposing a bright square over it. The participants' task was to determine whether the probed object was a target or a distractor. According to Pylyshyn and Storm (1988), the indexing of the target objects would allow each of them to be simultaneously tracked and identified throughout the motion phase of the experiment, despite the fact that the targets were perceptually indistinguishable from the distractors. Pylyshyn and Storm (1988) found that performance in this multiple object tracking task was extremely high for subsets of up to 5 elements—in fact, participants could simultaneously track up to five target objects at an accuracy approaching 90%. McKeever and Pylyshyn (1993), Yantis (1992), Scholl and Pylyshyn (in press), Viswanathan and Mingolla (1989), Cavanagh (in press), and Culham et al. (in press) have all reported similar results. Moreover, using a simulation of the task, analyses by Pylyshyn and Storm (1988) and McKeever and Pylyshyn (1993) indicate that a single spotlight of focused attention moving rapidly among the target objects and updating a record of their locations could not produce this level of tracking performance in the setup used in their study. In the Pylyshyn and Storm (1989) simulations, for example, tracking performance was no higher than 50% based on extremely conservative assumptions. One such assumption was an attentional scan velocity as high as 250 degrees per second (the highest estimated scan velocity in the previous literature). Another assumption was that participants stored (with zero encoding time) the predicted locations based on the direction and speed of the targets' motion, and that they used a guessing strategy when they were uncertain. These results suggest that it is not possible for the task to be performed at the observed level of accuracy without some parallel tracking of the target objects. Pylyshyn

and Storm (1988) concluded that their results confirmed a key prediction of the visual indexing model; namely, that a small number of objects can be indexed and that the indexes are used to keep track of them in parallel without attentional scanning and without encoding their locations.

We should note that Yantis (1992) has added an additional mechanism to explain how multiple target objects are tracked in this task. Yantis (1992) argues that participants spontaneously group the targets together to form a virtual polygon, whose vertices correspond to the continually changing positions of the targets, and that it is this single "object" that is tracked throughout the trial. While it may be that observers conceptually group elements into a polygon it is still the case that the individual targets themselves must be tracked in order to keep track of the location of the vertices of such a virtual polygon. Consequently, we do not consider Yantis's (1992) account to be incompatible with Pylyshyn and Storm's (1988) analysis. Indeed, we have offered a closely related proposal for what we call an "error recovery" stage of the tracking process, wherein a polygon-like representation of the relative location of targets is maintained and referred to when the loss of a target is detected (McKeever & Pylyshyn, 1993; Pylyshyn et al., 1994).

The purpose of the present investigation was to explore the relation between visual indexing and attentional processing in the context of the multiple-object tracking paradigm. According to Pylyshyn (1989), visual indexing provides a means of keeping track of a number of objects, in the sense of providing a means for querying them, without first having to ascertain their positions through attentional scanning. We assume that, in order to focus unitary attention on an object, one must first find the object and move focal attention to it. However, if the object has already been indexed, focal attention can be allocated to it *directly*, without prior search. Consequently, shifting attention to indexed objects should be faster than shifting attention to non-indexed objects. Because of this we expect that in a task requiring focal attention, a response to an indexed object will generally occur before a response to other objects in the visual field.

This hypothesis was investigated in Experiment 1. Participants tracked a set of target objects and made a speeded identification response when they detected that a target or a distractor object underwent a subtle form transformation. According to our hypothesis, targets are indexed during the tracking task, and therefore unitary focal attention can be moved directly to them. If focal atten-

tion is shifted to a target, either on some regular basis or when some global change-detector reports a change, then any change at that target is more readily recognized. Thus, if targets undergoing changes are found to be more rapidly identified this would support our hypothesis that visual indexing facilitates attentional processing.

Experiment 1

Method

Participants. Seventeen individuals participated in this experiment. All had normal or corrected-to-normal vision, and were paid \$15.00 for their participation.

Apparatus and stimuli. Stimuli were presented on a 19-inch Sun Microsystems monochrome monitor with a resolution of 1056 by 900 pixels. A Sun Microsystems 3/50 microcomputer controlled the stimulus presentation and randomization of trials. Response times were collected with a three button mouse and a dedicated Zytec timing board (Danzig, 1988) which provided response latencies accurate to one millisecond.

Stimuli consisted of rectangular, seven-segment box figure eights and the capital letters E and H. The E and H characters were created by removing lines from the figure eights. The line segments used to construct these stimuli were a single pixel in width. The E and H figures were used because of their high degree of similarity to the figure eight character.

All the stimuli were created as pixel-drawn "objects" in video memory that could be moved across the screen without being continually recreated. Each object subtended a visual angle of 0.84 degrees in height and 0.63 degrees in width from a viewing distance of 50 cm. Stimuli were white and were presented on a dark background.

The animation of the objects in the display was controlled by an algorithm which simulated brownian motion, creating a random, independent, and continuous pattern of movement for each object. Motion sequences were generated in real-time during each trial and all the participants reported the perception of smooth and continuous motion during the practice trials. Each object was surrounded by an invisible circular barrier which insured that no two objects could collide or superimpose themselves over one another. Each object moved at a randomly determined velocity of between 4 and 8 degrees per second. Once determined, the velocity of each object did not vary throughout a trial. A wall repulsion force retained all the objects within a 15 by 15 degree area by bouncing them off these invisible borders. Although the total number of target and distractor objects displayed varied according to the trial type, the movement and velocity of these objects did not vary as a function of the number of targets and distractors present in a given trial. This was accomplished by having the maximum number of objects (16) moving in the display on every trial, but making only a subset of them visible for the particular condition concerned. This technique eliminated any relation between the density of objects in the display and the freedom of their movement.

Procedure. Participants were seated in a darkened room approximately 50 cm from the display and used a chinrest to reduce head movements and control viewing distance. A three button mouse was used to collect responses. Participants were given written instructions prior to the experiment, which outlined the general procedure and emphasized the importance of maintaining fixation throughout the session. Each participant was then given a demonstration and explanation of a trial sequence. They were instructed to note the positions of the blinking target objects at the start of each trial, because the task was to keep track of these objects when they began moving. During the motion phase, they were instructed to track the target objects without moving their gaze from the fixation cross (eye position was not monitored). At some point in the trial, one of the target or distractor objects would transform into an E or an H, and participants were to respond by pressing the E or H button as quickly as possible when they had identified this form change. Participants were informed that the target and distractor objects were equally likely to undergo form changes. Each participant completed 20 practice trials prior to the experiment.

Figure 1 depicts a trial sequence. Each trial began with the presentation of an isolated fixation cross for 2 s. Then, depending upon the condition, from 7 to 16 figure-eight objects appeared on the screen. The placement of the objects was randomly determined on each trial, subject to the constraint that none could overlap one another or their invisible "barriers". Three or four of these objects then began to blink on and off for 3 s, designating them as the target objects to be tracked. The appearance of the remaining objects (the distractors) did not change during this target designation phase. All of the objects then began to move in a random and continuous fashion about the screen (subject to the previously mentioned constraints), and the participant attempted to simultaneously track each of the target objects. The target and distractor objects were indistinguishable from one another during this tracking phase. Participants tracked the targets for a randomly determined interval of between 5 and 9 s, after which either a target or a distractor object was transformed into an E or an H. On 50% of the trials, a target object underwent a form change, and on the remaining 50% of the trials a distractor underwent a form change. The remaining targets and distractors, as well as the new letter in the display (E or H), maintained their movement during the transformation and continued moving until the participant responded. After a response had been made, the screen was cleared and a new trial was initiated following a 3 s inter-trial interval.

Each participant completed eight blocks of 45 trials. The order of trials was randomized separately for each participant, and there was a five-minute rest period between each block.

Design. There were three factors manipulated in this experiment: the type of form change (target or distractor form change), the number of targets tracked (3 or 4) and the number of distractors present in the display (4, 8, or 12). There were 30 trials in each of the twelve

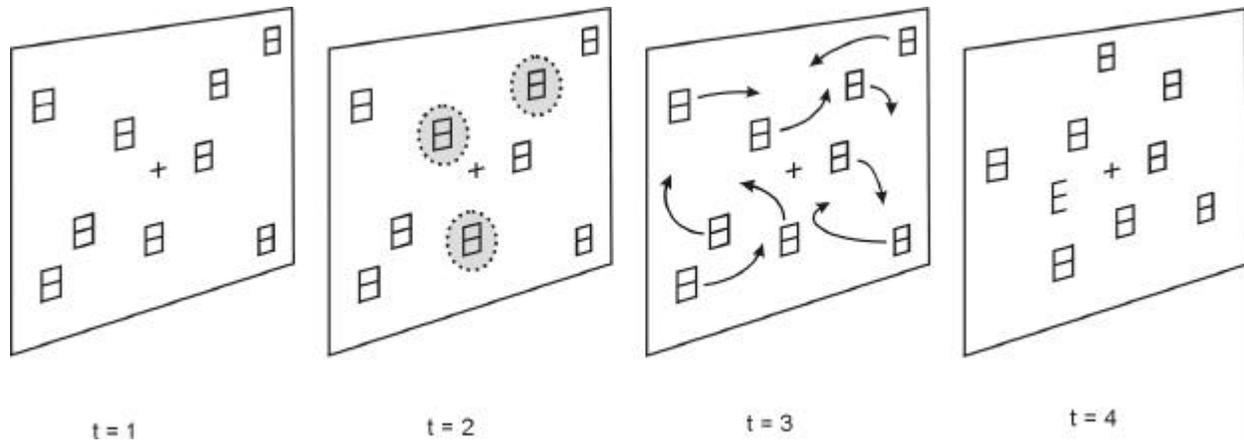


Figure 1: A schematic representation of a trial sequence. Participants view items on a video monitor. In the target designation phase ($t=2$) the target objects were flashed for 3 seconds (the selected targets are shown in circles in this illustration). The targets were then tracked for several seconds during the motion phase ($t=3$), and then a target or a distractor object underwent a form change by dropping two segments and ending up as an E or an H ($t=4$). The participants' task was to identify this form change as quickly and as accurately as possible. In this example, there are three target objects, and one of these target object undergoes a form change to an E shape.

conditions, for a total of 360 trials.

Results

Response latencies greater than 3 seconds were classified as outliers and were removed from the dataset. This procedure resulted in the removal of 3.6% of the data. Response latencies and error rates were submitted to a 2 (Type of Form Change) \times 2 (Number of Targets) \times 3 (Number of Distractors) repeated measures analysis of variance (ANOVA). The mean identification error rate (E or H) across all the conditions was 3.8%, and a three-factor repeated measures ANOVA yielded no significant effects, all F 's < 1 . (Unless otherwise stated, the p values for all significant statistics reported in the text are less than .05.)

In the analysis of the response latencies, the main effect for the number of targets tracked (3 or 4) was not significant, $F < 1$. That is, response latencies were similar whether three or four target objects were being tracked. The mean response latencies in Experiment 1, collapsed across the number of targets tracked (3 or 4), are listed in Table 1.

There was a significant main effect for the type of form change (target or distractor object), which confirmed that participants responded more rapidly to target form changes than to distractor form changes, $F(1, 16) = 29.58$, $MSE = 25,084$. Responses to target form changes were an average of 121 ms faster than responses to distractor form

changes. The main effect for the number of distractors (4, 8, or 12) was significant, $F(2, 32) = 17.04$, $MSE = 9,009$, as response latencies increased as the number of distractors in the display increased.

The interaction between the type of form change and the number of distractors was not significant, $F(2, 32) = 1.91$, $MSE = 7,772$. The absence of this interaction suggests two things: (a) that participants responded to target form changes more rapidly than to distractor form changes regardless of the number of distractors in the display, and (b) that response latencies to both target and distractor form changes increased with increases in the number of distractors. This latter finding was unexpected, and will be discussed in greater detail in the Discussion section below.

The interaction between the number of targets and the number of distractors was not significant, $F < 1$, nor was the interaction between the type of form change and the number of targets, $F(2, 16) = 3.40$, $MSE = 7,692$. Finally, the three-way interaction was not significant, $F < 1$.

Analysis of region-bounded effects. The results of this experiment clearly indicate that target form changes are identified more rapidly than distractor form changes when participants are tracking the target objects. This result supports our hypothesis that visual indexing facilitates attentional processing.

Table 1

Mean Response Latencies (in Milliseconds) and Identification Error Rates (in %) to Target and Distractor Form Changes in Experiment 1

Type of Form Change	Number of Distractors		
	4	8	12
Target			
<i>M</i>	1022	1080	1145
<i>SE</i>	43.4	34.1	36.3
Errors	3.8	4.6	3.1
Distractor			
<i>M</i>	1175	1194	1241
<i>SE</i>	51.6	50.9	46.3
Errors	3.8	3.4	4.5

Note. The data are averaged across the Number of Targets (3 or 4).

Although we contend that the latency advantage for target form changes is a consequence of the visual indexing of the target objects, there is an alternative explanation for this finding. Specifically, it is possible that during the tracking task the participant's attention was focused on the continually changing triangular or polygonal figure whose vertices constituted the three or four targets in the display. Maximum attentional sensitivity may then have been allocated to the region bounded by and including the group of target objects. Such a process could be the consequence of a deliberate attentional strategy employed by participants to enable target tracking, in which a "zoom-lens" of visual attention was focused on this region (Eriksen & St. James, 1986). If this was the case, then we would expect that responses to target form changes would be facilitated, because all of the target form changes would have occurred within this region of focused attention. Moreover, distractors undergoing form changes within this region would be identified more rapidly than distractors undergoing form changes outside of this region, and the averaging of these two types of trials would produce an apparent delay in responding to distractor form changes relative to target form changes. Thus, a zoom-lens model could account for our data by postulating that participants dynamically allocate attention to the region encompassed by the targets.

To evaluate this hypothesis, for the last six of the seventeen participants in Experiment 1, the screen coordinates of the target and distractor objects were recorded following every response. The trials in which a distractor object underwent a form change were divided into two groups: trials in which the participant responded to the form change when the distractor was within the region

encompassed by the targets, and trials where the participant responded to the form change when the distractor was outside of this region. The target locations at the time of the response served as the vertices of the region in question, with the boundaries of the entire region defined by the convex hull of the target objects (the smallest convex polygon that contained all of the targets). These data were then submitted to a one-factor repeated measures ANOVA, with distractor position (distractor inside or outside the convex hull of the targets) as the single factor. A zoom-lens account would predict an effect of distractor position in this analysis, with distractor form changes being identified more rapidly when they occurred within the convex hull of the target set.

However, in this analysis the effect of distractor position was not significant, $F < 1$. That is, response latencies to distractor form changes were similar regardless of whether the distractor was inside (1215 ms) or outside (1238 ms) the convex hull of the target set.¹ Consequently, it does not appear that the facilitation in responding to target form changes can be explained by appealing to a zoom-lens mechanism. Note that a similar result was also reported by Intriligator and Cavanagh (1992), who used a variant of the multiple object tracking task involving only two targets moving in a rigid configuration. They reported that the detection of simple luminance changes was facilitated when they occurred *on* targets being tracked, but not in the region *between* the two tracked objects.

Discussion

Why are target form changes identified more rapidly than distractor form changes during the

tracking task? According to our hypothesis, the indexes bound to target objects confer an access priority to these objects, so that target objects can be checked via focal attention before distractor objects when a form change occurs. Distractor form changes will only be identified by way of a serial self-terminating attentional scan initiated after the indexed objects have been checked. Our explanation is similar to that of Yantis and Johnson (1990), who argued that a small number of abrupt onset items can be attentionally tagged and then examined before no-onset items in visual search displays, producing a processing advantage for abrupt onset items. Similarly, our claim is that the target objects are indexed at the start of each trial, and that this indexing allows them to be continually referenced when they are in motion. When a form change occurs, the indexes bound to the target objects allow them to be directly examined via focused attention, so that target objects are checked before distractor objects for the presence of a form change. Note that we are assuming that the form change itself can be detected without the aid of focused attention. More specifically, we contend that the initial registration of the form change event is detected preattentively, perhaps by a generalized difference operator which signals that some general change in the display has occurred but does not provide precise location or identity information (e.g., Atkinson & Braddick, 1989; Scialfa & Joffe, 1995). The detection of the form change would then permit the allocation of focal attention required to identify the type of form change (E or H).

If the target objects are being checked before the distractor objects when a form change occurs, then target form changes will be identified more rapidly than distractor form changes. This would seem to be a plausible account of the data at hand, and yet this explanation is not consonant with the fact that responses to target form changes were increasingly delayed as the number of distractors in the display increased. That is, if the target objects are always checked before the distractors, then increasing the number of distractor objects in the display should have no effect on the speed at which target form changes are identified.

This application of the visual indexing model assumes, however, that the target objects are flawlessly indexed and tracked throughout every trial. This assumption is almost certainly false, as the tracking task itself is subject to error and failures of tracking do occur. It is probably the case that indexes do not always stay permanently bound to the target objects throughout the tracking task. Although we maintain that indexing and tracking a

small number of objects is preattentive, it is probably the case that maintaining those indexes over an extended period of time requires some effort or reactivation to prevent their decay or loss. A reasonable assumption is that the probability of an index being lost or misplaced would increase with the number and/or density of distractors in the display, so that increases in the number of distractors would lead to decreases in tracking performance. It is known that the attentional resolution required for individuating objects drops off rapidly with eccentricity (Intriligator, 1998), increasing the chances that when a target and a distractor object pass close to one another in the periphery of the display, the two objects may then momentarily fail to be resolved. When the target and distractor subsequently move away from each other, the index could be shifted to the wrong object or lost altogether. If an index was shifted to a distractor, the participant might then consider this object a target, while the object that had been tracked might then become a distractor.

If the index bound to a target object was lost or shifted, and the same target subsequently underwent a form change, then responses to this form change would be delayed. In effect, this target would now be a distractor, and responses to these non-indexed target form changes would be slower than responses to indexed target form changes. During the course of the experiment a fair number of these events may have occurred, increasing the average response latencies to target form changes. More specifically, as the number of distractors in the display increased, the probability that an index would be lost or shifted to a distractor would also have increased. Averaging the response latencies to form-changing targets that were tracked with those that were not tracked would create a spurious increase in response latencies to target form changes with increases in the number of distractors. In Experiment 2 we tested this possibility.

Experiment 2

If imperfections in the process of indexing and tracking the target objects was responsible for the display size effect witnessed in Experiment 1, then such an effect should not occur if the targets are perfectly tracked throughout every trial. Although it is not possible to ensure that all the targets are perfectly tracked, we have designed a procedure that may, given certain assumptions, help us to determine whether a particular target that underwent a form-change had in fact been correctly tracked on that particular trial. To do this we developed a dual-response procedure in which we

asked participants to indicate, on each trial, whether the object that underwent a form change was a target or a distractor. Thus, the participant was required to indicate the identity of the changed object and also whether it was a target or a distractor.

In this experiment, participants tracked four target objects and made a speeded identification response to a form change in the display (E or H), after which they made a two-alternative forced choice categorization decision as to what type of object underwent a form change (target or distractor). Because participants had to indicate what type of object underwent a form change on each trial, the trials where participants had lost a form-changing target could be distinguished from the trials where a form-changing target was accurately tracked. If failures in tracking performance produced the display size effect for target form changes in Experiment 1, then excluding those trials where participants had lost a form-changing target from the RT analyses should greatly attenuate and perhaps even eliminate this display size effect.

Of course, the success of the method depends on participants being able to categorize the changed objects as ones they had or had not tracked. Because there will be some trials on which participants may not be certain as to whether the object that underwent a form change was a tracked object, their category assignment may contain both errors of omission and errors of commission. Such errors may occur for several reasons. For example, a target may be lost because the index simply becomes dissociated from it – in which case the forced-choice categorization response may merely reflect a guessing strategy. On the other hand, if the target is lost because the index shifts to a nearby distractor, the participant may unknowingly take the newly indexed object to be a target or might even be aware of the shift and exercise some wariness. Clearly, our ability to isolate those trials in which the participant correctly tracked a form-changing target will not be without some error. But so long as the errors are relatively infrequent and the bias in the use of the target versus distractor categories remains relatively fixed over the conditions of the experiment, this method may provide a useful way of restricting the trials used to compute the RT measure to those which the participant has actually tracked the form-changing target.

In summary, we can state our specific predictions as follows: (a) If participants are more likely to lose target objects as the number of distractors in the display increases, then target categorization

errors (incorrectly attributing target form changes to distractors) should increase with increases in the number of distractors; (b) Response latencies to form-changing targets that are no longer tracked (and thus not indexed) should be slower than those for accurately tracked targets; (c) When the accuracy of the participants' categorization decisions (target or distractor form change) is ignored, response latencies to both target and distractor form changes should increase with increases in the number of distractors in the display (as in Experiment 1); and (d) Excluding those trials where participants had lost a form-changing target should greatly attenuate the effect of display-size on identification RT.

Method

Participants. Twenty-four individuals participated in this experiment, and were paid \$10.00 upon completion of the 45 minute session. None of these individuals had participated in Experiment 1.

Apparatus and stimuli. The apparatus and stimuli were identical to those of Experiment 1.

Procedure. Participants tracked four targets during every trial, and after 5 to 9 s of tracking one of the objects was transformed into an E or an H figure. The time to identify this form change was measured. The probability that a target or a distractor would undergo a form change was identical (50%), and the participants were informed of these probabilities.

After the participant had responded to the form change, the screen was cleared and two boxes labeled "Tracker" (the target) and "Non-Tracker" (the distractor) appeared. The participant then indicated what type of form change had occurred (target or distractor form change) by moving a mouse pointer into the appropriate box and then pressing a mouse button. This categorization decision was made on every trial, and participants were instructed to guess when they were unsure of the correct response.

Design. Two factors were manipulated: the Type of Form Change (target or distractor object) and the Number of Distractors in the display (4, 8, or 12). There were 30 trials in each of the six conditions of this experiment.

Results and Discussion

The mean identification error rate (E or H) across all the conditions was 3.9%, and a two-factor repeated measures ANOVA (Type of Form Change x Number of Distractors) yielded no significant effects, all F 's < 1. Response latencies greater than 3 seconds were classified as outliers and were removed from the dataset. This procedure resulted in the removal of 3.1% of the data.

Identification data. Before looking at the categorization data, we replicated the analysis that

was done in the first study. In this analysis the accuracy of the participants' categorization decisions (i.e., attributing the form changes to target or distractor objects) was ignored when computing average response latencies. Thus, the response latencies to form-changing targets that were no longer being tracked (as indexed by target categorization errors) were averaged with the response latencies to targets that had been accurately tracked. This treatment of the data should reproduce the pattern of effects observed in Experiment 1: Responses to target form changes should be faster than responses to distractor form changes, and response latencies to both target and distractor form changes should increase with increases in the number of distractors.

The results of this analysis were in fact identical to those of Experiment 1. The mean response latencies and error rates are listed in Table 2. A two-factor (Type of Form Change x Number of Distractors) repeated measures ANOVA revealed a significant main effect for the type of form change, $F(1, 23) = 46.52$, $MSE = 34,500$, and the number of distractors, $F(2, 46) = 9.48$, $MSE = 9,920$, but no interaction between these two factors, $F(2, 46) = 2.28$, $MSE = 7,383$. Participants identified target form changes an average of 211 ms faster than distractor form changes, and response latencies to both target and distractor form changes increased with increases in the number of distractors. Note that response latencies were slower in this experiment than in Experiment 1, presumably because participants had to determine the identity of the changed object (E or H) as well as the type of changed object (target or distractor) prior to responding.

Categorization data. Earlier we argued that target objects are not always perfectly tracked during the tracking task, and that the probability of losing one or more targets increases with increases in the number of distractors. This assumption can now be directly tested. Because participants made a categorization decision on every trial to indicate whether a target or a distractor underwent a form change, the trials in which participants accurately tracked form-changing target objects could be distinguished from the trials in which they did not. For example, if a target object underwent a form change and the participant categorized the event as a "distractor" form change, then the participant had not accurately tracked this particular target throughout the trial. Responses in which the participant correctly identified the form change (E or H) but incorrectly categorized the type of form change (target or distractor) were classified as *categorization errors*.

Table 3 lists the percentage of target and distractor categorization errors as a function of the number of distractors in the display. As measured by the categorization data, tracking performance (100% – the percentage of categorization errors) was reliably greater than chance (50%) in each of the six conditions of this experiment (all p 's < .05), although it was lower than that observed in Pylyshyn and Storm's (1989) experiment (particularly when there were more than four distractors).

A two-factor (Type of Form Change x Number of Distractors) repeated measures ANOVA indicated that both target categorization errors and overall categorization errors increased with increasing number of distractors. There was a main effect for the type of form change (target or distractor object), $F(1, 23) = 5.36$, $MSE = 546$, as categorization errors were more frequent for target form changes. The main effect for the number of distractors (4, 8, or 12) was also significant, $F(2, 46) = 45.53$, $MSE = 54$. Importantly, the interaction between these two factors was significant, $F(2, 46) = 15.18$, $MSE = 97$. The source of this interaction is clear. As the number of distractors in the display increased, the number of target form changes incorrectly attributed to distractor objects increased, whereas the number of distractor form changes incorrectly attributed to target objects was relatively constant. These results confirm that participants were more likely to lose target objects in the tracking task as the number of distractors in the display increased.

According to our hypothesis, if the index bound to a particular target object is lost or shifted (producing a target categorization error), and the same target subsequently undergoes a form change, then the time to identify this form change will be longer relative to when the target is accurately indexed and tracked throughout the trial. Consequently, we would expect that the response latencies associated with target categorization errors (incorrectly attributing a target form change to a distractor) would be slower than the response latencies associated with correct target categorizations (correctly attributing a target form change to a target).

To test this prediction, we compared the response latencies associated with target categorization errors to those associated with correct target categorizations. These data were submitted to a one-factor (Trial Type: incorrect target categorization versus correct target categorization) repeated measures ANOVA.² The effect of trial type was significant, $F(1, 20) = 18.68$, $MSE = 29,148$; the

Table 2

Mean Response Latencies (in Milliseconds) and Identification Error Rates (in %) to Target and Distractor Form Changes in Experiment 2

Type of Form Change	Number of Distractors		
	4	8	12
Target			
<i>M</i>	1410	1481	1469
<i>SE</i>	65.1	62.1	59.7
Errors	4.4	4.1	3.6
Distractor			
<i>M</i>	1609	1662	1722
<i>SE</i>	64.7	67.4	65.6
Errors	3.1	4.3	3.8

response latencies associated with target categorization errors (1669 ms) were slower than the response latencies associated with correct target categorizations (1441 ms). Thus, when participants were no longer tracking a particular form-changing target (as indicated by a target categorization error), responses to this target were delayed.

Moreover, note that if the indexes bound to target objects are lost or are shifted to distractors, then the speed of responding to these non-indexed target form changes should be similar to the speed of responding to distractor form changes. Thus, the response latencies to target form changes that were incorrectly attributed to distractors should be very similar to the response latencies to distractor form changes that were correctly attributed to distractors.

To test this prediction, we compared the response latencies of trials where participants had made a target categorization error (attributing a target form change to a distractor) to those trials where the participants had correctly attributed a form change to a distractor. These data were submitted to one-factor (Trial Type: incorrect target categorization versus correct distractor categorization) repeated measures ANOVA. This analysis yielded no significant effect of trial type, $F < 1$. In fact, there was only a 30 ms difference between the response latencies associated with target categorization errors (1669 ms) and the response latencies associated with correctly categorized distractor form changes (1699 ms). This suggests that targets that were no longer indexed (and tracked) were essentially “distractors” to the participant, as they were identified as being distractors and were responded to no faster than “true” distractor form changes.

Identification and categorization data. Recall

that the pattern of categorization errors confirmed that participants were more likely to lose target objects as the number of distractors in the display increased. This is precisely what one would predict given the hypothesis that increases in the number of distractors increases the probability that indexes bound to targets will be lost or shifted to distractor objects. What remains to be shown, however, is that the loss of indexes bound to form-changing target objects produced the display size effect observed in Experiment 1.

Because participants had to indicate the type of form change on each trial (target or distractor), the trials where participants had lost a form-changing target could be distinguished from the trials where the form-changing target object was accurately tracked. According to our hypothesis, excluding those trials where participants had lost a form-changing target from the RT analyses should greatly attenuate the display size effect.

The response latencies on trials where the participants correctly identified (E or H) and then correctly categorized the type of form change (target or distractor) were submitted to a 2 (Type of Form Change) x 3 (Number of Distractors) repeated measures ANOVA. The results of this analysis indicated that when a form-changing target object was accurately tracked, the time to respond to this form change did not significantly vary with increases in the number of distractors. The data are listed in Table 4. The main effect for the type of form change (target or distractor) was significant, $F(1, 23) = 50.11$, $MSE = 59,042$. Responses to target form changes were an average of 287 ms faster than responses to distractor form changes. The main effect for the number of

Table 3
Mean Percentages of Target and Distractor Categorization Errors in Experiment 2.

Type of Form Change	Number of Distractors		
	4	8	12
Target			
<i>M</i>	14.9	33.6	39.4
<i>SE</i>	2.9	3.8	4.2
Distractor			
<i>M</i>	18.4	20.8	21.6
<i>SE</i>	2.6	2.6	3.6

distractors (4, 8, or 12) was marginally significant, $F(2, 46) = 3.40$, $MSE = 14,527$, $p = .064$. Finally, the interaction between the type of form change and the number of distractors was significant, $F(2, 46) = 3.61$, $MSE = 8,946$. An inspection of Table 4 suggests that there was a display size effect for correctly categorized distractor form changes, but that response latencies to correctly categorized target form changes appeared to be relatively unaffected by increases in the number of distractors. Simple main effects corroborated this interpretation. Response latencies to distractor form changes increased with increases in the number of distractor objects, $F(2, 46) = 6.25$, $MSE = 12,579$, whereas response latencies to target form changes were not significantly affected by the number of distractors in the display, $F < 1$.

Summary. The conclusions one can draw from these analyses of the identification, categorization, and joint identification/categorization data are clear. Tracking performance deteriorated as the number of distractors in the display increased (categorization data), and when form changes took place on targets no longer being tracked (and indexed), the response latencies to these form changes were longer than those for targets that were accurately tracked. When these two types of trials were averaged together (identification data), the net result was an increase in response latencies to target form changes with increases in the number of distractors. However, when the trials in which participants had lost a form-changing target were excluded from the response latency data (joint identification and categorization data), response latencies to target form changes appeared to be independent of the number of distractors in the display. Thus, it seems reasonable to conclude that the display size effect for target form changes observed in Experiment 1 was the consequence of failures in tracking performance.

Distractor categorization errors. The analy-

ses so far appear to establish that when objects are being correctly tracked, the latency to identify a form change on a tracked object is not influenced by the number of distractor objects around it, and that the identification of form changes on target objects that have lost indexes fares no better than on distractor objects. But what about the distractor objects that are erroneously classified as targets? Are distractors misidentified as targets treated as if they were targets, or is the misidentification itself somehow noticed? The prediction in this case is not as clear because it is not known whether an index can simply shift to another object or whether it is merely lost and then an alternative target for it is sought.

As a preliminary exploration of these questions, the response latencies on trials where participants incorrectly categorized distractor form changes as target form changes were compared to those trials where participants correctly categorized target form changes. If indexes merely shifted unnoticed from target to distractor objects, resulting in the distractor in question being treated as a target, then the response latencies associated with these incorrectly categorized objects should be similar to those associated with correctly categorized targets. That is, the presence of an index bound to a distractor would convey the same attention-priority benefits as an index bound to a target.

A one-factor repeated measures ANOVA (Trial Type: incorrect distractor categorization versus correct target categorization) revealed a statistically reliable difference between the response latencies associated with distractor categorization errors and those associated with correct target categorizations, $F(1, 20) = 12.43$, $MSE = 86,031$. That is, the response latencies associated with distractor categorization errors (1760 ms) were

Table 4

Mean Response Latencies (in Milliseconds) to Correctly Identified and Correctly Categorized Target and Distractor Form Changes in Experiment 2.

Type of Form Change	Number of Distractors		
	4	8	12
Target			
<i>M</i>	1389	1411	1396
<i>SE</i>	64.5	63.4	63.8
Distractor			
<i>M</i>	1621	1702	1732
<i>SE</i>	63.4	66.5	68.1

significantly slower than the response latencies associated with correct target categorizations (1441 ms). Consequently, it would appear that indexes did not merely shift unnoticed from target to distractor objects. (Also note that the response latencies associated with distractor categorization errors were similar to the response latencies associated with correct distractor categorizations; 1760 ms versus 1699 ms, respectively, $F < 1$.) This finding was not anticipated by the current model and we can only speculate on what goes on in these cases. There are, however, independent reasons for thinking that the loss of an index may be detected (as hypothesized in the error-recovery model described in Pylyshyn et. al. 1994). Thus, a reassigned index might be treated somewhat differently from a correctly tracked one by the decision stage of the tracking process. Perhaps the uncertainty associated with the detection of the loss of an index results in an increase in reaction time.

Analysis of the region-bounded attention assumption. Finally, as in Experiment 1, we evaluated the possibility that the allocation of a spreading region of focused attention (as in the “zoom-lens” model) could account for the facilitation in responding to target form changes. For all twenty-four participants, the position of form-changing distractors at the time of a response was recorded, and the trials in which a distractor form change occurred inside the convex hull of the target set were compared to the trials in which a distractor form change occurred outside of this region. A zoom-lens view would predict that distractor form changes occurring within this region would be identified more rapidly than distractor form changes occurring outside of this region. However, as before, response latencies to distractor form changes were similar regardless of whether the distractors were inside (1666 ms) or outside (1662 ms) of the region encompassed by the targets, $F <$

1. This result indicates that the allocation of a zoom-lens of focused attention was not responsible for the facilitation in responding to target form changes, because the identification of distractor form changes was not enhanced in the region encompassed by the targets.

General Discussion

The purpose of this investigation was to study the relation between visual indexing and attentional processing in the context of the multiple-object tracking paradigm. Specifically, we hypothesized that target objects being tracked (and thus indexed) would be assigned a higher priority for access to serial focused attention, so that target form changes would be identified more rapidly than distractor form changes. The results of Experiment 1 clearly demonstrated that during the tracking task target form changes are identified more rapidly than distractor form changes. However, we also found that responses to target form changes were increasingly delayed as the number of distractors in the display increased, a finding not consistent with the visual indexing model. That is, if the target objects are independently tracked and used to determine the responses, then the speed of responding to target form changes should not be affected by the number of distractors in the display.

In Experiment 2, we evaluated the possibility that failures in tracking performance were responsible for this display size effect. We hypothesized that on some trials participants had not accurately tracked the form-changing targets, and that their responses to these “lost” targets would be delayed. We reasoned that if participants indicated the type of form change on every trial (target or distractor), then the trials where participants had accurately tracked a form-changing target could be distin-

gushed from the trials in which they had not. Excluding those trials where participants had lost a form-changing target from the RT analysis would then eliminate or greatly attenuate this display size effect.

The results of Experiment 2 confirmed these predictions. Tracking performance deteriorated as the number of distractors in the display increased, and responses to form-changing targets that were not accurately tracked were slower than response to targets that were accurately tracked. When only the trials in which participants had correctly tracked the form-changing targets were examined, the response latencies to target form changes were unaffected by the number of distractors in the display.

Finally, our experiments have shown that the facilitation afforded by the indexes applies only to the actual targets and does not spread to the regions between them, so the results cannot be explained by the assumption that attention is allocated broadly to the region encompassed by the target objects (as in the “zoom-lens” model). Responses to distractor objects undergoing form changes within the convex hull of the target set were no faster than responses to distractors undergoing form changes outside of this region.

These findings are consistent with Pylyshyn’s visual indexing model (Pylyshyn, 1989, 1994, 1998). According to that model, certain salient properties – a sudden onset being the most obvious such property – result in an index or pointer being assigned to the objects. There is a limit of 4 or 5 such indexes available. The indexes are object-based in that they do not point to the locations of objects but to the objects themselves, and, as a result, indexes keep referencing objects as the latter move around. The purpose of indexes is to bind objects to internal references or names. Just as the arguments of a function have to be bound before the function can be evaluated, so cognitive processes cannot be applied to visual objects unless the objects are bound to the arguments of the processes. For example, in order to evaluate the n -ary predicate “collinear”, the objects over which the predicate applies must be indexed or bound to arguments of the predicate.

The multiple object tracking task is one of the purest and most direct tests of the assumptions of this model. This is because, as Pylyshyn and Storm (1988) showed, tracking multiple independently-moving identical objects could not be done in their study by recording and updating an encoding of the locations of the objects being tracked, so a more direct tracking mechanism must be available. In the present study a number of additional

predictions of this model were examined. The first prediction was that tracked objects can be examined more quickly than non-tracked objects in a task that requires focal attention (i.e., form identification). The second prediction was that tracked objects can be accessed more quickly than non-tracked objects even when the latter are located in the same region of the display as the tracked objects (contrary to what might be expected from the “zoom lens” model of attention allocation). The third prediction was that the presence of different numbers of non-tracked objects does not directly affect the speed of accessing and processing of tracked objects. All of these predictions were verified, although the last one required that we use a dual-response methodology that allowed us to remove mistracked objects from the pool used to compute response latencies.

Other Models of Multiple Parallel Access in Visual Attention

At least one other model has attempted to account for the pattern of reaction times in a situation similar to the one we used. As noted, Yantis and Johnson (1990) found that a limited number of abrupt onset items could be processed before non-onset items during visual search, and they proposed that abrupt onsets are attentionally “tagged” and then assigned a higher priority for access to focused attention. This tagging model is compatible with our proposal, although we have emphasized the dynamic changing location and tracking aspect of the access and have proposed a different mechanism to explain the results. Instead of appealing to the existence of a “tag” located on a stimulus object, we have proposed a mechanism that is part of the early vision system (in the sense of Pylyshyn, in press-a). We have argued that such a mechanism is independently required for other early-vision purposes (such as to enable vision to be “situated” and causally connected in the right way with objects in the world – Pylyshyn, in press-b). Although we have not attempted directly to compare tagging models (see also Watson and Humphreys, 1997; Ullman, 1994) with the visual indexing model, we note that the tagging models make no predictions about the ability to track multiple objects nor about the locality property of the access provided by tracked objects. Also, it appears that tags have a rather short-lived effect (Yantis & Jones, 1991), whereas our tracking studies have run as long as 30 seconds. And finally, tag models assume that there is some place where the tags actually reside – because tags cannot actually be written on the world, presumably they are writ-

ten into some iconic store which itself can be scanned. Because little is known about such a store, and the very idea of an iconic store presents serious empirical problems (cf., O'Regan, 1992; Pylyshyn, in press-b), the visual indexing model offers some advantages as a basis for understanding the phenomenon of priority access.

Perhaps closer in spirit to the indexing model is the object file theory of Kahneman, Treisman, and Gibbs (1992). Like visual indexes, object files are object-centered and continue to reference objects across changes in spatial position. However, object files are memory organizations and the theory does not deal with the mechanism by which object files are attached to the objects with which they are associated, and it makes no predictions concerning the time for accessing tracked objects. In contrast, although Yantis and colleagues (Jonides & Yantis, 1988, Yantis & Jones, 1991; Yantis & Jonides, 1984) do deal with the access issue, they have used static displays to study the attentional priority conferred to abrupt onset items, and to our knowledge have not explored the possibility that attentional tags are object-centered and apply to moving objects. Consequently, the visual indexing model (which was developed independently in a very different context – Pylyshyn, Elcock, Marmor, and Sander, 1978) augments these existing models and makes specific predictions about the attentional priority conferred to a number of objects under dynamic display conditions. Moreover, it has been applied to a wide range of novel visually related phenomena, from subitizing to mental imagery and visual stability (Pylyshyn, 1998).

References

- Atkinson, J., & Braddick, O. J. (1989). "Where" and "what" in visual search. *Perception*, *18*, 181-189.
- Baylis, G. C., & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 451-470.
- Burkell, J. A., & Pylyshyn, Z. W. (1997). Searching through subsets: A test of the visual indexing hypothesis. *Spatial Vision*, *11*, 225-258.
- Castiello, U., & Umiltà, C. (1990). Size of the attentional focus and efficiency of processing. *Acta Psychologica*, *73*, 195-209.
- Cavanagh, P. (in press) Attention: Exporting vision to the mind. In C. Taddei-Ferretti (Ed.), *Neuronal basis and psychological aspects of consciousness*. Singapore: World Scientific.
- Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. H. (in press). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*.
- Danzig, P. (1988). A programmable timer for the Sun workstation. Available from Peter Danzig, University of Alberta, Canada.
- Downing, C. J. (1988). Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 188-202.
- Eriksen, C., & Hoffman, J. (1974). The extent of processing of noise elements during selective encoding of visual displays. *Perception & Psychophysics*, *12*, 201-204.
- Eriksen, C., & St. James, J. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*, 225-240.
- Eriksen, C., & Yeh, Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 583-597.
- Hawkins, H. L., Hillyard, S. A., Luck, S. J., Mouloua, M., Downing, C. J., & Woodward, D. P. (1990). Visual attention modulates signal detectability. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 802-811.
- Intriligator, J. M. (1997). *The spatial resolution of attention*. Unpublished Ph.D. Dissertation, Department of Psychology, Harvard University, Cambridge, MA.
- Intriligator, J., & Cavanagh, P. (1992). *An object-specific attentional facilitation that does not spread to adjacent spatial locations*. Poster presented at the annual meeting of the Association for Research in Vision and Ophthalmology, Sarasota, Florida.
- Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, *43*, 346-354.
- Juola, J. F., Bouwhuis, D. G., Cooper, E. E., & Warner, C. B. (1991). Control of attention around the fovea. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 125-141.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175-219.
- McKeever, P., & Pylyshyn, Z. W. (1993). *Nontarget numerosity and identity maintenance with FINSTs: A two component account of multiple target tracking*. Centre for Cognitive Science, University of Western Ontario, Technical Report: Cogmem 65.
- O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*, 461-488.
- Pashler, H. E. (1998). *The Psychology of Attention*. Cambridge, MA: MIT Press.
- Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and Performance X* (pp. 531-556). Hillsdale, NJ: Erlbaum.
- Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*, 160-174.
- Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial indexing model. *Cognition*, *32*, 65-97.

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. (in press-a). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*.

Pylyshyn, Z. W. (in press-b). Connecting vision and the word: Tracking the missing link. In A. Marques, B. J., & J. Saagua (Eds.), *The Foundations of Cognitive Science*. Oxford, UK: Oxford University Press.

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

Pylyshyn, Z. W., Elcock, E. W., Marmor, M., & Sander, P. (1978). *Explorations in visual-motor spaces*. Paper presented at the Proceedings of the Second International Conference of the Canadian Society for Computational Studies of Intelligence, University of Toronto.

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

Scholl, B. J., & Pylyshyn, Z. W. (in press). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*.

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

Tsal, Y., & Lavie, N. (1993). Location dominance in attending to color and shape. *Journal of Experimental Psychology*, 19, 131-139.

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

Viswanathan, L., & Mingolla, E. (1998). *Attention in depth: disparity and occlusion cues facilitate multi-element visual tracking*. Poster presented at the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL.

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, 90-122.

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

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

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

Yantis, S., & Jones, E. (1991). Mechanisms of attentional selection: Temporally modulated priority tags. *Perception & Psychophysics*, 50, 166-178.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 601-621.

Authors' Note

This work was supported by grants to ZWP from the Natural Sciences and Engineering Research Council, from the Institute for Robotics and Intelligent Systems (IRIS Project B-4), and from MIH grant 1R01 MH60924 to ZWP.

We thank Peter Keller for his programming assistance. We also thank Jacquie Burkell, Roy Eagleson, Brian Fisher, Rick Gurnsey, Stephen Lupker, and Paul Mckeever for their many useful comments on previous versions of this paper, and to Pierre Jolicoeur and Steven Yantis for their very helpful reviews. Correspondence should be addressed to Christopher R. Sears, Department of Psychology, University of Calgary, 2500 University Drive N.W., Calgary, Alberta, Canada, T2N 1N4, or to Zenon W. Pylyshyn, Rutgers Center for Cognitive Science, Psychology Annex, Rutgers University, New Brunswick, NJ 08905. Email inquiries may be sent to either sears@ucalgary.ca or zenon@rucss.rutgers.edu.

Footnotes

1. Because the data from only six of the seventeen participants in this experiment were available for this analysis, one could argue that the lack of an effect of distractor position (inside or outside the convex hull of the target set) was due to a reduction in statistical power. However, the data from the same six participants indicated that target form changes were responded to more rapidly than distractor form changes, $F(1, 5) = 14.19$, $MSE = 22,103$. Moreover, there was no effect of distractor position in Experiment 2 either, where the data from twenty-four participants were examined.

2. Three of the twenty-four participants made no target categorization errors in one or more of the conditions in this experiment, and so all of the response latency analyses involving target categorization errors are based on the data from the remaining twenty-one participants.