This paper appears (with neater figures) as: Cognitive Psychology, **38**, 259 - 290. (1999)

Tracking Multiple Items Through Occlusion: Clues to Visual Objecthood

Brian J. Scholl and Zenon W. Pylyshyn

Rutgers Center for Cognitive Science

In three experiments, subjects attempted to track multiple items as they moved independently and unpredictably about a display. Performance was not impaired when the items were briefly (but completely) occluded at various times during their motion, suggesting that occlusion is taken into account when computing enduring perceptual objecthood. Unimpaired performance required the presence of accretion and deletion cues along fixed contours at the occluding boundaries. Performance was impaired when items were present on the visual field at the same times and to the same degrees as in the occlusion conditions, but disappeared and reappeared in ways which did not implicate the presence of occluding surfaces (e.g. by imploding and exploding into and out of existence, instead of accreting and deleting along a fixed contour). Unimpaired performance did *not* require visible occluders (i.e. Michotte's *tunnel effect*) or globally consistent occluder positions. We discuss implications of these results for theories of objecthood in visual attention.

What is an object? This is a question which draws together researchers from many corners of cognitive science, and which can be asked relative to many different cognitive and perceptual processes. The experiments reported in this paper address this question from the perspective of dynamic visual attention.

A wide variety of studies (e.g. Burkell & Pylyshyn, 1997; Pylyshyn, 1989, 1994; Pylyshyn et al., 1994) have shown that the visual system can have simultaneous access to several discrete visual objects in its field of view. The process of providing such access, called *indexing*, has been shown to be *object-based*, meaning that the indexes remain attached to particular visual objects independent of the objects' movements and such properties as their color, shape, or position. What

---

We thank James Intriligator and Raynald Comtois for programming advice and software support during the early stages of this project, Tanya Edwards for assistance with data collection, and the members of the Rutgers / Princeton / NYU Object Group for helpful suggestions regarding this work. Special thanks to Alan Leslie, Anne Treisman, and two anonymous reviewers for comments on earlier drafts. Some of these results were presented at the 1998 meeting of the *Association for Research in Vision and Ophthalmology* and in a Master's thesis at Rutgers University by the first author.

Address correspondence and reprint requests to Brian J. Scholl, Center for Cognitive Science, Psychology Bldg. Addition, Busch Campus, Rutgers University - New Brunswick, 152 Frelinghuysen Road, Piscataway, NJ 08854-8020. Email: scholl@ruccs.rutgers.edu.

does it mean for a process to be *object-based*? Although the general notion of an 'object' (e.g. Hirsch, 1982; Wiggins, 1980) is beyond the scope of this paper, there is a notion of a *visual object* that has been widely used to refer to visually-primitive punctate spatio-temporal clusters. It is roughly this notion that we have in mind when we say that indexes are 'object-based'.

In this paper we investigate some of the spatiotemporal constraints which the visual system uses to compute enduring visual objecthood. For certain parts of the visual system, visual *objects* are parts of the visual field which respect certain dynamic spatiotemporal constraints — in particular constraints relating to *occlusion*. This means, for example, that objects can be continuously tracked through spatiotemporal gaps only of certain types and certain magnitudes. Such spatiotemporal parameters in effect define what it means to be a persisting visual object. This has important implications for several theories of visual attention, which are examined in the General Discussion. We will motivate our experiments in the context of Pylyshyn's *FINST* theory of visual indexing (e.g. Pylyshyn, 1989, 1994; Pylyshyn et al., 1994), which has previously been supported by *multiple-object tracking* experiments similar to those reported in this paper.

The FINST Visual Indexing Theory

Visual attention is associated with a limitation on our capacity to process visual information. Although we know that this limitation exists, it is not clear at the outset what the correct *units* are for characterizing it. It was traditionally assumed that attention simply restricts various types of visual processing to certain *spatial areas* of the visual field — e.g. in the manner of a spotlight (e.g. Eriksen & Hoffman, 1972; Posner, Snyder, & Davidson, 1980) or a zoom-lens (e.g. Eriksen & St. James, 1986). It has recently been demonstrated, however, that there must also be an *object-based* component to visual attention, in which attentional limitations are characterized in terms of the number of discrete *objects* which can be simultaneously processed (for reviews, see Egeth & Yantis, 1997, and Kanwisher & Driver, 1992; cf. Scholl & Leslie, in press).

Several recent theories have been concerned with how these visual objects are individuated, accessed, and used as the basis for memory retrieval. These theories include Kahneman and Treisman's *Object File* theory (Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992), Yantis' theory of *Attentional Priority Tags* (Yantis & Johnson, 1990; Yantis & Jones, 1991), the notion of *Object Tokens* (Chun & Cavanagh, 1997; Kanwisher, 1987), and Wolfe's theory of *Preattentive Object Files* (Wolfe & Bennett, 1997). Pylyshyn's 'FINST' theory of visual indexing (e.g. Pylyshyn, 1989, 1994) complements these other theories by postulating a mechanism by which object-based individuation, tracking, and access is realized.

In order to detect even simple geometrical properties among the elements of a visual scene (e.g. being inside, or being collinear), Pylyshyn (1989, 1994) argues that the visual system must be able to simultaneously reference — or 'index' — multiple objects. This is exactly analogous to the requirement that all variables in a function be bound before a function can be evaluated, and indeed indexes can be viewed as a variable-binding mechanism for perceptual/motor predicates or procedures. They bind internal variables to objects to be evaluated or acted upon.

Similarly, although focal attention may be scanned in a prescribed direction until it finds objects, it cannot orient directly to a particular object unless that object has already been indexed. Pylyshyn's model is based on *visual indexes* which can be both exogenously and endogenously assigned to various items in the visual field, and serve as a means of *access* to those items for higher-level processes that allocate focal attention. In this regard, they function rather like pointers in a computer data structure: they reference certain items in the visual field (identifying them *as* distinct objects), without themselves directly revealing properties of those objects.

(The visual indexes in Pylyshyn's theory were historically called *FINSTs*, for *fing*ers of *inst*antiation, due to the fact that physical fingers work in an analogous way: "Even if you do not know anything at all about *what* is located at the places that your fingers are touching, you are still in a position to determine such things as whether the object that finger number 1 is touching is to the left of or above the object that finger number 2 is touching. . . . [T]he access that the finger contact gives makes it inherently possible to track a *particular* token, that is, to keep referring to what is, in virtue of its historical trace, the *same* object" [Pylyshyn, 1989, p. 68].)

Visual indexes can be assigned to objects in the visual field regardless of their spatial contiguity (in contrast with the assumptions of a spotlight model), with the following restriction: the architecture of the visual system provides only about *four* indexes. Furthermore, the indexes are *sticky*: if an indexed item in the visual field moves, the index moves with it. The visual indexes bestow a processing priority to the indexed items, insofar as they allow focal attention to be shifted to indexed (and possibly moving) items without first searching for them by scanning through the intervening space. (Note that the visual indexing theory thus *complements* — rather than competes with — theories that posit a single locus of focal attention; cf. Pylyshyn & Storm, 1988, p. 180).

Several different experimental tasks have been used to adduce support for the indexing framework, including evidence from *multiple-object tracking* (Pylyshyn & Storm, 1988; Sears & Pylyshyn, in press), *subitizing* (Trick & Pylyshyn, 1993, 1994), *visual search* (Burkell & Pylyshyn, 1997), and the *line-motion illusion* (Schmidt, Fisher, & Pylyshyn, 1998). For concise reviews of this experimental support, see Pylyshyn (1994) and Pylyshyn et al. (1994). The current experiments employ the multiple-object tracking paradigm, to which we now turn.

Multiple Object Tracking

Pylyshyn & Storm (1988) introduced the *multiple-object tracking* paradigm as a direct test of the visual indexing theory. Subjects in their first experiment viewed a display initially consisting of a field of identical white crosses (each subtending 0.42 deg). A certain subset of the crosses were then flashed several times to indicate their status as targets. All of the crosses then began moving independently and unpredictably about the screen, constrained only so that they could not pass nearer than 0.75 deg to each other, and could not move off the display. (All trajectories were generated offline in advance. When trajectories passed too near to each other, the last few frames of random motion were discarded and recalculated. Crosses whose trajectories carried them to the edge of the display were reflected off that edge.) This motion continued for an interval ranging from 7 to 15 s. At predetermined times, a small square was flashed on the display, and subjects pressed keys to indicate whether the square had been flashed at the location of a target, a non-target, or neither. Note that since all the crosses were identical during the motion interval, subjects could only succeed by picking out (or, in our terms, *indexing*) the targets when they were initially flashed, and then tracking them through the motion interval. See Figure 1 for a schematic representation of the multiple-object tracking task.

Subjects were successful (never less than 85.6% accurate) in these experiments when tracking up to five targets in a field of ten identical independently and unpredictably moving items. Pylyshyn and Storm ruled out a class of alternate explanations for this result in which a single spotlight of attention sequentially and repeatedly visits each item in turn: even at the fastest reported scan velocities (250 deg/s), a simulated attentional spotlight (augmented with several location-prediction and guessing heuristics) was unable to approach the actual performance of human subjects. Even allowing for a strategy wherein only a certain subset of the targets were tracked (and the rest simply guessed at), they calculated that at least *four* items would still have to be independently tracked in parallel. (See Pylyshyn & Storm, 1988, for the details of this simulation.)

Further studies support the view that it is the items *themselves* that are being indexed and tracked. Attention is typically thought to improve various sorts of low-level visual processing, for example enhancing acuity (e.g. Nakayama & Mackeben, 1989), and speeding response times to attended objects or areas (e.g. Downing & Pinker, 1985; Posner et al., 1980). Similarly, visual indexes bestow a processing advantage to the indexed items, since they can be immediately accessed by higher-level processes without a serial search. Intriligator (1997, Experiment 2; Intriligator & Cavanagh, 1992) and Sears and Pylyshyn (in press)

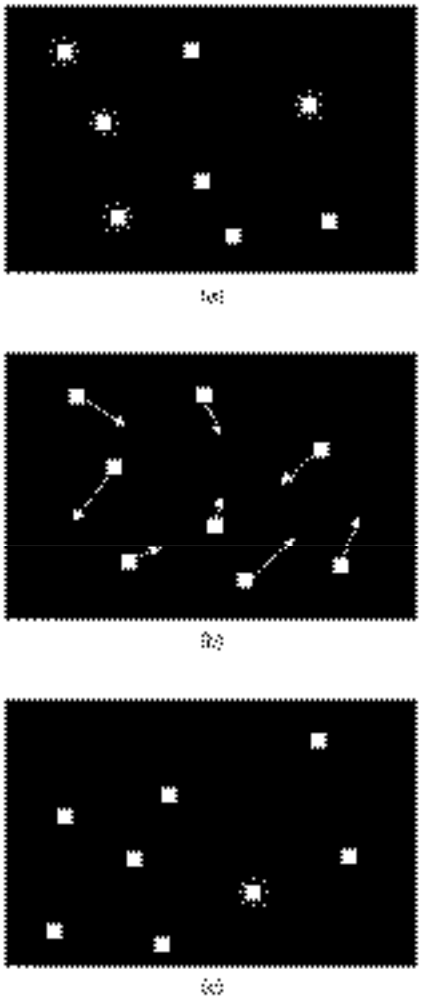

FIG. 1. A schematic depiction of the multiple-object tracking task (not to scale). A number of items are presented, and a subset of them are flashed several times to indicate their status as targets. All of the items then begin moving randomly and unpredictably about the screen. After a predetermined interval, the motion stops, and one of the items is flashed again several times to indicate its status as the probe. Subjects are to decide if the probe item is one of the target items, and respond appropriately.

explored this processing advantage in the context of multiple-object tracking, demonstrating that it is target-specific; in particular, it doesn't hold for non-targets — even those which are located within the convex polygon bounded by the moving targets. Thus, it must be the items *themselves* which are being indexed and tracked, and not the region of space in which they're located. (In Intriligator's terms, the processing advantage is *split* between these items rather than being *spread* between them.)

The crucial result of these experiments — that subjects can successfully track up to about five identical unpredictably moving targets in a field of about 10 total identical items — has since been replicated several times in multiple laboratories, using several different types of stimuli and motion trajectories (e.g. Intriligator, 1997; Sears & Pylyshyn, in press; Viswanathan & Mignolla, 1998; Yantis, 1992).**[1]**

We now motivate the present experiments by focusing on *how* the items in these experiments might be tracked by visual indexes.

How Are Multiple Items Tracked?: Objecthood and Occlusion

Pylyshyn and his colleagues have demonstrated that subjects are able to track up to five independently and unpredictably moving items in a field of identical distractors, and they have explained this ability by appeal to visual indexes, but they have not specified exactly *how* the indexes manage to track the moving objects (though see Acton & Eagleson, 1993, and Eagleson & Pylyshyn, 1988, for preliminary models of the indexing mechanism, as well as Koch & Ullman, 1985, for a related network implementation).

We prefer to view the 'stickiness' of the indexes as a side-effect of the way in which object individuation and indexing are implemented. If the individuation of visual objects is done by a local computation, as seems plausible (e.g. by something like the neural network model of Koch & Ullman, 1985), then continuous motion need not disturb the stability of the object-cluster formed in this way: the local salience-determining winner-take-all network will simply keep the cluster intact as it moves. This method might often be effective, since objects on the visual field *do* tend to move on spatiotemporally continuous paths — objects in naturalistic scenes, after all, cannot jump in and out of existence.

---

[1] Several other complexities of multiple-object tracking — such as the apparent fact that performance deteriorates with increasing numbers of distractors — are not pertinent to the current experiments, and are not reviewed here. See Pylyshyn (1994) and Sears and Pylyshyn (in press) for discussion of these complexities, and of how they can be accommodated by the FINST visual indexing theory. Yantis (1992) replicated the basic result of Pylyshyn & Storm (1988), and reported additional experiments supporting the view that subjects tracked the targets by perceptually grouping them into a coherent but nonrigid virtual shape, and then simply tracking deformations of this virtual shape. This notion is discussed in detail below, and has been controlled for in the current experiments.

The current project begins by noticing that *occlusion* is a salient exception to this regularity, and is thus a problem for this scheme. Objects may often be (partially or wholly) occluded by other objects as they move about on the visual field, and such occlusion will continuously disrupt the spatiotemporal continuity of items on the visual field. (A network model like that of Koch and Ullman, 1985, will lose the cluster if the object disappears from view, unless some additional process comes into play to compensate.) Yet occlusion does not disrupt the continuing identities of the objects in the scene, and thus should not disrupt perceptual objecthood. Occlusion should therefore not disrupt any indexing mechanism that is going to be useful for perceiving real-world scenes. As Nakayama and his colleagues have noted,

> [O]cclusion varies greatly, depending on seemingly arbitrary factors—the relative positions of the distant surface, the closer surface, the viewing eye. Yet, various aspects of visual perception remain remarkably unimpaired. Because animals, including ourselves, seem to see so well under such conditions and since this fact of occlusion is always with us, it would seem that many problems associated with occlusion would have been solved by visual systems throughout the course of evolution. (Nakayama & Shimojo, 1990, quoted in Nakayama, He, & Shimojo, 1995, p. 62)

These considerations suggest that if the indexing machinery uses spatiotemporal proximity to compute continuing objecthood (and thus to track multiple moving objects), it had better make allowances for occlusion: objects aren't 'allowed' to disappear and reappear, *unless* the presence of an occluder provides an explanation for this behavior.

Does the indexing system take occlusion into account? Below we report experiments suggesting that indeed it does: observers are able to track multiple items (maintaining their continuing objecthood) as they move about the visual field, despite the fact that the items frequently disappear behind occluders and are completely absent from the visual field until they emerge. Performance in such cases is found not to differ significantly from performance with identical trajectories but without occlusion, suggesting that the indexing system does indeed track items about the visual field by employing spatiotemporal constraints on objecthood that recognize and make allowances for occlusion.

EXPERIMENT 1

If the visual indexing system does indeed take occlusion into account, then performance on the multiple-object-tracking task should not be disrupted when there are occluders present on the display, behind which objects disappear for some period of time. In Experiment 1 we thus compared tracking performance (where subjects tracked four targets in a field of eight items total) both with and without occluders, predicting no difference in performance.

This alone, however, is not enough to show that the mechanisms responsible for tracking recognize and make allowances for occlusion. It could be argued, for instance, that the tracking mechanism is simply robust enough to withstand *any* modest interruptions in spatiotemporal continuity, without making allowances for occlusion per se. To rule out this alternate explanation, this experiment also included two control conditions which varied the character of the spatiotemporal interruptions. There were thus four conditions total, comprising the four values of our single independent variable of 'Occlusion Method'.

Condition 1: No Occlusion (Visible Outlines). With this baseline condition we simply attempt to replicate the result that subjects can successfully track four target items in a field of eight total items. To afford a more direct comparison with performance in the various control conditions, we included two stationary pseudo-occluders in the display, presented as outlined rectangles, with a width twice the size of the items themselves. Items in this condition, however, simply traveled right through the rectangular outlines. At the end of every trial, a single item was probed, and subjects were to indicate whether or not the probed item was one of the four targets. Based on earlier reports (e.g. Intriligator, 1997, Experiment 2; Pylyshyn & Storm, 1988; Yantis, 1992), we predict that there will be few if any errors in this condition. Performance in this condition will serve as a comparison for the occlusion condition as well as the various control conditions.

Condition 2: Occlusion. To test tracking performance under occlusion, we simply duplicated the 'No Occlusion (Visible Outlines)' condition, but now had the rectangular outlines actually act as occluding surfaces, such that the items disappeared and reappeared from 'behind' them. Items were thus absent from the display during their motion to the degree that they intersected one of these occluders. The accretion and deletion cues provided by the item motion provided a strong phenomenal percept of occlusion. We predict that performance in this condition will also be near ceiling, since we hypothesize that occlusion is recognized — and does not disrupt — the mechanisms responsible for tracking.

Condition 3: Instantaneous Disappearance/Reappearance. Since equivalent levels of performance in the above two conditions needn't reflect a tolerance for occlusion per se, we included control conditions in which the local details of the interactions between items and occluders were altered, in an attempt to zero in on the crucial factors of the occlusion events. In this condition, we start at the opposite extreme from the 'Occlusion' condition, removing various aspects of the occlusion events. First, the visible occluder outlines were removed from the display (although they were still functionally present). Second, items in this condition disappeared instantaneously whenever their positions intersected an occluder position (i.e. as soon as their leading edges encountered the occluder boundary), and then eventually reappeared instantaneously as soon as the positions of their trailing edges no longer intersected the occluder position. Since this form of sudden disappearance also fails to support the interpretation of an object disappearing behind an occluding surface, we predict that performance will also be significantly impaired in this condition, relative to the 'No Occlusion (Visible Outlines)' and 'Occlusion' conditions.

Condition 4: Implosion/Explosion. There are independent reasons (discussed below) why performance could be impaired in the 'Instantaneous Disappearance / Reappearance' condition, however, so here we tried a case which was more similar to the original 'Occlusion' condition. We again removed the visible occluder outlines (so that the occluders were functionally present but invisible), but disappearances and reappearances were again gradual. Instead of accreting and deleting along a fixed contour, as in the 'Occlusion' condition, however, items in this condition imploded and exploded into and out of existence from their *centers* when they reached the boundaries of the (invisible, 'virtual') occluders. This control preserved the local characteristics of occlusion — items were present on the display at the same times, and disappeared progressively at the same rates as in the 'Occlusion' condition. Yet this form of gradual disappearance and reappearance does not readily support the interpretation of an object going behind an occluding surface. As noted by J. J. Gibson,

> there are two quite different . . . ways in which an object may disappear and appear. It may *go out of sight* or *come into sight*, on the one hand, and it may *go out of existence* or *come into existence* on the other. (Gibson, Kaplan, Reynolds, & Wheeler, 1969)

Gibson here is referring to the profound *perceptual* difference that can result from objectively quite similar manners of disappearance. We predict that just such a profound difference will manifest itself in subjects' relative performance between these conditions, despite their local similarity. We hypothesize that performance will be near ceiling in the 'Occlusion' condition because the tracking system makes allowances for occlusion *qua* occlusion — and does not simply have a general robust tolerance for interruptions of continuity. We thus predict that performance in this condition (which does not support the interpretation of an occluding surface, as noted by Gibson) will be impaired relative to the 'No Occlusion (Visible Outlines)' and 'Occlusion' conditions.

Schematic 'snapshot' diagrams of these various conditions are presented in Figure 2.

To ensure that any differences in performance between these conditions are due to the 'occlusion method' manipulations, and not to other accidental differ-

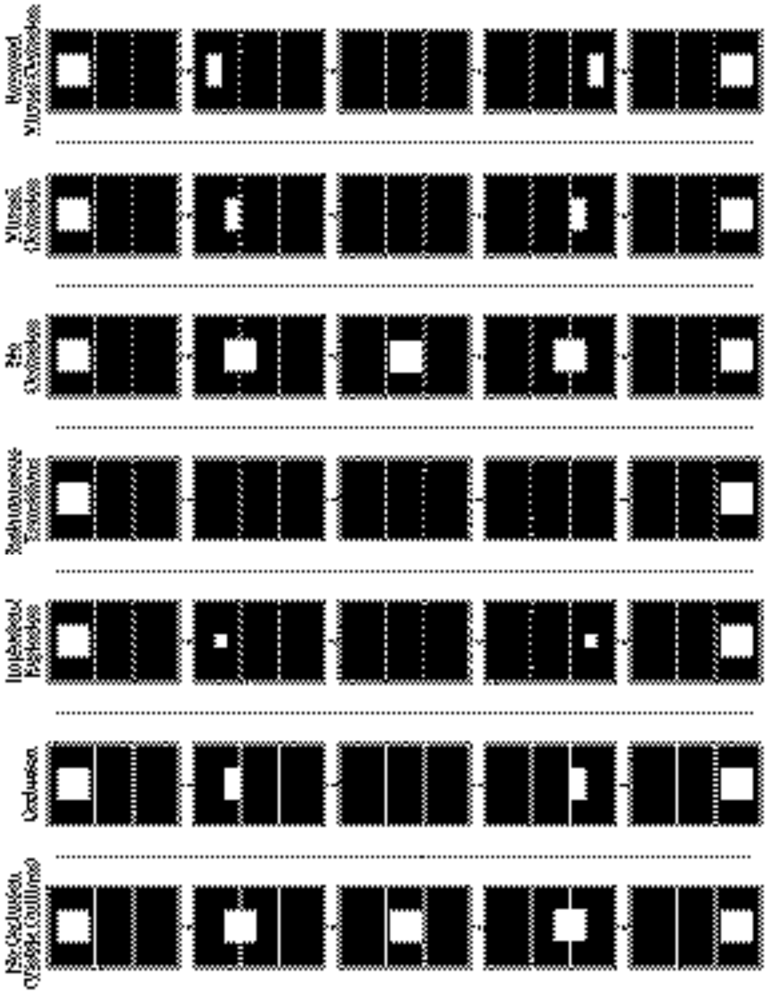

FIG. 2. The inherently dynamic nature of the occlusion conditions makes them difficult to represent in a static medium, but here we present the different occlusion conditions of from all Experiments as sequences of static ‘snapshot’ diagrams. In each condition, an item travels downward throughout five sequential frames of motion, interacting with a hypothetical occluder position (not to scale). Solid occluder boundaries represent visible occluders, while dashed occluder boundaries represent invisible occluders (presented to aid comprehension). See the text for detailed descriptions of each occlusion method. Note that the *Local Virtual Occlusion* condition of Experiment 3 is not depicted, since it cannot be distinguished here from the *Virtual Occlusion* condition.

ences between the trials (such as coincidental differences in inter-item spacing or trajectory patterns), we employed *identical trajectory sets, target selections, and probe selections* in each condition. The trials comprising each condition were thus identical in every way to the trials of the other conditions, excepting only the method of ‘occlusion’. (All comparisons were within-subjects. Trial order was randomized separately for each subject. Enough trials were included so that subjects did not notice the repeated trajectories in different conditions.) This manipulation rules out several other possible alternate explanations based on accidental differences between the trials for each condition.

One such alternate explanation, for instance, could involve Yantis’ (1992) notion of ‘virtual polygon’ tracking. In the original presentation of the FINST indexing theory, Pylyshyn (1989) noted that to track the items in a multiple-object tracking task, “The encoding of relative positions might be facilitated by noticing a pattern formed by the points, thereby ‘chunking’ the set in a single mnemonic pattern” (p. 81). Yantis (1992) expanded upon this point, suggesting that subjects track multiple items by perceptually grouping them into a ‘virtual polygon’ and then tracking deformations in that virtual shape. Yantis demonstrated that tracking performance is improved whenever experimental manipulations facilitate the formation of the ‘virtual target polygon (e.g. by explicitly describing the strategy to subjects, or by constraining the object motion such that the virtual target polygon never collapsed upon itself). Sears and Pylyshyn (in press) argue that this sort of perceptual grouping itself requires visual indexing, but for present purposes the point is that this sort of factor cannot be involved in any of the present results, since the identical trajectory sets were used in each condition. The ease with which ‘virtual target polygons’ could be formed and maintained may thus vary between the different trials, but does not vary between the different ‘occlusion method’ conditions.

In summary, we have two crucial predictions regarding this first experiment. First, tracking performance should not differ depending on whether or not items are occluded at the rectangle boundaries. Second, tracking performance should be disrupted by the ‘Implosion/Explosion’ and ‘Instantaneous Disappearance/Reappearance’ controls, in which items behave in ways that are locally similar to occlusion, without supporting the interpretation of an occluding surface.

Method

Participants. Fifteen Rutgers University undergraduates participated in one individual 40-min session to fulfill an introductory psychology course requirement. All subjects had normal or corrected-to-normal vision. One subject chose to terminate the experiment before completion, and was replaced.

Apparatus. The tracking displays were presented on a monitor controlled by a Macintosh Quadra microcomputer. Subjects were positioned with their heads in a chinrest 57 cm from the display monitor, the viewable extent of which subtended 30.52 by 22.89 deg. All software was written in the C programming language, using the VisionShell libraries of programming routines (Comtois, 1994).

Stimuli. Each trial employed four target items and four distractor items, each consisting of a solid white square subtending 0.48 deg. Initial item positions were generated randomly, with the constraint that each had to be at least 2.38 deg from the edge of the display monitor, at least 2.38 deg from each other, and at least 0.48 deg from the edge of an occluder position. On trials containing visible occluders (see below), two horizontal occluders were drawn on the initial display as outlined rectangles. Each occluder spanned the entire width of the screen (30.52 deg), and was 1 deg in vertical extent (slightly more than twice the extent of the items themselves). The lines comprising the outlined occluder rectangles were each 1 pixel (0.05 deg) wide, and clearly visible. Occluder positions were randomized for each trial, with the constraint that each had to be at least 4.77 deg from the top and bottom edges of the screen, at least 2.86 deg from the other occluder, and could not overlap the fixation. An empty framed square (0.38 deg) was present on every trial in the center of the display as the fixation.

A ten-second animation sequence was generated for every trial to produce unpredictable trajectories for each item, as follows. Items were each assigned initial random velocity vectors, which were updated on each frame. To prevent items from getting too close or intersecting one another, each item was repulsed(via an inverse-distance-squared ‘force field’) by each other item, and by the edges of the display. This repulsion affected the items’ velocities and directions, often causing approaching items to veer away from each other, or even reverse direction. The new position of an item on each frame was a function of the item’s current position, updated by (a) the item’s velocity vector, (b) the sum of the repulsion vectors at that item on that frame, and (c) a random adjustment to the item’s velocity vector. This method resulted in unpredictable motion trajectories. Trajectories were generated randomly, with the constraint that no item ever crossed the fixation, or ended in a position that intersected an occluder. The resulting set of trajectories for a trial (along with randomly selected target and probe items, as well as occluder positions) were stored offline as 335 static frames, to be presented for 30 ms each for a total of 10 s of motion.

In the resulting motion, items could move a maximum of 0.119 deg/frame. Since frames were displayed for 30 ms each, the resulting item velocities were in the range from 0 to 4 deg/s, with an average velocity (over all trials and items) of 3.26 deg/s. No item ever came nearer than 0.48 deg to another item.

At the beginning of each trial, the items, fixation, and occluders (if present) were displayed. After 1 s, the four target items flashed off and on five times (disappearing and reappearing for 167 ms each on each flash). The ten seconds of item motion then ensued, after which the probe item was flashed off and on five times (in the same manner as the target flashes), against the background of items (and possibly occluders) in their final positions.

Occluder Conditions. Each trial employed one of four possible occlusion methods, depicted schematically in Figure 2. On *No Occlusion (Visible Outlines)* trials, items moved continuously throughout the motion phase, always passing right through the visible outlined ‘occluder’ rectangles without interruption. On *Occlusion* trials, visible occluders appeared on the screen throughout the trial, as described above. Item motion proceeded as in the *No Occlusion (Visible Outlines)* condition, except that those portions of items that happened to intersect (i.e. were 'behind') occluders were not visible. This resulted in perceptual occlusion during item motion, with highly salient accretion and deletion cues at the edges of the occluder boundaries. Across all trials, items were completely occluded — and thus not visible — for a minimum of 60 ms per occlusion, and a maximum of 2640 ms per occlusion. The average duration of total occlusion for an item (per instance of occlusion) was 322.38 ms. (Of course, items did not need to pass through occluders completely in each instance of occlusion. Items often merely 'brushed' occluder boundaries, or even stayed half-occluded during a brief interval in which the item moved roughly parallel to the occluding edge. The actual item trajectories were calculated without regard for the positions of the occluders, excepting their final positions, as described above.) Across all trials, each item was completely occluded an average of 2.45 times per trial, with an average of 19.6 complete occlusion events per trial.

On *Implosion/Explosion* trials, items were present on the display at all the same times and to the same degrees as in the *Occlusion* trials, but the items were drawn differently when they intersected an occluder. Instead of accreting and deleting along a fixed contour (i.e. the edge of the occluder), objects in this condition imploded and exploded into and out of existence from their *centers* when they reached the boundaries of the (invisible, or 'virtual') occluders. Thus, whereas the last visible portion of a nearly-completely-occluded item in the *Occlusion* condition might be a single horizontal strip immediately adjacent to the occluding boundary, in the *Implosion/Explosion* condition it might rather be a small square — sometimes with a few added pixels added to one or two edges to equate the item's area with the other conditions — slightly displaced off of the occluding boundary; see Figure 2. This method resulted in the same amounts of visible item area during each frame of motion as in the other conditions, and the same rates of accretion and deletion.

Items in the *Instantaneous Disappearance/Reappearance* condition were drawn on the display (always in their entirety) only when they did not intersect a ('virtual', invisible) occluder position. Items disappeared instantaneously upon intersection with an occluder position, and reappeared instantaneously as soon as the intersection ceased.

Design and Procedure. Twenty sets of trajectories (along with target selections, probe selection, and occluder positions) were generated and stored offline. We refer to these as the 20 *trajectory files.* On ten of these trials, the probe item was one of the four target items; on the other ten trials, the probe item was one of the four distractors. Each of the trajectory files was then combined with each of the four occluder conditions, for a total of 80 trials.

At the beginning of each experimental session, subjects were instructed to maintain fixation throughout the experiment, and to respond on each trial as soon as the probe item was flashed, by pressing one of two keys to indicate whether the probe item was one of the target items or one of the nontarget items. Eye movements were not monitored, since this has not been found to affect performance on this task. (Pylyshyn & Storm, 1988, monitored and ensured fixation by discarding trials on which subjects made eye-movements, and obtained qualitatively identical results to other investigators who either employed no special constraints or instructions concerning fixation — e.g. Intriligator, 1997; Yantis, 1992 — or else instructed subjects to maintain fixation but did not monitor eye-movements — e.g. Sears & Pylyshyn, in press.) Subjects first completed ten practice trials for which data were not collected, and then completed the 80 experimental trials in a randomized order (different for each subject), with a self-timed break every 27 trials. The entire experimental session took about 35 minutes.

Results

Accuracy, measured in terms of the frequency with which subjects correctly identified whether the probe was one of the targets, was recorded for each experimental trial, with a total of 20 possible correct responses for each of the four occluder conditions. (Note that since this measure included both positive and negative trials, it is not confounded with any response bias.) The mean accuracy and standard errors for each condition are shown in Figure 3. An analysis of variance on these accuracy data (i.e. number correct, out of a possible 20) revealed a significant effect of the occlusion method factor ($F(3, 42) = 36.64$, $p < .001$).

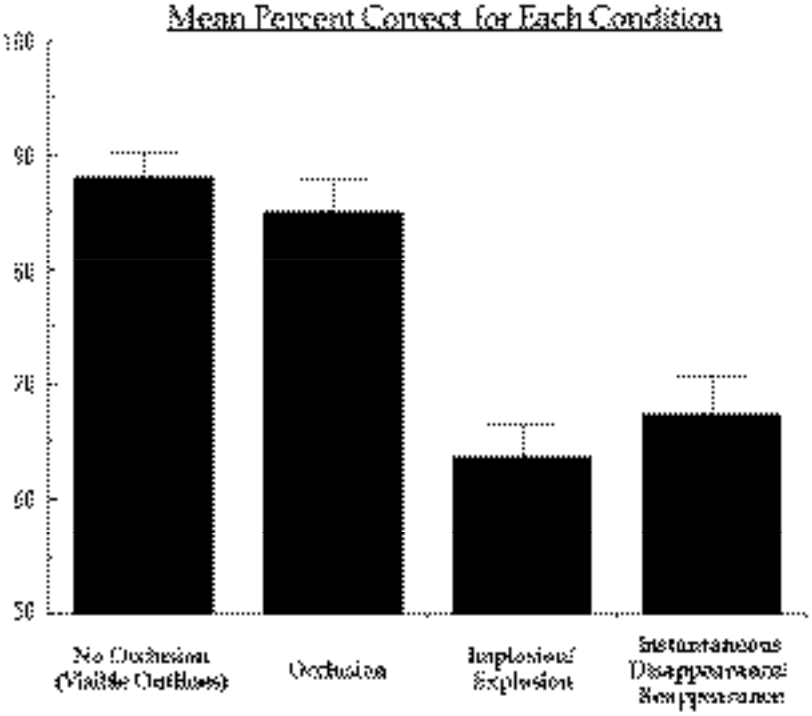

FIG 3. Mean Percent Correct and Standard Errors For Occluder Conditions in Experiment 1.

Additional planned comparisons indicated that (a) performance in the *Occlusion* and *No Occlusion (Visible Outlines)* conditions did not significantly differ ($F(1, 14) = 3.20$, $p > .05$); (b) performance in the *Implosion/Explosion* condition was significantly worse than in both the *Occlusion* condition ($F(1, 14) = 51.02$, $p < .001$) and the *No Occlusion (Visible Outlines)* condition ($F(1, 14) = 75.67$, $p < .001$); and (c) performance in the *Instantaneous Disappearance/Reappearance* condition was significantly worse than in both the *Occlusion* condition ($F(1, 14) = 34.62$, $p < .001$) and the *No Occlusion (Visible Outlines)* condition ($F(1, 14) = 42.85$, $p < .001$).

Discussion

The results of Experiment 1 are consistent with the hypothesis that the mechanisms responsible for multiple-object tracking make allowances for occlusion and are able to track items despite their brief disappearances behind occluding surfaces. Overall performance was quite high (around 90%) in the *No Occlusion (Visible Outlines)* baseline condition, and was qualitatively similar to other such demonstrations (e.g. Intriligator, 1997; Pylyshyn & Storm, 1988; Sears & Pylyshyn, in press; Yantis, 1992), replicating the standard result that subjects can successfully track four independently and unpredictably moving items in a total field of eight identical items.

Both main predictions were confirmed. Tracking performance was not adversely affected when items briefly disappeared behind occluders, suggesting that occlusion does not disrupt the responsible mechanism: items are continuously tracked through occlusion (inasmuch as the same visual index reacquires the item), even though they are briefly absent from the visual field. Significantly impaired performance was observed, however, in each of the other two control conditions which did not support an interpretation of occlusion, indicating that successful tracking through occlusion is not due to a robust tolerance for spatiotemporal interruption in general.

We would explain the impairment of performance in the *Instantaneous Disappearance/Reappearance* condition (a) by appeal to the hypothesis that items can be tracked through interruptions in spatiotemporal continuity only when the interruption is perceived as being caused by the object moving behind an occluder, along with (b) the fact that instantaneous transitions (without accretion / deletion cues) do not implicate the presence of an occluding contour. (As one reviewer noted, however, this may be more or less true for different item sizes, for example single pixels.) In fact, such instantaneous appearances probably do actively interfere with the attentional mechanisms responsible for multiple-object tracking. Yantis and his colleagues have demonstrated that such ‘sudden onsets’ exogenously attract attention (Jonides & Yantis, 1988; Yantis & Jonides, 1984, 1990; see also Scholl, in press; Theeuwes, 1991), and it is an explicit part of Pylyshyn's spatial indexing theory that sudden onsets attract indexes. Such attraction would presumably interfere with maintaining current index assignments, particularly when the items to which those indexes have been assigned have recently (and instantaneously) disappeared from the visual field (cf. Sears & Pylyshyn, in press).

The impaired performance in the *Implosion/Explosion* condition indicates that even the presence of gradual accretion and deletion does not by itself allow successful tracking. Items were present on the display at the same times and to the same degrees in the *Implosion/Explosion* and *Occlusion* conditions, but only the latter condition supported successful tracking.

One remaining difference between the *Implosion/Explosion* and *Occlusion* conditions is that the geometric center of the visible portion of a disappearing item continued moving in the *Occlusion* case until finally disappearing, but stopped briefly (at least in that component of its motion which was orthogonal to the occluder boundary) while imploding in the *Implosion/Explosion* case, slightly offset from the occluder boundary— see Figure 2. (Similarly for accretion / exploding.) This difference could fuel several alternate explanations of the impaired performance in the *Implosion/Explosion* condition — for example, it could be taken to indicate that the motion continued *in depth*, with the item rapidly receding during the implosion phase.**[2]** Such alternate explanations seem unlikely, given the small objective differences between these conditions, but we tested them as follows. Ten additional subjects viewed 20 *Implosion/Explosion* trials and 20 new trials which were identical except that the imploding or exploding item was always drawn immediately adjacent to the occluder position. In other words, an item now imploded while its geometric center continued to move toward the occluder, and eventually disappeared immediately adjacent to the occluder instead of slightly offset (and likewise for the explosion case). These two conditions employed the same 20 trajectory sets as the conditions in the previous main experiments. Subjects viewed these 40 trials (in a different random order for each subject) using the same procedure as above. Performance in the two conditions did not significantly differ ($t(9) = 1.03$, $p > .1$), failing to support the alternate explanations, and suggesting that the slight differences in the motion of the item's geometric centers did not affect the results of the other experiments. Indeed, performance in the new control condition was slightly worse than performance in the *Implosion/Explosion* condition, though not significantly so. (We also note informally that it was difficult even to distinguish these two types of trials in practice, while engaged in the experimental task.)

Based on the impaired performance in the *Implosion/Explosion* condition, then, the crucial aspect of occlusion appears not to be the presence of accretion and

---

[2] Thanks to several colleagues — including Dave Melcher and two anonymous reviewers — for stressing these alternate explanations.

deletion in general, but rather the presence of such cues along a fixed contour, which supports the perception of an occluding surface. We suggest, with Gibson, that the implosion and explosion cues indicate items which are going 'in and out of existence' rather than 'in and out of sight'. This is also consistent with Yantis' (1993) findings that certain types of "onsets capture attention not because they are accompanied by a luminance increment, but because they mark the appearance of a new perceptual object" (p. 157).

EXPERIMENT 2

In Experiment 2 we seek to further refine the nature of occlusion, as it is exploited by the mechanisms responsible for tracking, when computing enduring objecthood across spatiotemporal gaps. First, we seek to determine if the visible occluder outlines are themselves necessary for unimpaired tracking, or if the local accretion and deletion cues themselves can support the presence of an occluding surface. Second, we examine whether accretion and deletion must be along a *particular* fixed contour in order to support successful tracking. Experiment 2 is identical to Experiment 1 except in the number and types of occluder conditions. There were five different occluder conditions in Experiment 2:

Condition 1: No Occlusion. This baseline condition is identical to the *No Occlusion (Visible Outlines)* condition of Experiment 1, except that the occluder rectangles were not visible. This condition is thus a more exact replication of the standard multiple-object tracking paradigm (e.g. Intriligator, 1997; Pylyshyn & Storm, 1988; Yantis, 1992). We expect few errors in this condition, consonant with these previous demonstrations, and with the results of Experiment 1.

Condition 2: Occlusion. This condition was identical to the *Occlusion* condition of Experiment 1.

Condition 3: Virtual Occlusion. This condition was identical to the 'Occlusion' condition, except that the occluder rectangles — while functionally present — were not visible. Although it may seem that this would provide another control for occlusion, pilot work confirmed that the phenomenal sense of occlusion in this condition is not at all attenuated. This is supported by traditional demonstrations such as Michotte's *tunnel effect* (e.g. Michotte, Thinès, & Crabbé, 1964/1991) wherein accretion and deletion cues *themselves* result in an interpretation of occlusion (see discussion below). We thus hypothesize that this 'virtual occlusion' condition will be treated as a case of actual occlusion and thus will not impair performance.

Condition 4: Implosion/Explosion. This condition was identical to the *Implosion/Explosion* condition of Experiment 1.

Condition 5: Reversed Virtual Occlusion. Invisible virtual occluders were present in this condition just as in the 'Virtual Occlusion' and 'Implosion/Explosion' conditions, but again the character of the items' disappearances and reappearances in relation to these virtual boundaries did not support the perception of occlusion. Whenever an item in this control condition reached a virtual 'occluding' edge, it disappeared or reappeared — to the same degree and at the same rate as in the previous conditions — but from the *opposite* fixed contour as in the 'Occlusion' and 'Virtual Occlusion' conditions. In other words, the accretion and deletion cues in this condition were from the 'wrong' sides, producing exact mirror images of the local item behavior in the 'Occlusion' and 'Virtual Occlusion' conditions: when an item reached an occluding boundary, the trailing edge of the item actually stopped while the leading edge moved 'backwards', away from the occluder. As in the 'Implosion / Explosion' condition, we predict that performance in this condition will be significantly impaired relative to the 'No Occlusion' and 'Occlusion' conditions. See Figure 2 for 'snapshot' depictions of these different occlusion methods.

Note that *identical trajectory sets and target selections* were again used in all conditions, to ensure that no differences existed between the trials of each condition except for the manipulation of occlusion method. (Again, there were enough trials, in randomized order, so that subjects did not notice the repetition of trajectory files.) Moreover, we used all and only the trajectory files from Experiment 1, in order to afford an informal comparison of performance in conditions across experiments.

In summary, we have three crucial predictions regarding Experiment 2. First, we again predict unimpaired performance in the 'No Occlusion' and 'Occlusion' conditions, replicating the result from Experiment 1 that tracking performance is not disrupted by the presence of occlusion. Second, we predict that tracking performance should be disrupted in the 'Reversed Virtual Occlusion' control, in which the items behave locally in ways that are quite similar to occlusion, without supporting the interpretation of an occluding surface. (Experiment 2 also allows us to replicate the result of impaired performance in the 'Implosion / Explosion' condition.) Third, we predict that there will be little or no disruption in tracking performance in the 'Virtual Occlusion' case, since there is reason to expect that the visual system may treat this condition as one of actual occlusion, as in Michotte's *tunnel effect*.

Method

This experiment was identical to Experiment 1, except where noted.

Participants. Fifteen Rutgers University undergraduates, none of whom had participated in Experiment 1, participated in one individual 50-min session to fulfill an introductory psychology course requirement. All subjects had normal or corrected-to-normal vision.

Occluder Conditions. Each trial employed one of five possible occlusion methods, depicted schematically in Figure 2. On *No Occlusion* trials, items moved continuously throughout the motion phase, with no visible occluders or interruptions. The *Occlusion* and *Implosion/Explosion* conditions were identical to those in Experiment 1. *Virtual Occlusion* trials were identical to *Occlusion* trials, except that the occluders themselves were not drawn. Items still behaved as if there were occluders present, with the associated disappearances, reappearances, and accretion / deletion cues. *Reversed Virtual Occlusion* trials were identical to the *Virtual Occlusion* trials, except that, upon intersecting an occluder position, an item accreted or deleted (again, at the same times, to the same degrees) along the *opposite* fixed contour, such that the trailing edge of the item actually stopped while the leading edge 'retracted' away from the occluder. The accretion and deletion cues were thus from the 'wrong' sides of the items in this condition, producing exact mirror images of the local item behavior in the *Occlusion* and *Virtual Occlusion* conditions.

Design and Procedure. Each of the 20 trajectory files from Experiment 1 (including item trajectories, target selections, probe selection, and occluder positions) was combined with each of the five occluder conditions, for a total of 100 trials. Subjects first completed ten practice trials for which data were not collected, and then completed the 100 experimental trials in a randomized order (different for each subject), with a self-timed break every 25 trials. The entire experimental session took about 45 minutes.

Results

Accuracy was recorded for each experimental trial, again with a total of 20 possible correct responses for each of the five occluder conditions. The mean accuracy and standard errors for each condition are shown in Figure 4. An analysis of variance on these accuracy data revealed a significant effect of the occlusion method factor ($F(4, 56) = 19.31$, $p < .001$).

Additional planned comparisons indicated that (a) performance in the *Occlusion* and *No Occlusion* conditions did not significantly differ ($F(1, 14) = 4.15$, $p > .05$); (b) performance in the *Virtual Occlusion* condition did not significantly differ from either the *No Occlusion* condition ($F(1, 14) = 2.41$, $p > .1$) or the *Occlusion* condition ($F(1, 14) = .01$, $p > .1$); (c) performance in the *Implosion/Explosion* condition was significantly worse than in both the *No Occlusion* condition ($F(1, 14) = 62.26$, $p < .05$) and the *Occlusion* condition ($F(1, 14) = 27.97$, $p < .05$); and (d) performance in the *Reversed Virtual Occlusion* condition was significantly worse than in both the *No Occlusion* condition ($F(1, 14) = 14.56$, $p < .05$) and the *Occlusion* condition ($F(1, 14) = 5.24$, $p < .05$).

Discussion

High levels of performance were observed in both the *Occlusion* and *No Occlusion* conditions, further supporting the hypothesis that the mechanisms responsible for tracking take occlusion into account when computing continuing objecthood. The observed levels of performance are again qualitatively similar to previous demonstrations of successful multiple-object-tracking. These results also replicate the result from Experiment 1 that performance is impaired when the accretion/deletion cues are based on implosion and explosion rather than along a fixed contour.

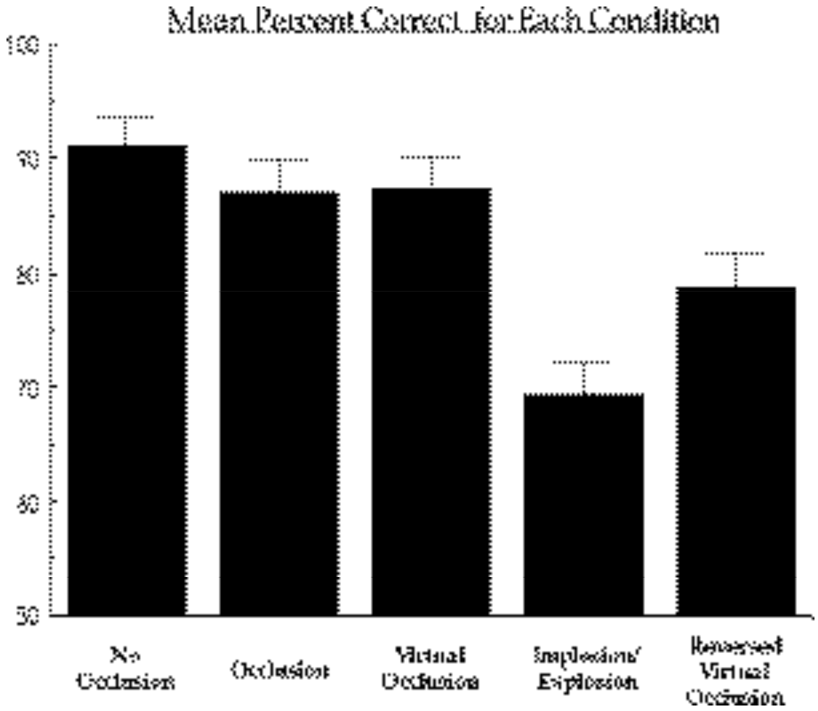

FIG. 4. Mean Percent Correct and Standard Errors For Occluder Conditions in Experiment 2.

The impaired performance in the *Reversed Virtual Occlusion* condition further refines this interpretation. Items in this condition did accrete and delete along a fixed contour when ‘occluded’, but did not support successful tracking. It thus appears that successful tracking requires accretion/deletion cues along a particular fixed contour, namely that contour which is adjacent to the (possibly virtual) occluder.[3]

---

[3] This condition was motivated simply as another control in which the items behaved locally in ways which were very similar — but not identical — to occlusion. Informal inspection of these displays, however, suggests that the impaired performance may be due to occlusion after all, via anomalous

Our prediction regarding 'virtual occlusion' was also confirmed. The presence of *visible* occluders (i.e. visible edges of occluding surfaces) does not appear to be necessary for successful tracking, since performance in the *Virtual Occlusion* case did not differ significantly from the baseline, or from the condition with visible occluders. We would argue that this is because the visual system interprets this 'virtual' occlusion as *bona fide* occlusion, consonant with Michotte's *tunnel effect* (Michotte et al., 1964 / 1991). In the words of Kahneman et al. (1992),

> If an object is seen to disappear (with gradual occlusion [i.e. deletion cues]) and then to reappear some distance away (with gradual disocclusion [i.e. accretion cues]), subjects report a compelling impression that a single object disappeared into a tunnel or behind a wall, traveled invisibly in the interval, and finally reappeared at the other end. (Kahneman et al., 1992, p. 180)

This is precisely what happens in our *Virtual Occlusion* condition. Despite the lack of visible occluders, subjects perceive items as disappearing behind invisible occluding surfaces. The high level of performance on this condition thus serves as an additional demonstration that multiple objects can be successfully tracked despite their disappearance from view, so long as the character of their disappearance supports the perception that they have moved behind an opaque occluding surface.

EXPERIMENT 3

Subjects in the previous two experiments were able to track multiple independently and unpredictably moving items equally successfully regardless of whether those items passed completely behind stationary occluders as they moved. These results suggest that this successful performance requires accretion and deletion cues along a fixed contour adjacent to the occluder position, but does not require visible occluders. In this experiment we seek to refine this picture further, by asking whether successful performance requires globally consistent occluder positions.

In the previous experiments, the locations of the (possibly virtual) occluders were randomly assigned in each trajectory file. Nevertheless, each occluder still remained in the same stationary location *throughout* each trial, and all of the items

---

motion cues from the 'reversed' accretion and deletion, in the opposite direction from the item's previous trajectory. In other words, the reversed deletion cues (wherein the trailing edge of an item stops, and the leading edge moves 'backward', away from the occluder) may be interpreted as a sudden 180 deg change in the direction of motion, coupled with the sudden perceptual appearance of a 'virtual' occluder in the location just traversed by the item (perhaps at a different depth plane). Thus, it might be argued that this condition involves an interpretation of occlusion after all, rather than the desired control. We are happy to endorse this as a possible interpretation, since such a sensitivity to occlusion is exactly what we are trying to demonstrate.

in a trial were occluded in these same locations. Experiment 3 is similar to the previous two experiments, with the addition of a new condition, *Local Virtual Occlusion*, in which global consistency is violated.[4] Each instance of disappearance and reappearance in this condition employed local accretion and deletion cues which were identical to those in the previous *Occlusion* and *Virtual Occlusion* conditions, but here these events did not take place in the same locations for every item. Rather, each moving item occluded and disoccluded relative to its own private set of occluders, none of which were in the same locations as the occluders for any other item. Thus, while each individual item still occluded and disoccluded at consistent locations throughout a trial, one item could undergo a brief occlusion while passing from point A to point B, while another item (perhaps moving in parallel to the first) moved from A to B without interruption. Furthermore, this situation could be reversed in a new location a moment later, such that there was no consistent (static) depth ordering of items and occluders that would 'make sense' of the patterns of occlusion. Performance in this condition is compared to performance in *No Occlusion*, *Occlusion*, and *Virtual Occlusion* conditions.

If global properties of the display (such as a globally consistent occluder position) are inferred from the behavior of the items as a group and used to enhance tracking performance, then we would expect performance in this condition (in which this global property is not consistent) to be impaired relative to other conditions. This experiment thus addresses the question of whether a purely locally-based mechanism (centered at each visual index), with no access to information about the global scene, can account for tracking through occlusion.

Method

This experiment was identical to Experiments 1 and 2, except where noted.

Participants. Fifteen Rutgers University undergraduates, none of whom had participated in the previous experiments, participated in one individual 40-min session to fulfill an introductory psychology course requirement. All subjects had normal or corrected-to-normal vision.

Occluder Conditions. Each trial employed one of four possible occlusion methods, depicted schematically in Figure 2. The *No Occlusion*, *Occlusion*, and *Virtual Occlusion* conditions were identical to those in Experiment 2. In the *Local Virtual Occlusion* condition, items behaved locally just as in the previous *Virtual Occlusion* condition, except that the vertical positions of the two horizontal virtual occluders were different for each item. For each item in a trial, two occluder positions were chosen randomly (without replacement) from all of the occluder positions in the 20 base trajectory files, again constrained so that items did not intersect occluders at the beginning or end of their

---

[4] We thank Thomas Papathomas for suggesting this condition.

motion trajectories. Each item was drawn on the screen throughout its trajectory, excepting those portions which intersected one of that item's 'private' occluder positions.

Design and Procedure. Each of the 20 trajectory files from Experiment 1 (now including item trajectories, target selections, and probe selection, but *not* occluder positions) was combined with each of the four occluder conditions, for a total of 80 trials. Subjects first completed ten practice trials for which data were not collected, and then completed the 80 experimental trials in a randomized order (different for each subject), with a self-timed break every 27 trials. The entire experimental session took about 35 minutes.

Results and Discussion

Accuracy was recorded for each experimental trial, again with a total of 20 possible correct responses for each of the four occluder conditions. The mean accuracy and standard errors for each condition are shown in Figure 5. An analysis of variance on these accuracy data revealed no effect of the occlusion method factor ($F(3, 42) = 2.07$, $p > .1$).

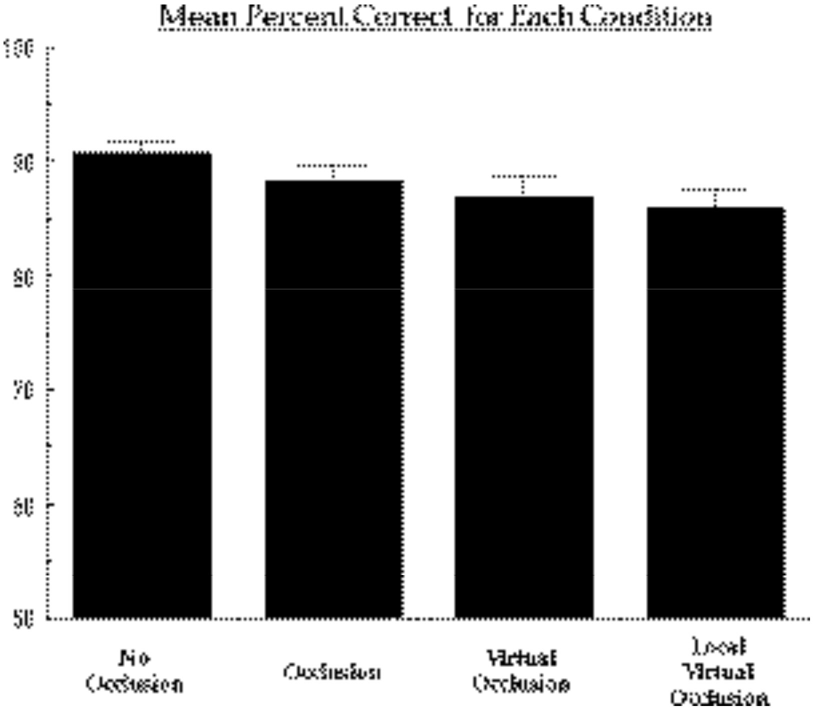

FIG. 5. Mean Percent Correct and Standard Errors For Occluder Conditions in Experiment 3.

There was no significant difference between performance on any of the occlusion conditions in this experiment. This result has two important implications. First, note that this constitutes a second replication of the crucial finding of Experiment 1 — that the presence of occluders (visible or virtual) does not disrupt tracking performance relative to a baseline condition with no occlusion. Second, the fact that performance in the *Local Virtual Occlusion* condition did not differ significantly from the other conditions indicates that tracking through occlusion is mediated by a locally-supported mechanism that need not use global information about the scene.

GENERAL DISCUSSION

Before summarizing the implications of our experiments, we turn to a brief discussion of some related issues, including: (a) relations to other recent experimental demonstrations that a sensitivity to occlusion permeates visual processing, even at its lowest levels; (b) a discussion of the neurophysiological plausibility of tracking through occlusion; and (c) implications for other theories of object-based attention, such as Kahneman and Treisman's theory of *object files*.

Other Evidence that Allowances for Occlusion Permeate Low-Level Visual Processing

We have suggested that allowances for occlusion *should* characterize early visual processing, based on the fact that occlusion clearly permeates visual experience: "occlusion is one of the most fundamental facts about vision in daily life" (Shimojo & Nakayama, 1990, p. 285). These sorts of general considerations have been borne out by the experiments reported in this paper, and also by several recent related demonstrations. Most of these demonstrations show that "preattentive processes do more than simply register and group together elementary properties of the two-dimensional image — they are also capable of determining properties of the corresponding three-dimensional scene" (Enns & Rensink, 1991, p. 346).

Many of these experiments involve static stimuli in the context of visual search (Davis & Driver, 1994; Enns & Rensink, 1998; He & Nakayama, 1992, 1994), but similar results have been reported in other paradigms involving dynamic stimuli (Shimojo & Nakayama, 1990; Shimojo, Silverman, & Nakayama, 1989; Tipper, Brehaut, & Driver, 1990; Watamaniuk & McKee, 1995; Yantis, 1995). We briefly describe three of these related studies.

Yantis (1995) demonstrated that the character of apparent motion in Ternus configurations is affected by the presence of an occluding surface. Subjects in his experiments viewed bistable apparent motion displays such as those in Figure 6. Frames 1 and 2 are alternated in quick succession, and observers see either *element motion*, in which a single moving dot is seen to 'hop' back and forth over a stationary dot in the middle, or *group motion*, in which both dots are seen to move back and forth as a pair. At short ISIs (< 20 ms), subjects tend to see the former; at longer ISIs (> 150 ms), they tend to see the latter. At intermediate ISIs, the display is bistable.

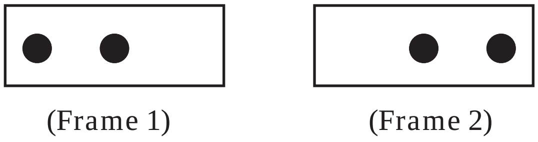

FIG. 6. A simple bistable apparent motion display. See text for details.

Yantis (1995) had subjects view bistable apparent motion displays in which the ISI sometimes contained an occluder in the center of the display, made up of illusory contours, and presented stereoscopically either behind or in front of the other items. When the occluder was seen as being in front of the apparently-moving items, subjects were more likely to see element motion at traditionally bistable ISIs. This indicates "the influence of 3-D surface-based representations even in very simple visual tasks" (p. 182). Yantis (1995) refers to this phenomenon as *amodal integration*, wherein "a momentarily occluded object is perceived as continuing behind the occluder through time" (p. 182). We have avoided using this term to describe our experiments because we wish to stress the importance of accretion / deletion cues (see above), which were not present in Yantis' apparent-motion experiment.

Tipper, Brehaut, and Driver (1990) report an experiment with actual motion which can be given a similar interpretation: they found that the *negative priming effect* (NPE) was a function of objecthood (as opposed to spatial locations), and that the NPE 'stuck' to moving objects, even when those objects passed behind occluders. "[W]hen an object moves . . ., inhibition appears to 'move' with it, even when the objects undergo occlusion" (p. 499). Tipper et al. (1990) explain this phenomenon by appeal to object-files surviving occlusion, an interpretation which we endorse below.

Finally, Viswanathan and Mignolla (1998) recently provided converging evidence that the mechanisms responsible for multiple object tracking are sensitive to occlusion cues. Subjects in their experiment tracked multiple items whose trajectories were linear (and 'bounced' off of the walls and the fixation point), but were allowed to overlap. Performance with such trajectories was impaired, but this impairment was greatly attenuated when the items' relative depths were made apparent via either T-junctions at the intersecting items' border, or by disparity information in conditions where the items were tracked in three-dimensional space.

The Neurophysiological Plausibility of Tracking Through Occlusion

Our account requires that there exist visual mechanisms that can track objects even when they are not visible on the retina. Is this view neurophysiologically tenable? Some recent evidence suggests that it is. While there are many neurophysiological demonstrations of motion extrapolation in general (e.g. Duhamel, Colby, & Goldberg, 1992; Nijhawan, 1994), a recent paper by Assad and Maunsell (1995) reports the discovery of groups of neurons which fire as though signaling the presence of motion behind an occluder, even during intervals in which there is no motion on the retinal image. Rhesus monkeys in their experiment viewed the motion of a small stimulus in three different conditions, while maintaining fixation. In their 'full vision trials', the object simply moved from the periphery to the center of gaze. In their 'occlusion trials' the object followed the same path, but disappeared briefly during it's trajectory. Albright (1995) notes that "the visible target positions and the timing of the intervening delay were consistent with target motion occurring behind an opaque occluder, along a path that was identical to the real movement of the full-vision condition" (p. 332). Trials in their 'blink' control condition were identical to the occlusion trials, except that there was no motion cue: the item reappeared in the same place in which it originally disappeared.

The motion paths in each condition were aligned so that the stimulus passed in the preferred direction through the receptive fields of neurons in posterior parietal cortex. Assad & Maunsell recorded the responses of these neurons. While each neuron responded to the 'full vision' trials, approximately half of the sampled cells responded even during those intervals of the 'occlusion trials' which contained no visible stimulus. (In contrast, no neurons responded during the corresponding interval of the 'blink' control condition.) Albright (1995) emphasizes the uniqueness and import of this result: "by manipulating crucial contextual cues, [Assad and Maunsell] have cut the direct links between sensory and perceptual events that are characteristic of most neurophysiological studies of the primate visual system" (p. 333).

The exact function of this part of posterior parietal cortex remains unknown at present, but it is possible that these neurons are part of the underlying neurophysiological implementation of an attentive tracking system such as Pylyshyn's visual indexing mechanism. This possibility is supported by recent neuroimaging research indicating increased (and bilateral) parietal activation in a multiple-object tracking task, relative to a passive viewing condition and controlling for eye movements (Culham, Brandt, Cavanagh, Kanwisher, Dale, &

Tootell, 1998; Culham, Cavanagh, & Kanwisher, 1998). This scenario is further supported by recent neuropsychological evidence that patients with parietal lesions cannot perform attentive tracking tasks (Michel, Henaff, & Intriligator, 1997).

Implications for Object Files

We have motivated our experiments in the context of Pylyshyn's 'FINST' theory of visual indexing (e.g. Pylyshyn, 1989, 1994), and in this context our results indicate that visual indexes can indeed survive occlusion (as indicated by the appropriate accretion / deletion cues), but not similar interruptions in spatiotemporal continuity which do not implicate the presence of an occluding surface. These sorts of implications also hold in the case of Kahneman and Treisman's theory of *object files* (Kahneman & Treisman, 1984; Kahneman et al, 1992; Treisman, 1988).[5]

One traditional model of visual experience goes something like this, to a first approximation: Visual stimuli are identified as objects when their visual projections activate semantic representations in long-term memory. Visual experience, then, consists simply in various shifting patterns of this type of LTM activation. Kahneman et al. (1992) call this the 'display-board model of the mind', and note that it has a number of distressing shortcomings. It appears to be the case, for instance, that objects can be perceived and tracked through space even when they remain unidentified. Furthermore, when objects are initially mis-identified, and later correctly recognized, there is still never any doubt that the object involved was the same object. Identification *of* a particular object, in other words, is distinct from identification *as* an object.

Kahneman and Treisman have argued that an intermediate level of representation is needed to mediate this latter task. In their theoretical framework, this role is played by *object files*. According to their theory, attending to an object in the visual field causes a temporary representation called an *object file* to be created. This object file stores information about the object's properties (including its color, shape, and current location), and this information is continually updated when the world changes. Each time an object's properties change, the new state of the object is compared with the previous state of the object file. If these two states are similar enough, then the object is seen as continuous, and the object file is updated appropriately. If the two states are

---

[5] Note that we assume, as do Kahneman and Treisman, that visual indexes and object-files are both parts of a single indexing system. There are plenty of details to iron out, but Kahneman et al. (1992) suggest that "We might think of [a visual 'FINST' index] as the initial spatiotemporal label that is entered in the object file and that is used to address it. . . . [A] FINST might be the initial phase of a simple object file before any features have been attached to it" (p. 216). More precisely, the FINST system is the *mechanism* by which properties are bound to objects. Such a mechanism is tacitly presupposed by most object-based theories of attention.

dissimilar enough, however, then the previous object file decays, and a new object file is opened to represent the 'new' object. Kahneman et al. (1992) describe three operations which are involved in managing object files: (a) a *correspondence* operation, which determines for each object whether it is novel, or whether it moved from a previous location, as indicated in an object file; (b) a *reviewing* operation, which retrieves an object's previous characteristics, some of which may no longer be visible; and finally (c) an *impletion* operation which uses both current and reviewed information to construct a phenomenal percept, perhaps of object motion.

Kahneman et al. (1992) demonstrated that object files survive real and apparent motion, but what about occlusion? Object files are thought to be eventually discarded after their associated objects disappear from the visual field, although this decay is clearly slow enough so that the files survive the brief ISIs in apparent motion displays. Kahneman et al. (1992) hypothesized that object files might survive occlusion (see pp. 177, 180), but did not test this empirically. Our results suggest that object files may indeed be able to survive longer intervals of interruption, so long as the character of the interruption signals the presence of an occluder.

The following is a reconstruction of how the object file theory might accommodate the present results. In the context of Kahneman and Treisman's theory, subjects perform multiple-object tracking by opening object-files for each of the targets. Since the items are featurally similar, these object-files must track their objects solely via spatiotemporal continuity cues. When an item disappears suddenly or anomalously, the corresponding object file is discarded, and a new object file is created in response to the ensuing sudden or anomalous reappearance. Since this new object file does not possess the historical location trace of the discarded file, performance on the tracking task is impaired. When an object disappears and reappears via accretion/deletion cues along the appropriate fixed contour, however, this signals to the object file system that an occluder has been encountered, which causes the object file not to be discarded. Rather, the object file reacquires the object when it reappears (via the *correspondence* operation), and then a percept of continuous motion behind the occluder is constructed via the *impletion* operation. The result is successful tracking through occlusion.[6]

---

[6] We are also pursuing the question of motion extrapolation during occlusion. We have suggested that discrete representations such as visual indexes or object files continue to track items through occlusion, but exactly what trajectories do these items track? In the experiments reported here, the occlusion durations were short enough (and the item densities low enough) that there was rarely any ambiguity involved in the reacquisition of an item following occlusion. Recall that the item trajectories in our experiments were generated without regard for the positions of the occluders. In other experiments not reported here, we have allowed the item trajectories and occluder positions to interact in more complex ways, such that items reappear at the 'right' times but the 'wrong' places — i.e. slightly displaced along the occluder boundaries from where they would have appeared if they

Summary and Conclusions

Insofar as multiple-object tracking studies provide evidence for the individuation and tracking of what we have referred to as ‘visual objects’, the present studies may be viewed as the first steps in an investigation of the nature of such visual objecthood. We have hypothesized that *spatiotemporal continuity* is one of the hallmarks of objecthood at this level of the visual system, but that *occlusion* is taken into account when deciding if an interruption in spatiotemporal continuity constitutes a disruption in continuing objecthood. We have reported three multiple-object tracking experiments supporting this hypothesis, and in the process have refined the notion of occlusion, as it appears to be employed in visual tracking.

Note that although our results have direct implications only for notions of *visual* objecthood, the indirect implications of the experiments reported here may involve other more general notions of objecthood, such as the so-called *object concept* in infancy. For a discussion of how the results of these experiments impact theories of infants’ notions of physical objecthood, see Leslie, Xu, Tremoulet, & Scholl (1998), Scholl (1997), and Scholl and Leslie (in press).

These experiments indicate that tracking survives occlusion *qua* occlusion, and not just any modest interruptions in spatiotemporal continuity. In particular, the results from these experiments allow us to evaluate the importance of five different properties for successful tracking through interruptions in spatiotemporal continuity.

- First, successful tracking in such situations requires the presence of accretion / deletion cues. When such cues are absent, as in the *Instantaneous Disappearance/Reappearance* condition of Experiment 1, performance is significantly impaired.

- Second, these accretion / deletion cues do not require the presence of visible occluding surfaces or edges. Objects can be perceived as moving behind an occluding surface even in the absence of independently-perceived surfaces, as occurs in the so-called *tunnel effect*. When occluders are invisible but functionally present, as in the *Virtual Occlusion* conditions of Experiments 2 and 3, performance is no less successful than when tracking proceeds without any interruption at all in spatiotemporal continuity.

---

had kept the same speeds and trajectories. In other conditions, items reappear at the ‘right’ places but the ‘wrong’ times — i.e. as if they had gotten momentarily ‘stuck’ behind the occluders. By varying the duration and extent of this sort of anomalous behavior, we hope to determine how tolerant the tracking mechanisms are to changes in item trajectories during occlusion.

- Third, for occlusion to be perceived, these accretion / deletion cues must be associated with a fixed contour. When this is not the case, as in the *Implosion/Explosion* conditions of Experiments 1 and 2, performance is significantly impaired.

- Fourth, these accretion / deletion cues must be locally consistent with the perception of a target moving behind an occluding surface, and therefore the accretion / deletion must be from the 'right direction' — i.e. along that fixed contour which is adjacent to the boundary of the occluding surface. When items accrete and delete from the opposite fixed contours, as in the *Reversed Virtual Occlusion* condition of Experiment 2, performance is significantly impaired.

- Fifth, successful tracking in such situations does *not* require globally consistent occluder positions. When each item is occluded in locations distinct from those of each other item, as in the *Local Virtual Occlusion* condition of Experiment 3, performance is no less successful than when tracking proceeds without any interruption at all in spatiotemporal continuity.

These factors are all crucial for the interpretation of an occluding surface, but neglect of such factors has sometimes been encouraged by the use of single-pixel 'point' stimuli in occlusion experiments, which disguise the crucial role of accretion / deletion cues.

Our interpretation of these results is that multiple-object tracking survives interruptions of spatiotemporal continuity only when the character of the interruption supports the interpretation of an occluding surface (to 'explain' the interruption). In other words, the mechanisms responsible for tracking recognize occlusion, and make allowances for it. In such situations, perceptual objecthood is continuously maintained throughout an item's trajectory via some internal representation — such as a visual index or an *object file* — even though the object may frequently disappear completely from the visual field.

REFERENCES

Acton, B., & Eagleson, R. (1993). A neural network model of spatial indexing. *Investigative Ophthalmology and Visual Science, 34*, 413 (abstract).

Albright, T. D. (1995). 'My most true mind thus makes mine eye untrue.' *Trends in Neurosciences, 18,* 331 - 333.

Assad, J. A., & Maunsell, J. (1995). Neuronal correlates of inferred motion in primate posterior parietal cortex. *Nature, 373,* 518 - 521.

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

Chun, M., & Cavanagh, P. (1997). Seeing two as one: Linking apparent motion and repetition blindness. *Psychological Science, 8*, 74 - 79.

Comtois, R. (1994). VisionShell. [Software libraries]. Cambridge, MA: author.

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

Culham, J. C., Cavanagh, P., & Kanwisher, N. (1998). Attention response functions of the human brain measured with fMRI. Manuscript submitted for publication.

Davis, G., & Driver, J. (1994). Parallel detection of Kanizsa subjective figures in the human visual system. *Nature, 371*, 791 - 793.

Downing, C., & Pinker, S. (1985). The spatial structure of visual attention. In M. Posner & O. S. M. Marin (Eds.), *Attention and Performance XI* (pp. 171 - 187). London: Erlbaum.

Duhamel, J., Colby, C., & Goldberg, M. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science, 255*, 90 - 92.

Eagleson, R., & Pylyshyn, Z. W. (1988). A computational model of a 2D ('FINST') tracking mechanism using spatio-temporal operators and a predictive filter. University of Western Ontario, Centre for Cognitive Science, Technical Report No. 38.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48*, 269 - 297.

Enns, J., & Rensink, R. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review, 98*, 335 - 351.

Enns, J., & Rensink, R. (1998). Early completion of occluded objects. *Vision Research, 38*, 2489 - 2505.

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics, 12,* 201 - 204.

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

Gibson, J. J., Kaplan, G. A., Reynolds, H. N, & Wheeler, K. (1969). The change from visible to invisible: A study of optical transitions. *Perception & Psychophysics, 5*, 113 - 116.

He, Z., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature, 359*, 231 - 233.

He, Z., & Nakayama, K. (1994). Perceiving textures: Beyond filtering. *Vision Research, 34*, 151 - 162.

Hirsch, E. (1982). *The concept of identity.* New York: Oxford University Press.

Intriligator, J. M. (1997). The spatial resolution of visual attention. Unpublished doctoral dissertation, Harvard University.

Intriligator, J. M, & Cavanagh, P. (1992). An object-specific spatial attentional facilitation that does not travel to adjacent spatial locations (Abstract). *Investigative Ophthalmology and Visual Science* (Suppl.), *33*, 1263.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics, 43,* 346 - 354.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 29 - 61). New York: Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object -specific integration of information. *Cognitive Psychology, 24*, 174 - 219.

Kanwisher, N. (1987). Repetition blindness: Type recognition without token individuation. *Cognition, 27*, 117 - 143.

Kanwisher, N., and Driver, J. (1992). Objects, attributes, and visual attention: Which, what, and where. *Current Directions in Psychological Science, 1*, 26 - 31.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4,* 219 - 227.

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

Michel, F., Henaff, M., & Intriligator, J. (1997). Posterior callosum split and left parietal lesion reveal visual deficits in the right visual field (Abstract). *Cognitive Neuroscience Society Abstracts*.

Michotte, A., Thinès, G., & Crabbé, G. (1964 / 1991). *Les complements amodaux des structures perceptives*. Louvain: Publications Universitaires. Excerpted in G. Thinès, A. Costall, & G. Butterworth (Eds.), *Michotte's experimental phenomenology of perception* (pp. 140 - 167). Hillsdale, NJ: Erlbaum, 1991.

Nakayama, K., He, Z., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S. M. Kosslyn & D. Osherson (Eds.), *Visual Cognition*, Vol 2. of An Invitation to Cognitive Science, 2nd Ed (pp. 1 - 70). Cambridge, MA: MIT Press.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, *29*, 1631 - 1647.

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature, 370*, 256 - 257.

Posner, M., Snyder, C., & Davidson, B. (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 index model. *Cognition*, *32*, 65 - 97.

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

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

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

Schmidt, W. C., Fisher, B. D., & Pylyshyn, Z. W. (1998). Multiple-location access in vision: Evidence from illusory line motion. *Journal of Experimental Psychology: Human Perception & Performance, 24,* 505 - 525.

Scholl, B. J. (1997). Cognitive development and cognitive architecture: Two senses of ‘surprise’. Paper read at the 23rd annual meeting of the Society for Philosophy and Psychology, 5/5/97, New York.

Scholl, B. J. (in press). Attenuated change blindness for exogenously attended items in a flicker paradigm. *Visual Cognition*.

Scholl, B. J., & Leslie, A. M. (in press). Explaining the infant’s object concept: Beyond the perception/cognition dichotomy. In E. Lepore & Z. Pylyshyn (Eds.), *Rutgers Lectures on Cognitive Science*. Oxford: Blackwell.

Sears, C. R., & Pylyshyn, Z. W. (in press). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology*.

Shimojo, S., & Nakayama, K. (1990). Amodal presence of partially occluded surfaces determines apparent motion. *Perception, 19*, 285 - 299.

Shimojo, S., Silverman, G., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research, 29,* 619 - 626.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics, 49,* 83 - 90.

Tipper, S., Brehaut, J., & Driver, J. (1990). Selection of moving and static object for the control of spatially directed action. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 492 - 504.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lectures. *The Quarterly Journal of Experimental Psychology, 40,* 201 - 237.

Trick, L., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 331 - 351.

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

Viswanathan, L., & Mignolla, E. (1998). Attention in depth: Disparity and occlusion cues facilitate multi-element visual tracking. Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems Technical Report #CAS/CNS-98-012.

Watamaniuk, S., & McKee, S. (1995). Seeing motion behind occluders. *Nature, 377*, 729 - 730.

Wiggins, D. (1980). *Sameness and substance*. Oxford, England: Basic Blackwell.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research, 37*, 25 - 43.

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

Yantis, S. (1993). Stimulus-driven attentional capture. *Current Directions in Psychological Science, 2,* 156 - 161.

Yantis, S. (1995). Perceived continuity of occluded visual objects. *Psychological Science, 6*, 182 - 186.

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

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 & Performance, 10*, 601 - 621.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 121 - 134.

(Accepted September 16, 1998)