Some Primitive Mechanisms of Spatial Attention[1]

Zenon Pylyshyn
Rutgers University and The University of Western Ontario

ABSTRACT

Our approach to studying the architecture of mind has been to look for certain extremely simple mechanisms which we have good reason to suspect must exist, and to confirm these empirically. We have been concerned primarily with certain low-level mechanisms in vision which allow the visual system to simultaneously index items at multiple spatial locations, and have developed a provisional model (called the FINST model) of these mechanisms.

Among the studies we have carried out to support these ideas are ones showing the subjects can track multiple independent moving targets in a field of identical distractors, that their ability to track these targets and detect changes occurring on them does not generalize to nontargets nor to items lying inside the convex polygon that they form (so that a zoom-lens of attention does not fit the data). We have used a visual search paradigm to show that (serial or parallel) search can be confined to a subset of indexed items and the layout of these items is of little importance. We have also carried out a large number of studies on the phenomenon known as subitizing and have shown that subitizing occurs only when items can be preattentively individuated and in those cases location precuing has little effect, compared with when counting occurs, which suggests that subitizing may be carried out by counting active indexes rather than items in the visual field. And finally we have run studies showing that a certain motion effect which is sensitive to attention can occur at multiple precued loci.

We believe that taken as a whole the evidence is most parsimoniously accounted for in terms of the hypothesis that there is an early preattentive stage in vision where a small number of

---

[1] The work reported herein was carried out under the author's direction at the Center for Cognitive Science, University of Western Ontario. Much of it has been reported in theses in the department of psychology and the department of electrical engineering at the University of Western Ontario, by graduate students Brian Acton, Jacquie Burkell, Roy Eagleson, Paul McKeever, William Schmidt, Chris Sears, and Lana Trick. These students, along with our Research Scientist Brian Fisher, contributed most of the empirical and theoretical work discussed in this paper. This research was supported by the Canadian Natural Science and Engineering Research Council (Grant A2600) and by a Project B-4 grant from the Institute for Robotics and Intelligent Systems.

For reprints write to the author at the Rutgers Center for Cognitive Science, Psychology Building, Busch Campus, P.O. Box 1179, Piscataway NJ 08855-1179. Internet: zenon@ruccs.rutgers.edu salient items in the visual field are indexed and thereby made readily accessible for a variety of visual tasks.

Background: The search for cognitive architecture

On the last occassion such as this — the 10th anniversary issue of Cognition — I described the guiding assumption of my research as the belief that there are two kinds of explanatory principles in psychology (Pylyshyn, 1981b). One set of principles advert to the beliefs and goals that the person has and the other is concerned with the kind of computational device that a human organism instantiates — the kinds of basic cognitive operations of which it is capable, the constraints under which it must operate and other "cognitively impenetrable" properties that underwrite its capacities. Another way to put this is to say that there are knowledge-based explanations and architecture-based explanations. I have devoted much of my critical energy to arguing in various forums (e.g., Pylyshyn, 1991) that (1) as scientific psychologists (though never in our role as astute observers of human behavior) we have underplayed the importance of knowledge-based explanations in accounting for various cognitive phenomena and (2) we have underestimated the extent to which our being knowledge-based systems constrains the form that an adequate cognitive model can take. As an example of (1), I have argued that results such as those observed in experiments on "mental scanning", as well as several other mental image manipulation experiments, tell us nothing about the nature of mind, or of the form of representation underlying mental images (Pylyshyn, 1981a). Rather, they tell us what subjects believe about the nature of the task and what they know about the world. I argued that these beliefs, together with the rationality of most mental processes, and the possession of certain minimal psychophysical capacities, explains the relevant regularities observed in such experiments. As an example of (2), Fodor and I (Fodor and Pylyshyn, 1988) argued that the general boundary conditions that must be met by any intelligent knowledge-using or reasoning system requires that its architecture have certain properties lacking in connectionist models, as well as analogue models (Pylyshyn, 1984) and systems whose basic operating principle is something like "resonance" (as is the case with J.J. Gibson's proposal; see Fodor & Pylyshyn, 1981).

If this general picture is true, it suggests that a useful, indeed perhaps a *necessary* strategy for understanding human behavior is first to try to discover some properties of the mechanisms which operate on knowledge and which bring the organism in contact with its environment. This is the strategy advocated by the late Allen Newell, although he argued that one ought to attempt to design an architecture for the entire cognitive system — even if much of it had to be highly provisional. But it is not the only strategy that has been advocated. Indeed, David Marr (Marr, 1982) argued that one should postpone hypothesizing architectural mechanisms until one has a clearer idea of what problem that architecture is designed to solve. However, it is not clear that outside a few problem areas one can make a reasonable attempt to understand "what problem a system is designed to solve". We must acknowledge that in general a cognitive system (or any biological system) is saddled willy-nilly with a certain architecture which may not be optimal for what one assumes to be the problem being solved. This may be so because nature simply made use of a mechanism that happened to be available — handed down from earlier evolutionary stages where it served some unknown function — or it may be so because we do not know what method of cost accounting should be used to compute optimality (see the remarks at the end of Pylyshyn, 1991). In any case in actual practice these different methodological points of view (or doctrines) constitute different heuristics for achieving the same ultimate goal, and it is good that a variety of such strategies be pursued in parallel. Even Marr, despite his injunction to the contrary, was concerned with inferring forms of representation and mechanisms in early vision.

My own approach over the past decade or so has been to look for certain extremely simple mechanisms which we have good reason to suspect that the human visual system must have. These reasons stem from the nature of the tasks that we know the visual system can accomplish, as well as from certain limitations that the visual system appears to have. The initial hypothesis typically comes from attempting to answer the *minimal requirement* question: What must the visual system do early in its analysis of the stimulus in order that it can easily do the things that we know it eventually does well? This, in turn, raises certain related questions: Which visual tasks are carried out in parallel by processes operating over local information across the visual array? Which tasks require serial processing by their very nature (see the discussion by Ullman, 1984) and which happen to be serial because of constraints imposed by the architecture of the visual system? If the processing is done serially, what controls the order in which the operations are applied and where they are applied, and how is the serial information merged into a single percept? Which tasks require that general world-knowledge be brought to bear and which are "modular" and cognitively impenetrable? Which tasks have to be accomplished in order for the visual information to interact across modalities, and especially to provide guidance for the motor system?

Many of these questions relate to how the visual system handles information about space and how it detects certain visual properties and relates them to their location in the visual field. To deal with questions such as these we have developed a highly incomplete but provocative theory of visual attention and visual *indexing* called the FINST theory. Over the past decade or so we have gathered converging evidence for the view that there is an early pre-attentive stage of vision in which the location of properties in space are selected or *individuated* and *indexed* prior to any serial process being applied to the visual array — in fact even prior to assigning focal attention to places in the array. Although the basic ideas are extremely simple, they have to be introduced carefully since they invite misunderstanding of what is being claimed. For this reason I shall devote some space to the explication of the basic assumptions of FINST theory before very briefly summarizing some of the data that supports it. A more complete description of the theory is given in Pylyshyn (1989).

In order to detect simple geometrical properties among elements in a visual scene — properties like being inside, being collinear, forming an equilateral triangle or a square, and so on — the visual system must be able in some way to simultaneously reference more than one item[2], since the relations in question apply synchronically over several items. Even if the elements are scanned in sequence, information about their relative location or their geometry must be available simultaneously at some level. In fact there is evidence, some of which we shall cite later, that rather than being available in some internal representation, the information can be accessed directly from the scene in arbitrary order — i.e., in an order not governed by the location of the items in the scene but by factors internal to the process. In other words, many of our visual tasks are solved not by internalizing some image of a visual scene and operating on it, but by leaving the information where it is in the world and operating directly on it as needed. This suggests that some perceptual problems remain "situated", to use the terminology popular in some quarters. To do this, however, requires at the very least that there be a mechanism that allows multiple access, or potential parallel access to several salient places.

We need to be clear about what it means to have potential parallel access to several places. Having *access* to an item or place is quite different from carrying out an operation on it and is in fact a prerequisite for the latter. The clearest example of an access mechanism is a pointer in a computer data structure. A pointer is a symbol which some processes *can* use to carry out operations on the data referenced by the pointer. A pointer typically occurs as a variable symbol or an argument in a function. Before the function can be evaluated the argument must be bound to some item token, so when the function is evaluated it is the item pointed to that serves as the operand. If the function has several such arguments then we need to have access to all the items; they must all be bound to their arguments, even if the arguments are subsequently treated in

---

[2] I use the neutral term "item" in place of such descriptive terms as "feature" or "feature-cluster" or "visual object". The question of precisely what kind of item can be indexed is an empirical one which is currently being explored experimentally. The present discussion is independent of the outcome of this investigation.

some sequence rather than all at once. This is why we refer to the pointers as providing "potential parallel access".

When we say that the visual system has *access* to items we mean that it has a way to interrogate them or to apply some operation or process to these items without first having to search for them, or more precisely without having to search for items that have certain specified properties. To have potentially parallel access does not require that all items be processed simultaneously — they may or may not be processed in parallel depending on the nature of the task or of the process that operates upon those items. What parallel access does entail is that the access mechanism itself does not constrain the processing to occur in any particular order related to the spatial layout of the items — indeed it does not constrain processing of these items to be serially ordered at all. If the processing is serial then some other constraint related to the process itself would be the determining factor in the order in which the items were treated.

Notice that if access were through a conventional unitary "spotlight" of attention, then the attentional beam would have to be moved from place to place. We would then need a control scheme for moving the attentional beam. A simple and widely adopted scheme is one in which the beam is continuously scanned through space by some analog means. It might be instructed to, say, scan in a certain direction from its current locus. On the other hand if attention can be moved to disparate items in arbitrary order or simultaneously to a set of items in the field, then we would need an account of the mechanism by which attention could be sent directly to some item or to some set of items. The control process would have to pick out or in some way specify the particular item where the attentional beam must move to and that entails a way to refer to or *index* that item.

What this comes down to is that we need a mechanism for allowing items in a visual scene to appear in arguments to both visual predicates (predicates such as $Collinear(x_1,x_2,x_3,\ldots,x_n)$ for example) as well as in commands to allocate processing or to move focal attention or gaze to a certain item. Of course an obvious way of specifying the item is by providing its coordinates in some Cartesian space. Although this would allow us to do the job, it — like some proposed pictorial models of mental imagery (Kosslyn, Pinker, Smith & Shwartz, 1979) — provides both too much and too little information. Cartesian coordinates contain enough information to locate items in an absolute way and to compute absolute distances among them. Since humans do not appear to have access to such absolute metrical information we would have to have some way of blocking the use of coordinates for computing absolute distances and retinal locations. On the other hand, the coordinate-storage option does not allow the visual system to do multiple-object scanning, as we showed in Pylyshyn & Storm (1988). A much better option would be a mechanism that does not have this absolute information in the first place and which provides indexes that track moving objects without storing their coordinates. The indexes that we hypothesize (FINSTs) do exactly that; they provide a way to access or interrogate primitive visual entities without providing absolute coordinates. We shall argue that there is independent evidence for such an indexing mechanism and that such a mechanism has certain interesting side effects.

Before sketching the evidence it may be useful to provide another concrete illustration of how a FINST differs from a spotlight-of-attention and of exactly what is being claimed when we say that FINSTs provide access-without-scanning and access-without-coordinates. We have recently implemented a network model of FINST indexes (Acton & Eagleson, 1993; Pylyshyn, Eagleson & Acton, 1993), largely to show that the claims can be instantiated in a reasonable mechanism. The model takes an activation map as input. No claims are made about the function that takes us from the retinal image to the map, though the map itself is a common hypothesis which appears in Treisman’s Feature Integration Theory (Treisman & Gelade, 1980), Wolf’s Controlled Search Model (Wolf, Cave and Franzel, 1989) and other theories of search and visual attention. The FINST implementation is based on a proposal by Koch & Ullman (1985). Both the Koch & Ullman network and ours use a winner-take-all network for finding the most active item in the activation map. When this element has been identified, activity is transmitted to a global node in the network. (In our case the 4 most active elements are identified and activity is transmitted through 4 separate networks that are kept from interfering with one another through a temporal-priority scheme.) A side-effect of the convergence of the winner-take-all network is that a set of auxiliary shadow-units are also activated. These auxiliary units now allow a signal to be propagated from the global node and routed precisely to the base element which that particular winner-take-all network had settled on. What this simple idea provides is a circuit that (1) selects the most active element in an activation map, and (2) allows that element to be individually pulsed or strobed. Note that the network does not in any way identify or recognize the element or encode its location or any of its properties. It simply provides a way to direct a signal to the item onto which the winner-take-all network converged, and in so doing it allows that item to be addressed by other operations. For example it allows that item to be probed for some of its properties. One way that this could be done is as follows. Suppose that at the same time as the item in question is pulsed, a global property detector (say a redness detector) is also pulsed. If these pulses are both sub-threshold and yet the property detector fires, this establishes that *the very item that was selected by the winner-take-all network has the property RED*. And this conclusion is reached *without any other knowledge about where the element is nor what other properties it has*. The process can of course be repeated for other properties, in this way establishing that the element in question has a certain conjunction of properties. This way of probing for properties of particular items is related to Feldman and Ballard's (1982) proposal for dealing with the cross-talk problem in recognizing sets of co-located properties, and may explain why conjunction search requires that items be visited serially.

Our network model allows 4 different items to be strobed in this way and also provides a way for the individual strobe-paths to the 4 most active items to maintain consistency by tracking the *same* elements over time — i.e. it provides a way to keep the items always distinct and to keep them consistently indexed as they move about. The "stickiness" of the index binding in this network is the result of a low-level locally-computed motion extrapolation filter (e.g., a Kalman filter) associated with each indexed item, which enhances the predicted location of that item. While many of the design choices in this model are provisional, the model illustrates what we mean when we say that a FINST is purely an index: It allows the element that it indexes to be tracked and strobed or interrogated, and it can be used to direct operators to particular elements in a visual scene. It is a way of telling operators or processes *which* elements to operate over. And it does so while making minimal assumptions and providing no more information to subsequent processes than the absolute minimum needed. It is a truly pre-processing indexing or binding mechanism.

Some empirical evidence for particular properties of the indexing mechanism.

The FINST hypothesis is really a series of proposals for a primitive mechanism which operates on an early somatotopic map (or maps) — such as Marr's Primal Sketch or Treisman's Feature Map or Wolf's Activation Map — and which precedes the allocation of focused attention. It has wide implications, some of which are spelled out in Pylyshyn (1989). We have been engaged in testing some of the principle ways in which the FINST idea differs from the current unitary attention (spotlight beam) views. The main empirical prediction from the FINST hypothesis that we have been examining is that it is possible to arrange for a small number of items to be directly (and therefore rapidly) accessed by preselecting them, and hence assigning FINSTs to them, in one of several ways — either exogenously by making those items salient (say by abrupt onset) or endogenously by providing a temporary cue to allow subjects to pick them out. These indexes need not be assigned to contiguous items or even to nearby items, but can get assigned to several disparate items scattered throughout the visual field. Consequently we expect that items so selected will be accessed without requiring a search over the display (though the indexed items themselves may be visited serially), whereas other items in the display will have to be located by scanning and searching the display. Another assumption of the FINST theory, as it stands at the present time, is that a FINST index is “sticky” and once assigned will tend to remain with the item to which it has been assigned (though not necessarily without effort nor without having to be periodically refreshed) even when the latter moves about.

If indexes provide access to a number of items across the visual field, then one of the consequences should be that it is possible to select several visual elements and treat them as though (a) they were distinct — e.g., as though each had a different colour, and (b) they were essentially (to a first approximation) the only ones in the display. This is in contrast with the view which says that attention is not only unitary, but has to be scanned continuously from place to place to locate items and/or zoomed to cover smaller or larger regions.

Among the studies we have carried out in recent years to support these ideas are ones showing that subjects can track multiple independent moving targets in a field of identical distractors, that their ability to track these targets and detect changes occurring on them does not generalize to nontargets nor to items lying inside the convex polygon that they form (so that a zoom-lens of attention does not fit the data). We have also used a visual search paradigm to show that (serial or parallel) search can be confined to a subset of indexed items and the layout of these items is of little importance. We have also carried out a large number of studies on the phenomenon known as subitizing and have shown that subitizing occurs only when items can be preattentively individuated and in those cases location precuing has little effect, compared with when counting occurs, which suggests that subitizing may be carried out by counting active indexes rather than items in the visual field. And finally we have run some recent studies showing that a certain motion effect which is sensitive to attention can occur at what appears to be multiple attention loci, or what in our terminology are multiple indexed items. We describe these briefly in turn, but the reader will have to refer to the referenced papers and reports for details.

Multiple-object tracking studies.

The first direct experimental test of one of the assumptions of FINST theory was a series of studies employing the multiple object tracking task (Pylyshyn & Storm, 1988). In this task subjects visually tracked a prespecified subset (3 to 6) of a larger number of identical, randomly moving objects (+ signs) in a display. The items in the subset to be tracked (the targets) were identified prior to the onset of movement, by flashing them on and off several times. While in motion the targets were indistinguishable from the non-target items, which made the historical continuity of each target's motion the only mark of its identity. After a predetermined interval of tracking (between 7 and 15 secs) a square was flashed on the screen and subjects indicated whether the flash had occurred on a target object, a non-target object, or somewhere else in the display. Accuracy was surprising high in this task, although it decreased with increasing numbers of targets tracked.

In one study which had 4 targets and 8 distractors a detailed analysis was performed to determine whether a serial process involving scanning a spotlight of attention might have reproduced the observed results. A round-robin scan-and-update strategy was simulated. In this simulation, locations of targets were stored in a table, starting with their initial positions. A single attentional beam was continuously scanned from one stored target location to the next, and the object nearest that location was found and taken as the target (of course this could actually be a nontarget and thus later lead to an error). The actual coordinates of this item were then stored as the updated coordinates for the next cycle. The simulation was carried out using the actual dynamic displays that were used in the multiple-object tracking study and was run with different assumed attention scan velocities, as well as several location-prediction and guessing strategies. In all cases the performance of the simulation asymptotes at around 45% at scan speeds below about 100 deg/sec, and increases to 50% at scan speeds as high as 250 deg/sec — the highest rate we have been able to find in the scan literature. Both of these performance figures are far below the 87% correct mean identification rate actually observed. It was concluded that even with the fastest reported scan speed it was not possible for the task to be carried out at the level of performance actually observed without some parallel tracking. In addition, we computed predicted performance levels that might be expected if subjects were tracking some subset of n objects at a time and guessing for the remainder (this was based on the conservative assumption that subjects knew which items they were correctly tracking). The result suggested that the number of objects that would have to be tracked in parallel was most likely at least 4.

The multiple object tracking paradigm has been used many times in various laboratories and the basic findings of Pylyshyn & Storm (1988) have been confirmed (e.g., Yantis, 1992). In our own laboratory we have found that subjects can simultaneously track several independently moving objects under a wide variety of conditions. For example, McKeever & Pylyshyn (1993) used a variant of the tracking paradigm that minimized the opportunity for subjects to employ a successful guessing strategy. In these experiments subjects tracked 3 or 4 target items and had to identify *all* the tracked items at the end of every trial. Performance continued to be very much higher that expected by chance under these conditions.

Among the findings that were not predicted by the pure version of the FINST theory is the fact that performance depends on the number and nature of non-targets, as well as on certain restrictions on the trajectories of the targets (Yantis, 1992) — contrary to the assumption that individual indexed items are independent and the assumption that non-indexed items are not processed (since only indexed items can be addressed by further processes). In fact both McKeever & Pylyshyn (1993) and Sears & Pylyshyn (1993) found that performance deteriorated when there were more nontargets. If indexes allow one to filter out unindexed objects then the filter appears to leak — much as traditional attentional filters have generally been found to leak. But it is possible that there may be a more principled explanation for this apparent leakiness.

Mckeever & Pylyshyn argued that the decrease in performance with increasing number of distractors might be attributed to both an increase in the number of cases in which an index was transposed to a nontarget, and to the operation of a post-tracking process, such as an error-recovery routine. There is reason to believe that the tracking *task* — which involves much more than just indexing moving objects — is effortful and that this is the case because even though indexing an object may be preattentive, maintaining the index requires effort inasmuch as it involves warding off competing events that would take the index away to another object and may perhaps even involve periodically refreshing the index to prevent inhibition or decay.

Suppose subjects are able to detect when a target has been lost — at least on some significant proportion of trials. Then they might attempt to recover the lost target by searching for the most likely item that *might* be the lost target. In that case various factors such as item density and the particular visual properties of the objects on the screen, along with perhaps configural properties of the target set and predictability of their relative motions, might be expected to play a role in the recoverability of the lost target. In particular, the more nontargets there are in the vicinity, the poorer the recovery might be, and the more distinct the nontargets the better the recovery might be expected to be as well. The fact that McKeever & Pylyshyn found no decrease in performance with increase in number of nontargets under conditions when nontargets were visually distinguishable from the targets, adds some support to the view that the decreased performance may have been due to a post-tracking error-recovery stage. The same might plausibly be said of the role of the “convexity constraint” in the motions of targets found by Yantis (1992). These phenomena are very likely all the result of some post-index stage of the tracking task. For example, at such a stage there may be mechanisms and decision strategies which not only help in error recovery, but could also facilitate the tracking task by anticipating its movement: by shadowing the actual tracking in an internal model of the display. Indeed, McKeever & Pylyshyn propose just such a model-updating scheme, combined with error-recovery, as does Yantis — although Yantis’ proposal does not view this process as ancillary to the indexing mechanism itself.

The multiple-item tracking studies also affirmed one other salient property of FINST index theory not shared by attention-beam theories; that is the prediction that the visual system may be able to rapidly access discrete items distributed in space without being able to similarly access points in between. An attention beam offers enhanced processing over a single contiguous region, although several attentional-beam theorists have postulated that the scope of the attentional beam can be varied. This so-called zoom-lens view was forced upon attention theorists by a considerable amount of evidence suggesting that in some tasks subjects are able to attend to more widely dispersed cues than in other tasks (Eriksen & St. James, 1986). What such a zoom-lens view remains committed to (along with all other unitary-attention theories), is the idea that attention applies to a contiguous region, even though it might alter its scope (presumably with some decrease in the resources available at each point within the region covered by the attentional beam).

The assumption of a unitary contiguous region of attention being the range of application of all visual processes is not shared by FINST theory. Indeed, the FINST idea is based on the assumption that a number of distinct and relatively punctate filled places are simultaneously indexed, making it possible for subsequent processes to access them without accessing intervening places. In the Sears & Pylyshyn (1993) tracking studies, direct evidence was found for this assumption. Their data showed that the detection of form changes was enhanced (in terms of latency measures) only on the actual items being tracked, and not in the region bounded by the tracked target items. The detection latencies for objects lying within the convex polygon region defined by the targets were no faster than those for objects lying outside this region. A similar result was also reported by Intriligator & Cavanagh (1992), who used a variant of the multiple object tracking task involving only two targets moving in a rigid configuration. They reported that while detection latencies for the tracked items decreased, latencies for items between the two tracked items was no faster than elsewhere.

Thus multiple-item tracking studies provide strong support for one of the more counterintuitive predictions of FINST theory — viz, that the identity of items can be maintained by the visual system even when the items are visually indiscriminable from their neighbors and when their locations are constantly changing. Moreover it appears that it is the indexed objects themselves and not some contiguous region which contains them, that is selected. The discrete nature of the indexing mechanism, as well as the FINST assumption that several items can be indexed in parallel, was also demonstrated clearly in a series of quite different studies by Burkell & Pylyshyn (1993) involving stationary objects and a rapid-search paradigm described below.

Multiple-cue studies.

In a series of studies, Burkell & Pylyshyn (1993) showed that a number of disparate items could be precued from among a larger set of similar items and the precued subset could, in a number of important ways, be accessed by the visual system as though they were the only items present. The studies also showed that *all* precued items (of which there were 2 to 5) were available — that it was not a case that improved performance in the cued condition arose from sampling from the subset nor of scanning and searching for the items. The data also showed that cued items further apart did not produce longer access latencies. These results are incompatible with the proposal that items are accessed by moving around a single spotlight of attention. Instead they provide strong evidence in favor of primitive multiple indexing mechanisms such as FINSTs.

The experiments all involved a visual search paradigm (Treisman & Galade, 1980). In all cases subjects were presented a set of items (totaling 12, 15 or 24), together with a target of the sort that would define a *Conjunction Search* condition. In such a condition, items vary on pairs of properties (left-vs-right oblique lines, red-vs-green colors), and the target is an item which shares each of its properties with at least one other member of the set — so that it takes a conjunction of properties to identify the target. Many investigators have shown that in such conjunction search tasks the time to locate the target increases linearly with the size of the search set, with a slope of about 30 or more msec/item in the exhaustive search case when there is no target present and about half that when a target is present (although the exact slope varies a great deal with type of properties used for the disjuncts), thus suggesting a serial, self-terminating search in these cases. If the target was defined by the presence of a single feature (the *Feature Search* condition) the time to locate it is relatively insensitive to the number of items in the search set: the slope is generally found to be only around 5-10 msec/item and the target is said to "pop out" from the background distractor set. Under these conditions the search is often assumed to be parallel and preattentive since a slope of under 10 msec/item is faster than any known scan-search process. The precise difference between single-feature and conjunction search tasks is not important for the present purpose; all that matters is that they do differ markedly, and that the single-feature search condition shows the "popout" phenomenon, so that if there is any search it is extremely rapid and therefore it is unlikely that the items are searched by a serial scan process.

In the Burkell & Pylyshyn studies, if it were not for the precues the experiments would all be of the conjunction-search type. The task was to indicate whether there was a target among the cued items. All items were preceded by place markers. Cuing was accomplished by the sudden onset of place markers for the cued subset since Yantis & Jonides (1990) showed that abrupt onsets were particularly effective in attracting automatic attention. We found that precuing a subset of 3-5 items resulted in a considerable speedup of search time. Exactly how much the search was speeded up depended on the nature of the subset. The subset itself could constitute either a feature or a conjunction set — the precued items could differ from the target in only one feature or they might share each of two features with the target so that it would require the conjunction of two features to specify the target from among the cued subset items. The reliable finding was that the time to locate the target was significantly longer when the subset was a conjunction subset than when it was a feature subset.

In one of the experiments the size of the cued subset was varied from 3 to 5. When the subset was a feature search set the slope of the latency vs cued set size was found to be between 9 and 18 msec/item regardless of the cued set size. When the set was a conjunction search set, the slopes were 57 msec/item when the subset did not include a target, and 37 msec/item when there was a target in the subset. Recall that finding the target within the entire set of items always constituted a conjunction search. Consequently the fact that within the cued subset we find the same difference between feature and conjunction search as occurs in the basic search paradigm shows that the subset was being treated as the entire search set. It appears that the noncued items were being ignored, except for a general increase of RT relative to the control case (also examined) where the noncued items were actually absent from the display. This constant increase, called "cost of filtering" by Treisman, Kahneman & Burkell (1983), is expected whenever some aspect of a display has to be filtered out.

Note that in order to do this task subjects had to keep track of and access all the items in the cued subset. Even if unitary attention had to visit each item in the subset, membership in the subset had to be kept track of since it had been marked only by a transitory event (onset of position markers corresponding to the cued subset appeared 100 msec before the subset items themselves and then disappeared). There is no way to do the task of indicating whether the subset contains a target without preselecting all and only the items in that subset, particularly since in some of the experiments a target was actually present among the noncued items, though it was not to be counted as a target in that case. Moreover, the only way that the observed difference between the feature and conjunction subset could arise is if the cued subset was being treated as the search set. The fact that the slope for the feature subset was so shallow as to constitute "popout" also suggests that the items in the cued subset were being probed in parallel, in the kind of "registration" process that Treisman & Gelade (1980) site in explaining popout in feature search tasks in general. One way that this could happen is if the cued items were being simultaneously strobed or activated and a logical "or" of the outputs of the relevant feature detectors observed, along the lines already alluded to in the earlier discussion of our network model of indexing.

Another finding of the Burkell & Pylyshyn studies was that the latency on neither the feature nor the conjunction subsets increased with increasing distance among the cued items. By systematically manipulating the dispersion it was possible to measure RT as a function of mean distance. This RT did not increase with increasing distance as predicted by a scanning attention-beam model — in fact the RT actually decreased slightly, for reasons that are unclear though perhaps related to the diffusion of the effects of FINST operations and some ensuing interference among the closer indexes.

Subitizing studies.

Another set of experiments investigated a phenomenon referred to as "subitizing" wherein a small set ($n<5$) of items can be enumerated rapidly and accurately under certain conditions. There is good reason to believe that the process involved in subitizing is different from that involved in counting more than 4 or 5 items. We believe that this difference can be accounted for if we assume that the FINST mechanism is being deployed in a direct way in subitizing. A number of experiments (summarized in Trick & Pylyshyn, 1993a; 1993b) were carried out which provide evidence for the view that a small number of indexes are assigned to primitive distinct features (popout features) and that subitizing is accomplished by merely counting the number of active indexes, without having to spatially scan attention from one item to another. Two kinds of evidence support the claim that subitizing relies on preattentive information that can be obtained from FINST indexes whereas counting requires focal spatial attention. First, whenever spatial attention is needed to compute a spatial relation (c.f., Ullman, 1984) or perform feature integration (c.f., Treisman & Gelade, 1980), subitizing does not occur (Trick & Pylyshyn, 1993a). Second, the position of the attentional focus, as manipulated by location cues, has a greater effect on counting than on subitizing (Trick & Pylyshyn, 1993c; Trick & Pylyshyn, 1988).

The first set of experiments was designed to show that subitizing is not possible when the enumeration task is one in which item individuation requires attentive processing. One of the few earlier published studies that failed to produce strong evidence of subitizing had subjects enumerating concentric circles (e.g., Saltzman & Garner, 1948). However, stimuli involved in these failed subitizing cases have some rather special characteristics: the nearest and most similar contours come from different items, the items are necessarily of different sizes, and being concentric they share a common center. A number of experiments were performed to sort out which factors were responsible for the failure of subitizing in these cases. They provided clear evidence of subitizing when stimuli were squares that varied in size but no evidence of subitizing when the squares were concentric — in the latter condition the slope in the 1-3 range was approximately the same as the slope in the 5-7 range.

To exclude the possibility that the result arose because contours were closer together in the *Concentric* condition (thus perhaps resulting in lateral masking), additional studies were performed in which subjects were required to enumerate the straight lines and right angles that made up the sides and corners of the concentric rectangles. All subjects were able to subitize both corners and lines, even though the corners were of uniform size and the line lengths varied by a factor of 30. In fact, there was no significant difference between the latencies to count parallel lines and corners, and most particularly not in the subitizing range. Moreover, both subitizing slopes were within 2 msec of those for uniformly sized rectangles.

Additional studies investigated the ability to subitize when individuating items in the subitizing task required the serial computation of a spatial relation. For example, according to Ullman (1984) and (Jolicoeur, 1988), serial attentive processes are required to compute the “connected-to” relation. Based on this observation, a study was carried out in which subjects were presented with a winding contour superimposed over a number of colored blocks some of which had to be enumerated. In any display subjects were required to enumerate 1-8 blocks that were designated as targets while another 2-8 blocks that served as distractors. In the *Connected* condition subjects were required to enumerate items on a particular contour. Contours could be of three different lengths, ranging from short to long: 4 link, 5 link and 6 link. Distractors were defined as blocks that occurred after the break in the contour or on the orthogonal contour. In the *Color* condition subjects were shown the same displays, but their task was to enumerate items of a particular color regardless of which contour it was on. Attention is not required to detect an item of a different color from other items; color is assumed to be a primitive feature (e.g., Treisman & Gelade, 1980). Because preattentive information distinguished target items from distractors, subitizing was predicted in the *Color* condition. In contrast, attention is required to compute the connected relation so subjects should not be able to subitize the subset of connected items.

The result provided clear evidence of subitizing in the *Color* condition and no evidence of subitizing in the *Connected* condition. Moreover, in the *Connected* condition, though not the *Color* condition, latencies were affected by the length of the contour. Thus, subitizing does not occur when a spatial relation which forced serial processing was required to identify and individuate items in an enumeration task.

Another set of studies involved rapid visual search. As we have already seen in the discussion of the Burkell & Pylyshyn (1993) cued search experiments, attention is not required in order to detect the presence of an item that differs from others by a single primitive feature (feature search) or by a disjunction of features (disjunction search), but is required in order to locate an item that differs from others in the display by a conjunction of features (conjunction search). Consequently this search task provides another way to test for whether subitizing occurs when attention is required for individuation the items in question. In Trick and Pylyshyn (1993a) a search task was superimposed on an enumeration task. Subjects had to enumerate items in a field of distractors. There were two conditions. In the *Disjunction* condition subjects had to enumerate white or vertical lines in green horizontals. In the *Conjunction* condition subjects had to enumerate white vertical lines among green vertical and white horizontals.

Subitizing always occurred when there were no distractors in the display. But subjects were capable of subitizing even with 12 and 20 distractors in the *Disjunction* condition. In contrast, there was little evidence of subitizing in the presence of distractors in the *Conjunction* condition. The slope of the RT vs number of distractors function also differed markedly in the two conditions. Each distractor added approximately 65 msec to the time to enumerate an item in the *Conjunction* condition, whereas each distractor only added 6.2 ms in the *Disjunction* condition, suggesting serial processing of the *Conjunction* condition and parallel processing of the *Disjunction* condition.

From these studies it appears that subjects are able to subitize targets which are distributed among distractors when spatially serial analysis is not required to distinguish targets from distractors, as in the *Disjunction* search condition, or the color block condition of the previous experiment, but are incapable of subitizing targets among distractors when spatially-serial attentive analysis is required to distinguish targets from distractors, as in the *Conjunction* condition or the connected contour condition of the previous experiment.

These studies support the idea that subitizing uses a preattentive mechanism which indexes items to be subitized and which therefore does not require scanning. Another way to test whether attention scanning or attention zooming comes into play in these low cardinality enumeration phenomena, is by manipulating attentional focus or attentional spread in advance of enumeration. One way to manipulate where attention is focused, and perhaps also how widely it is focused, is by using a “cue validity” paradigm. In cue validity studies subjects are required to make some perceptual discrimination under different conditions of prior knowledge. In the *Valid Cue* condition they know beforehand both when and where a stimulus will appear; in the *Neutral Cue* condition they only know when the stimulus will appear (Neutral cuing); and in the *Invalid Cue* condition subjects are given incorrect information about where it will appear. Typically, subjects are faster and more accurate at making perceptual discriminations if they are given correct information about where the stimulus will fall. Performance is best in the *Valid Cue* condition, followed by the *Neutral* and *Invalid Cue* conditions. This finding has been interpreted as evidence that a processing focus, the “spotlight of attention” is moved through the stimulus array in response to subjects’ expectations about where the target item will appear.

A number of studies were carried out combining a cue validity paradigm with an enumeration task. The goal was to show first that subitizing would be possible whether attention was focused on a small area, as in the *Valid Cue* condition, or distributed throughout the display, as in the *Neutral* condition. If this were true then it would show that subitizing is not prevented when the attentional focus is narrowed. A second goal was to show that the position of the attentional focus would have a stronger effect on counting latencies than on subitizing latencies. Specifically, the difference between *Valid Cue* and *Invalid Cue* conditions should be more pronounced in the counting range than the subitizing range. This result would be expected if the counting process involves the attentional focus, and moving the attentional focus takes time. The position of the attentional focus should have a smaller effect in the subitizing range because subitizing doesn’t require the attentional focus.

In these studies subjects were required to count 1-8 dots. Colored rectangles were used to cue the area in which dots were to appear. In the *Neutral* condition all the rectangles were the same color. In the *Valid Cue* and *Invalid Cue* conditions, one rectangle was a different color from the others. In the *Valid Cue* condition the position of this rectangle predicted the position of the dots with 80% accuracy. The other 20% of the cases constituted the *Invalid Cue* condition. The results were quite clear. In all experiments, and for all subjects, subitizing was evident in *Valid, Invalid and Neutral* conditions. When cue validity had an effect, it was always a stronger effect when there was a large number of items. For example, when colored cuing rectangles were used there was no significant effect of spatial cuing in the 1-4 range, although there were significant effects in the 5-8 range. The average difference between invalid and valid latencies was 23 ms in the 1-4 range, as opposed to 125 ms in the 5-8 range. Consequently, the position of the attentional focus, as manipulated by spatial pre-cues, seems to have a greater effect in the counting range than the subitizing range, as would be expected if counting requires spatial attention. Moreover, the necessity of contracting the attentional focus in the *Valid Cue* and *Invalid Cue* conditions, relative to the *Neutral* condition, did not prevent subitizing, or even restrict the subitizing range.

Indexing and the line-motion illusion.

Finally, we have recently carried out a series of experiments (Schmidt & Pylyshyn, 1993) which use a line-motion illusion reported by Hikosaka, Miyauchi and Shimojo (1991), which is believed to be attention-sensitive. This technique provides another way to investigate the limits on multiple-locus processing in a visual field using a perceptual effect which minimizes the cognitive component of the task.

The illusory line-motion phenomenon occurs when attention to a target induces the perception of motion of a line which is suddenly presented with that target as an end point. In its simplest form subjects are asked to fixate a marker. A trial begins as a sudden onset cue occurs (and presumably draws attention to itself). After a brief ISI, a line is instantaneously drawn with the onset cue as one if its end points. Subjects consistently report that the line was “drawn” away from the cue.

In a series of studies (Fisher, Schmidt, and Pylyshyn, 1993), subjects fixated the central point in a high-speed low persistence calligraphic display. A number of locations in the display were cued with onset cues and then randomly probed for the occurrence of the illusion. Using this method it is possible to determine whether a number of such foci can simultaeously be effective in producing the illusion, and whether there are limits on this number. In the first experiment, the displays consisted of 1 to 8 cues evenly spaced around an imaginary circle. Subjects maintained gaze on a fixation point in the center of the display and all the cues onset simultaneously for 250 msec. After the cues were extinguished a line was drawn rapidly from the fixation point out to the circle’s circumference (i.e., in the direction opposite to that of the illusory line motion). On half the trials the line was drawn to the location where a randomly chosen cue had previously onset and on the other half of the trials the line was drawn to a location where there had been no onset cue (i.e., between where two cues had been). The line illusion was observed significantly more often in the former case than in the latter, demonstrating both that illusory direction of motion is less likely to be observed where there was no onset cue and also demonstrating that several cues could serve simultaneously to produce the illusion.

There was a strong decline in the frequency of the illusion once a certain minimum number of cues was reached (the number appears to be around 4 – 6, though we are still attempting to refine this estimate mathematically), providing support for the notion that a limited number of loci could operate simultaneously in producing the illusions. One straightforward way to account for both the multiplicity of effective loci and the limit to the number that can serve in this way is to assume that the onset cues are drawing FINST indexes and that the presence of such an index is either itself sufficient to produce the illusion, or else allows focal attention to be transferred rapidly to that location when the line is drawn and this indirectly produces the illusion. More careful timing and titration of these effects may sort out these alternatives.

A number of additional studies were also carried out using the line-motion illusion. They show that indexes may be assigned to items that become more distinct (e.g., brighter) in relation to their environment even though these items themselves do not change. Thus we get the illusion most strongly to items that become momentarily brighter that others in the display, and also to ones that remain bright when the other items briefly become dimmer. We also get the illusion most strongly to the item type which is numerically in a minority in the display consisting of two types of features (e.g., horizontal and vertical bars) and this effect reverses when the same item type becomes the most common one in the display. These and other studies currently under way may help to clarify the conditions under which items may get indexed and also the properties that accrue to indexed items.

Conclusions

The larger motivation for this research continues to be to discover some basic properties of the cognitive architecture — basic mechanisms that allow cognitive processes to be realized. We have chosen to investigate the early stages of the sensory interface with spatial information. Although the principal theoretical idea concerns an early stage in vision, the ultimate goal, as sketched in Pylyshyn (1989), is to integrate this mechanism with spatial cognition across modalities, including the motor system. As applied to vision, however, the hypothesis is that prior to the allocation of limited attentional resources by, a mechanism must first individuate a limited number of items in the visual field, maintain their individuality independent of their retinal position, and provide a way to directly access them for subsequent processing. We have presented evidence from several different areas which strongly suggests that whatever the facts may be concerning a unitary locus of processing, there is more going on in spatial access than the "single spotlight" view provides. In particular there must be a number of disparate, noncontiguous loci selected for special treatment by the early preattentive visual processor. These loci must be available potentially in parallel — i.e., they must be such that a parallel process could access them synchronically, much the way the retinal map is available and does not itself impose temporal constraints on accessing it. The loci are probably relatively localized (punctate), though little is known about their extent nor the way in which they may interact or inhibit one another if they are spatially close together.

The evidence we have presented comes from multiple object tracking studies, cued search studies, subitizing studies, and illusory line motion studies. All this evidence converges on several basic properties of visual spatial attention which implicate the FINST indexing scheme. One is that it is possible to track about 4 randomly moving objects and to keep them distinct from visually identical distractors, so that events taking place on the tracked targets can be quickly detected and identified. While the data are compatible with there being a detection/identification process which serially visits each indexed location, the data are also univocal in showing that the tracking itself could not be happening by a process of scanning attention across space from one object to another. They are also clear in showing that attention is not merely broadened to include a wider scope, since the advantage does not accrue to items within the general region occupied by the target items but only to the target items themselves.

The cuing studies go further in showing that several (up to 5) items can be precued from among a larger set and that the cued items can be treated by the visual system as though they were the only ones in the scene. The selected set is searched in parallel (in the feature search condition) or in serial (in the conjunction search condition) whenever they would be so searched if they were the only items present. These studies also showed that if items of the precued set are visited serially, they are not searched for by a scanning process, inasmuch as greater spatial dispersion does not lead to slower responses. So once again neither a spatial scan view of access nor a zoom lens view of access fits the evidence. Something else is going on and we suggest that it is the availability of FINST indexes which allows direct access to the indexed subset.

The subitizing studies provide yet another body of converging evidence leading to the conclusion that a small set of direct-access links are available and that these are computed preattentively and in parallel by the early vision system. When the items were ones that “pop out” within a set — so that attentional scanning and searching is not required for individuating them — then these items can be subitized up to a set size of about 4 items. The directness of this access is confirmed both by the difference in enumeration speed for small set sizes when serial attention is and is not required, and also by evidence that subitizing is less sensitive to location precuing than is counting.

And finally we presented some preliminary evidence that for at least one visual illusion that is sensitive to locus of attention the illusion can be controlled simultaneously at several disparate locations — i.e., the illusion acts as though there were up to 6 loci of attention.

No one type of evidence is conclusive. The FINST idea, while extremely simple, and we believe plausible prima facie, is also a proposal concerning a fundamental preattentive visual mechanism which is never observed directly. The tasks described herein all require much more than an indexing mechanism to produce a response. They involve decision and enumeration and discrimination and response selection stages, all of which are likely to contain serial components. Thus performance on these tasks — even when they demonstrate multiple and dispersed loci of processing advantage — can always be covered by some additions to a spotlight view. Nonetheless, we submit that taken as a whole the evidence is most parsimoniously accounted for in terms of the hypothesis that there is an early stage in processing when a small number of salient items in the visual field are indexed and thereby made available through a primitive index-binding mechanism for a variety of visual tasks.

References

Acton, B, and Eagleson, R. (1993). A Neural Network Model of Spatial Indexing. *Investigative Ophthamology and Visual Science, 34*(4), Item 413, p.784.

Atkinson, J., Campbell, F., & Francis, M. (1976). The magic number 4 +/- 0: A new look at visual numerosity judgements. *Perception,* 5, 327-334.

Burkell, J.A. & Pylyshyn, Z.W. (1993) Indexing multiple loci in the visual field: Evidence for simultaneous facilitation in visual search. Centre for Cognitive Science, University of Western Ontario, Technical Report: Cogmem 64.

Burkell, J. (1984). Perception of isoluminant colour changes. Master's Thesis, University of Western Ontario, London, Ontario.

Egly, R., & Homa, D. (1984). Sensitization of the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 778-793.

Eriksen, C., & St. James, J. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics, 40*(4), 225-240.

Feldman, J.A. & Ballard, D.H. (1982). Connectionist models and their properties. *Cognitive Science, 6*, 205-254.

Fisher, B. D., Schmidt, W. C., & Pylyshyn, Z. W. (1993). Multiple abrupt onset cues produce illusory line motion. *Investigative Ophthamlology and Visual Science, 34*(4), 1234.

Fodor, J. & Pylyshyn, Z. (1981). How direct is visual perception: Some reflections on Gibson's 'ecological approach'. *Cognition, 9*, 139-196.

Fodor, J., and Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28,* 3-71.

Frick, R. (1987). The homogeneity effect in counting. *Perception and Psychophysics, 41*(1), 8-16.

Hamilton, W. (1859). *Lectures on metaphysics and logic, Vol. 1*. Edinburgh: Blackwood.

Hikosaka, O., Miyauchi, S., & Shimojo, S. (1991). Focal visual attention produces motion sensation in lines. *Investigative Ophthamology and Visual Science, 32*(4), 176.

Hunter, W. & Sigler, M (1940). The span of visual discrimination as a function of time and intensity of stimulation. *Journal of Experimental Psychology, 26,* 160-179.

Intriligator, J. & Cavanaugh, P. (1992). An object-specific spatial attentional facilitation that does not travel to adjacent spatial locations. *Investigative Ophthamology and Visual Science, 33*(4), Item 2849, p 1263.

Jonides, J. & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics, 43*(4), 346-354.

Jevons, W. (1871). The power of numerical discrimination. *Nature, 3,* 281-282.

Jolicoeur, P. (1988). Curve tracing operation and the perception of spatial relations. In Z. Pylyshyn (Ed.) *Computational processes in human vision: An interdisciplinary approach*. Norwood, NJ: Ablex Publishing.

Jolicoeur, P, Ullman, S., & Mackay, L. (1986). Curve tracing operations and the perception of spatial relations. *Memory and Cognition, 14*, 129-140.

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

Kaufman, E., Lord, M., & Reese, T. (1949). The discrimination of visual number. *American Journal of Psychology, 62,* 498-525.

Klahr, D. (1973a). Quantification processes. In W.G. Chase (ed.) *Visual information processing*. (pp 3-34). New York: Academic Press.

Marr, D. (1982). Vision. San Francisco: W.H. Freeman and Co.

McKeever, P. & Pylyshyn, Z. (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.

Miyauchi, S., Hikosaka, O., & Shimojo, S. (1992). Visual attention can be assessed by illusory line motion sensation. *Investigative Ophthamology and Visual Science, 33*(4), 1262.

Pylyshyn, Z. (1981). The imagery debate: Analogue media *versus* tacit knowledge. *Psychological Review, 88*, 16-45.

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

Pylyshyn, Z. (1984). *Computation and Cognition: Toward a foundation for cognitive science*. Cambridge Mass: MIT Press, A Bradford Book.

Pylyshyn, Z. (1988). “Here” and “There” in the visual field. In Z. Pylyshyn (Ed.), Computational Processes in Human Vision: An Interdisciplinary Perspective. Norwood, N.J.: Ablex Publishing.

Pylyshyn, Z. (1991). The role of cognitive architecture in theories of cognition. In K. VanLehn (Ed.), Architectures for Intelligence. Hillsdale: Lawrence Erlbaum Associates Inc.

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

Saltzman, I., & Garner, W. (1948). Reaction time as a measure of the span of attention. *Journal of Psychology, 25,* 227-241.

Schmidt, W. C., & Pylyshyn, Z. W. (1993). Multiple location access in vision: Evidence from a line-motion illusion. Centre for Cognitive Science, University of Western Ontario, Technical Report: Cogmem 66.

Sears, C. R., & Pylyshyn, Z. W. (1993). Multiple Object Tracking and Visual Processing. Centre for Cognitive Science, University of Western Ontario, Technical Report: Cogmem 67.

Shrager, J., Klahr, D., & Chase, W. (1982). Segmentation and quantification of random dot patterns. Paper presented at the 23rd annual meeting of the Psychonomics Society.

Taves, E. (1941). Two mechanisms for the perception of numerousness. *Archives of Psychology*. New York. No. 1941.

Trick, L. & Pylyshyn, Z. (1988). When subitizing fails: the importance of preattentive item indexing for subitizing. Centre for Cognitive Science, University of Western Ontario, Technical Report: Cogmem 35.

Trick, L. M., & Pylyshyn, Z. W., (1993a). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance, 19*(2), 331-351.

Trick, L. M., & Pylyshyn, Z. W., (1993b). Why are small and large numbers enumerated differently? A limited capacity preattentive stage in vision. *Psychological Review*, in press.

Trick, L. M., & Pylyshyn, Z. W., (1993c). Cuing and counting: Sometimes the spotlight isn’t much help. Submitted.

Treisman, A. (1985). Preattentive processing in vision. *Computer vision, graphics and image processing, 31*, 156-177.

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

Treisman, A., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the cost of filtering. *Perception and Psychophysics, 33*, 527-532.

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

Van Oeffelen, M., & Vos, P. (1982a). A probabilistic model for the discrimination of visual number. *Perception and Psychophysics, 32*(2), 163-170.

Warren, H. (1897). The reaction time of counting. *Psychological Review, 4*, 569-591.

Wolfe, J., Cave, K., & Franzel, S. (1989). Guided search: An alternative to the Feature Integration model for visual search. *Journal of Experimental Psychology: Human perception and performance, 15*(3), 419-433.

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

Yantis, S. & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception & Performance, 16*(1), 121-134.