Multiple parallel access in visual attention

Pylyshyn, Z., Burkell, J., Fisher, B., Sears, C., Schmidt, W., and Trick, L.[1]
Center for Cognitive Science
University of Western Ontario

It has generally been assumed that there is a single focus of attention, representing the current locus of visual processing, which can be moved across the visual field independent of eye movements. The precise details of this “spotlight” view vary among investigators, although it is universally thought to apply to a single contiguous spatial region. There is some disagreement as to the possible shape of the attended region (whether circular or toroidal), whether the extent of the spotlight can be varied by “zooming”, under what conditions — voluntary or automatic — the spot can move, whether the spotlight must move continuously, and if so at what speed, or whether it can skip from place to place, and precisely what processing advantage accrues at the current locus. Despite these differences, the idea that a single locus is involved in allocation of processing resources seems widely accepted. And yet there is good reason to doubt that this is the entire story of how spatially local information is accessed in the visual field. In particular, there is reason to believe that even if attention is unitary and spatially focused, there is also a more primitive mechanism for simultaneously indexing several places in a visual field, thus individuating these places and making them directly accessible for further processing. There are both general considerations and some direct experimental data that are relevant to this issue.

General considerations suggesting the need for a multiple-locus indexing mechanism.

In order to detect simple relational geometrical properties in a visual scene — properties like insidedness, collinearity, and so on — the visual system must be able to in some way simultaneously reference more than one place or feature or object in the scene, since the relations

---

[1] Authors are members of the laboratory who contributed directly to the work reported herein. In addition we would like to thank Roy Eagleson whose theoretical analysis helped keep us on track, to Paul McKeever who contributed to the empirical work before leaving for other pursuits and to Brian Acton who has spent a lot of energy developing a compuitational model of these ideas. Lana Trick is now at the University of British Columbia and Kwatlyn College in Vancouver, BC.

in question apply synchronically over several places. Of course it could be that places are scanned in series and their coordinates (or distance and direction of scanning) recorded for each of the places, and then the property is subsequently computed from the internalized synchronic representation. But evidence conerning the speed and accuracy with which we can detect such properties as collinearity over a series of points seems to argue against an alternative that involves prior scanning to locate the elements of the relation (and in any case this leads to the prediction that with more elements or elements further apart it would take longer to decide on whether a property is present — a prediction that is not generally upheld (at least in the cases which we have studied, for example involving discriminating collinearity). Moreover the scanning-and-encoding option simply internalizes the problem which we have good reason to assume, at least in many cases, is solved by actively consulting the display. 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, at least requires a mechanism that allows multiple access, or potential parallel access to several salient places.

We need to be clear what it means to have potential parallel access to several places. Having *access* to an object or place is quite different from carrying out an operation on it. The clearest example of an access mechanism is a pointer in a computer data structure. A pointer is a symbol which some process 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 token or object, so when the function is evaluated it is the object pointed to that serves as the operand. If the function has several such arguments then we need to have access to all the objects bound to these arguments, even if the arguments are subsequently treated in some sequence rather than all at once. This is why we refer to the pointers as providing "potential parallel access".

The operands which the visual system operates on may be features or objects located at certain places in the visual field. Having access to these objects means that the visual system has a way to interrogate or to carry out some process over some or all of a set of objects in the visual field that have been *indexed*, or as we say, FINSTed. Because these features or objects have pointers to them they can be visited or queried in any order determined by the nature of the process which uses them, without first scanning or searching for those places. To have potential parallel access does not require that all places are 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 places. What parallel access does entail is that the access mechanism itself does not constrain the processing to occur in any particular order given by the layout of the places — or indeed to be serially ordered at all.

Notice that if access were through a conventional unitary "spotlight" of attention, then the attentional beam would have to be driven from place to place. So we must still have a theory of the 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 can be instructed to, say, scan in a certain direction from its current locus. But if attention can be moved to disparate places in arbitrary order or simultaneously to a set of places in the field then we would need an account of the mechanism by which attention could be sent directly to place some x or to some set of places. The control process would have to specify the particular x where the attentional beam must move to and that entails a way to refer to or index x.

What this comes down to is that we need a mechanism for allowing places in a visual scene to appear in arguments to both visual predicates (for example predicates such as *Collinear*(x1,x2,x3,…,xn) and to commands to allocate processing or attention or to move the eyes to a certain place. Unless there is a mechanism for indexing places then we would have to rely on some sort of scanning process to locate the places and to some kind of location-encoding process to retain their locations. If we want to avoid positing scanning and coordinate-encoding processes then we will have to have a primitive way of sending attention or processing to certain places or of foveating certain places. And we shall argue that there is independent evidence for an indexing mechanism with certain interesting properties.

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 the early 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 features/objects[2] to be directly (and therefore rapidly) accessible to further processing by preselecting, and hence assigning FINSTs to them, in one of several ways — either exogenously by making those features 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 objects scattered throughout the visual field. Consequently we expect that items so selected will be accessed without 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 the display for them. 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 feature 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 features across the visual field, then one of the consequences should be that it is possible to prepare for simultaneous (or at least rapid and location-independent) examination of several filled-places in vision. 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 and/or zoomed to cover smaller or larger regions. Among the evidence we shall discuss are experiments showing that observers can simultaneously track some 3-5 identical target objects moving randomly and independently among an array of identical moving (non-target) objects. Some of this evidence has been published (Pylyshyn & Storm, 1988; Yantis, 1992) and others will be summarized here for the first time. Among the lines of evidence we shall cite are those that show: (a) Subjects can visually track several randomly-moving objects under conditions where their performance could not be explained by a serial scan-and-sample process which scans in an analogue manner at speeds that are anywhere near the range that is generally accepted (Pylyshyn & Storm, 1988); (b) tracked objects undergoing form transformations are detected more rapidly than non-tracked objects undergoing form transformations, regardless of the attentional saliency of the form change itself (Sears & Pylyshyn, 1993) (c) A “zoom-lens” view of visual attention cannot account for this detection advantage, because detection performance is no better for objects lying inside the convex-hull region defined by the tracked objects than it is for objects lying outside that region (Sears & Pylyshyn, 1993) — they are only faster *on* the disjoint tracked objects themselves; (d) when we cue several (3-5) locations with rapid onset cues this allows a subject to treat those locations as though they were the only ones examined in the field of view (Burkell & Pylyshyn,1993); (e) in such precuing experiments, increasing the distance among precued objects, and the number of intervening objects, does not lead to an increase in the time to access them (as would be the case if they had to be visited by scanning); (f) subitizing phenomena appear to be nicely accounted for in terms of a FINST allocation view if we assume that subitizing can be carried out by an internal enumeration of active indexes, without having to scan the display — and the insensitivity of subitizing to location cues supports this analysis.

In what follows we shall briefly describe a few of the relevant experiments and results.

Review of some experimental findings

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 members of 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 other non-target objects, which made the historical continuity of each target's motion the only clue to it's identity. After a randomly determined interval of tracking (7 – 15 s in one experiment, 4 s in another) a square was flashed on the screen (for about 83 ms in one study, 50 ms in another) and subjects indicated whether the flash had occurred on an target object, a non-target object, or somewhere else in the display. Accuracy was surprising high in this task, although it decreased with increases in the number of targets tracked. This drop in accuracy rates was paralleled by an increase in reaction times.

In this study, one particular condition, with 4 targets and 8 distractors, was analyzed closely, both empirically and theoretically. We wished to determine whether a serial process involving scanning a spotlight of attention might be able to reproduce the observed results. A round-robin scan-and-update strategy was therefore simulated to serve as a baseline in comparing observed performance. 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. Its actual coordinates were then stored as the updated coordinates for the next cycle. There was always the possibility that at this stage a nontarget (or another target) would get mistaken for the intended target and thus one of the tracked items would be lost. The simulation was carried out using the coordinates and timing of the actual dynamic displays that were used in the multiple-object tracking study. The simulation was repeated 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 figures are far below the 87% correct mean identification rate actually observed. It was concluded that even with the fastest reported scan rates it was not possible for the task to be carried out at the level of performance actually observed without some parallel tracking. Further, analysis of performance levels that might be expected if subjects were extrapolating trajectories or tracking some subset of n objects at a time and guessing at the remainder, 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 his experiments subjects tracked 3 or 4 target items and identified all of the tracked items at the end of every trial. A trial ended with all the items in the display falling into a regular pattern (a clock face, with the extra (non-target) items arranged in the four corners of the screen). Subjects indicated where each of the tracked items came to rest in this pattern. McKeever & Pylyshyn (1993) found that performance continued to be high under these conditions: when subjects tracked four targets they could correctly report the final positions of all four objects on 20% to 30% of the trials. This was greater than would be expected under a several guessing strategies.

But not all prima facie predictions of the FINST hypothesis are born out in tracking studies. For example it has repeatedly been found 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 of independence of the individual tracked objects and the assumption that non-indexed items are not processed beyond the initial activation stage where indexes are initially set. But if only tracked targets were processed then one would expect the number and nature of non-targets to be irrelevant. In fact both McKeever & Pylyshyn (1993) and Sears & Pylyshyn (1993) found that performance deteriorated when there were more nontargets. If indexes allows one to filter out unindexed objects then the filter appeared 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 perhaps even periodically refreshing the index to prevent inhibition or decay.[3] Suppose subjects are able to detect when a target has been lost — at least on some significant number of trials. Then they might attempt to recover the lost target by searching for the most likely object that might be the target in question. In that case various factors such as the density and the feature 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. 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 decision mechanisms which not only help in error recovery, but could also facilitate the tracking task by anticipation: 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.

Sears and Pylyshyn (1993) argued that the increase in reaction times to target form transformations with increases in the number of non-targets was a consequence of (a) the loss of indices pointing to target objects over time (maintaining indexes — as opposed to the indexing process itself — may well require periodic active attention and refreshing or reactivation); (b) the probability that an index will be lost increases with increases in the number of non-targets; (c) when an index is lost subjects cannot distinguish the previously indexed target from the non-targets and thus treats them identically and, (d) as a result they respond to form transformations on previously indexed (“lost”) targets no faster than to form transformations non-targets.“ If this were true, then the RT’s to trials where subjects had accurately tracked form-changing targets (and responded quickly to these form changes) would be averaged with the RTs to trials were they had lost form-changing targets (and thus responded more slowly to the form changes), which would produce an apparent increase in RT with increases in the number of non-targets.

To test this hypothesis, Sears and Pylyshyn (1993) asked subjects not only to respond when they had detected an object undergo a form change, but also to indicate whether the object that underwent a form change was a target or non-target. They argued that by using this dual response method they could discriminate the trials where a form-changing target was successfully tracked (and indexed) from those in which it was lost.

Sears and Pylyshyn (1993) found that the number of target form changes erroneously attributed to non-targets increased with increases in the number of non-targets in the display, which they interpreted as confirming that the probability of a target being lost increased with increases in display size. Moreover, they found that the detection of form changes on these lost “targets” (which were identified as non-targets) was no faster than that for the non-targets themselves and significantly slower than that for the successfully tracked targets. Since the probability of a target being lost increased with display size, this could account for the display size effect for target form changes. In fact, Sears and Pylyshyn (1993) did find that excluding from the RT analysis those trials were subjects had no longer indexed a form-changing target eliminated the display size effect for target form changes (although there was still an effect of display size for non-target form changes).

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 access can be to features distributed in space without any access to 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 unitary attention (or attention-beam) theorists, is the ubiquitous idea that attention remains unitary (and hence 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 punctate (local) 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?) at the actual objects being tracked, and not in the region bounded by the tracked target objects. 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 detection latencies for places between the two tracked objects was no faster than elsewhere.

This multiple-item tracking studies provide strong support for one of the more counterintuitive predictions of FINST theory — viz, that so long as indexes are successfully maintained, the indexed items can be treated by the visual system as though they were the only ones in the display. Moreover it appears that it is the indexed objects themselves and not some contiguous region which contains them, that is selected. The punctate nature of the indexing mechanism, as well as the FINST assumption that several places 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. We now turn to these studies.

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 search paradigm (Treisman & Galade, 1980) and had the following property. In all cases subjects were presented a set of items (totaling 12 or 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 about 9 msec/item when there was no target in the subset and 18 msec/item when the set included a target (the effect of subset size was nonsignificant). 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 1500 msecs after the onset of the non-subset position markers). 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, since in at least 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 as a whole 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 may be subject to the kind of “registration” process that feature sets are in search tasks in general (Treisman & Gelade, 1980). 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. The actual mechanism involved is currently the subject of theoretical investigation (Acton & Eagleson 1993).

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, although may be related to the diffusion of the effects of FINST operations and perhaps some ensuing interference among the closer indexes.

Subitizing studies

There are some clear difference between the process of enumerating small and large numbers of items. These differences are manifest not only in latency functions, but in error rates and confidence ratings (Taves, 1941; Kaufman, Lord, Reese, & Volkmann, 1949), and subjects’ reports of what they do when they enumerate (e.g., Hamilton, 1859; Jevons, 1871; Shrager, Klahr, & Chase, 1983; Van Oeffelen & Vos, 1982a; Warren, 1897) as well as the susceptibility of the two kinds of enumerations to different variables seem (e.g., Frick, 1987; Hunter & Sigler, 1940). These various findings have lead researchers to conclude that there are two different enumeration processes. One process is specialized for small numbers of items and is effortless, fast and perfectly accurate. This process is called subitizing (Kaufman, Lord, & Reese, 1949; amended by Jensen, Reese, & Reese, 1950 for reaction time). The second process can handle large numbers of items, but is slow, effortful and error-prone. This process is called counting. The “elbow” in the reaction time curve is taken to be the boundary between the subitizing and counting ranges. Estimates for the subitizing range vary between 1-3 and 1-7 depending on the paradigm and criterion used to calculate the subitizing range (c.f., Mandler & Shebo, 1982; Atkinson, Campbell, & Francis, 1976). (Note that since the present experiments hinge on determining whether or not subitizing occurs with various kinds of stimuli it raises methodological questions about how to determine when subitizing has occurred. The method used is described in greater detail in Trick & Pylyshyn, 1993c.)

Warren discussed the differences between “perceptual” and “progressive” enumeration in 1897. Yet no one has definitively explained why two enumeration processes are necessary. In Trick & Pylyshyn (in press) we suggest that the difference between subitizing and counting arises from the architecture of the visual system. In particular, we argue that subitizing exploits a limited capacity parallel mechanism for item individuation, namely the FINST mechanism which we have already discussed.

We carried out a series of studies (summarized in Trick & Pylyshyn, 1993a; 1993c) providing 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 while counting requires 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 cue validity, has a greater effect on counting than subitizing latencies (Trick & Pylyshyn, 1993b; 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 published studies that failed to produce strong evidence of subitizing had subjects enumerating concentric circles (Saltzman & Garner, 1948; c.f. Woodworth & Schlosberg, 1954; Allport; 1975). However such stimuli have some rather special characteristics: the nearest and most similar contours come from different items, they are necessarily of different sizes, and being concentric they share a common center. For the present purposes in was important to determine which factor caused concentric items to be so difficult to enumerate. There were three conditions in these experiments. In the Same Size condition subjects were required to enumerate uniformly sized rectangles that were defined by four edges in a bounding contour (in each display all the rectangles might be small, medium, or large, but they were all the same). In the

Different Size condition at least one of the rectangles was different in size from the others. Finally, in the Concentric condition subjects were required to enumerate concentric rectangles.

There was clear evidence of subitizing in both the Same and Different Size conditions; 12 of 12 subjects showed the appropriate deviations from linearity resulting from an increase in slope after 3 or 4. In contrast, in the Concentric conditions, only 2 of 12 subjects showed deviations from linearity at all — and in particular the deviations resulting from an increase in slope. Moreover, for the Concentric condition the slope in the 1-3 range was approximately the same as the slope in the 5-7 range. Both slopes were comparable to the slope in the 5-7 range in the other two conditions. Thus, in cases where spatially serial analyses are required to resolve items as wholes, in cases where parallel preattentive processes would be expected to group contours from different items together to form units, subitizing was not evident. Yet none of the subjects had difficulty subitizing multiple contour items of different sizes.

Of course there is the possibility that the result arose from the fact that contours were closer together in the Concentric than Same and Different Size conditions, so that lateral masking between contours from different items might explain the absence of subitizing in the Concentric condition. For this reason a second study was 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 lines 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. Therefore, concentric items are difficult to subitize because the items have a common focus and are one inside another, rather than because of the variety of item sizes or the proximity of contours from different items.

A third study investigated the ability to subitize when the enumeration task required computing a spatial relation. According to Ullman (1984) and (Jolicoeur, 1988; Jolicoeur, Ullman, & Mackay, 1986), serial attentive processes are required to compute the “connected-to” relation. In the present experiment, subjects were presented with a winding contour imposed over an array of parallel lines. The items to be enumerated were small blocks that could be green or purple. 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. The items to be enumerated were specified by one of two conditions.

In the Connected condition subjects were required to enumerate items on a particular contour. At the beginning of the trial subjects were provided a lateral fixation marker to tell them the starting point of the contour they were supposed to attend to. Subjects were required to visually trace the contour, enumerating blocks until they came to the end of the 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. If attention is required to compute the connected relation then subjects should not be able to subitize the subset of connected items.

In the Color condition subjects were shown the same displays, and given the same lateral fixation point, 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 feature (e.g., Treisman & Gelade, 1980). Because preattentive information distinguished target items from distractors subitizing was predicted in the Color condition. (Once again the reader is referred to Trick & Pylyshyn (1993a) or the summary in Trick & Pylyshyn (1993c).)

There was clear evidence of subitizing in the Color condition. All subjects had the appropriate deviations from linearity. In contrast, there was little evidence of subitizing in the Connected condition; there was only one deviation from linearity, for one subject at the intermediate contour length only. Moreover, in the Connected condition, latencies were affected by the length of the contour. In the Color condition the contour length (though it did not define the set to be enumerated) did not affect enumeration latencies.

Thus, subitizing was not evident when a spatial relation was superimposed on the enumeration task, and that spatial relation forced spatially serial processing. Nonetheless, subjects could subitize the same displays if the task were to enumerate items of a particular color, a task that did not require spatially serial processing.

The final experiment in this series 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 by the individuation task. In Trick and Pylyshyn (1993a) a search task was superimposed on an enumeration task. Subjects were required to enumerate items in a field of distractors. There were two conditions. In the Disjunction condition subjects were required to enumerate white OR vertical lines in green horizontals. In the Conjunction condition subjects were required to enumerate white vertical lines in green vertical and white horizontals. There were 0-8 target items, and 0-20 distractor items.

Subitizing was always evident when there were no distractors in the display. However, 9 of the 10 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; 0/10 and 1/10 showed the telltale slope deviations in the 12 and 20 distractor conditions respectively. 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 1 item in the Conjunction condition whereas each distractor only added 6.2 ms in the Disjunction condition.

As we have already remarked, search experiments slopes of less than 10 ms/distractor are often taken as indications of spatially parallel (preattentive) processing, and slopes in excess of 10 ms/distractor as indications of spatially serial (attentive) processing. From this criterion, processing of distractors seems to be spatially serial in the Conjunction condition and spatially parallel in the Disjunction condition. It appears then 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 condition, but are incapable of subitizing targets among distractors when spatially serial analysis is required to distinguish targets from distractors, as in the Conjunction condition.[4]

These studies argued 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 preparation for 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 the cue validity study subjects are required to make a perceptual decision, e.g. press one key if there is a “B” in the display and another if there is a “D” under conditions in which they know beforehand both when and where a stimulus will appear (Valid cuing) or in which they only know when the stimulus will appear (Neutral cuing).

in addition there is also an invalid cuing condition in which subjects are given incorrect information about where it will appear. Typically, subjects are faster and more accurate at making perceptual decisions if they are given correct information about where the stimulus will fall. Performance is thus best in the Valid condition, followed by the Neutral and Invalid 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.

In Trick and Pylyshyn (1993b) a cue validity paradigm was combined 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 condition, or distributed throughout the display, as in the Neutral condition (Gawryszewski, Riggio, Rizzolatti, & Umilta (1987) have argued that processing resources are dispersed evenly throughout the display 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 subitizing latencies. Specifically, the difference between Valid and Invalid 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.

Five cue validity experiments were performed. In all of them subjects were required to count 1-8 items. The first study used colored rectangles to cue the area in which dots were to appear. There was one rectangle in each corner of the display: Cuing rectangles preceded the dots by 128 msec. In the Neutral condition all the rectangles were the same color. In the Valid and Invalid conditions, one rectangle was a different color from the others. The position of this rectangle predicted with 80% accuracy the position of the dots. The remaining four studies involved small alternations of this basic design, allowing the effects of other variables, such as central (exogenous) versus peripheral (endogenous) cues, and the presence or absence of distractors to be investigated. These effects are described in Trick & Pylyshyn (1993b) and will not be discussed here.

The results were quite clear. In all experiments, and for all subjects, subitizing was evident in Valid, Invalid and Neutral conditions. The necessity of contracting the attentional focus in the Valid and Invalid conditions did not prevent subitizing, or even restrict the subitizing range.

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. Similarly, in the other studies, when cues had any effect at all, the effect was stronger in the counting range than the subitizing range.

Indexing and the line-motion illusion

Finally, we have just begun a series of experiments which use a line-motion illusion reported by Hikosaka, Miyauchi and Shimojo (1991), which is believed to be attention-sensitive. This technique allows us to investigate questions concerning resource limitations on simultaneous processing of multiple loci in a visual field, the automatic assignment of indexes to features, and to test other assumptions of the FINST hypothesis 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 presented with that target as an endpoint. In its simplest form subjects are asked to fixate a marker. A trial begins as a cue onset exogenously draws attention. After a brief ISI, a line is instantaneously drawn with the onset cue as one if its endpoints. Subjects consistently report that the line was "drawn" away from the cue. This phenomenon has been used to explore the temporal characteristics and spatial extent of endogenous and exogenous attention (Miyauchi, Hikosaka, and Shimojo, 1992). In the latter study, Miyauchi and colleagues used the illusion to cancel a real motion of the line in the opposite direction of the effect. The level of real motion that is cancelled by the illusion gave them a measure of the strength of the effect in a variety of situations. They used this technique to explore a number of attentional phenomena, but did not test the effect of multiple attentional cues.

In a series of studies carried out in our laboratory (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 then cued with onset cues and then randomly probed for the occurrence of the illusion. Using this method we have been able to assess whether or not there are limitations on the number of loci of visual information processing, as well as the nature of those limitations. In the first experiment of this nature, displays consisting of 1 to 8 cues evenly spaced around an imaginary circle were used. Subjects maintained gaze on a fixation point in the centre of the display and all of the cues onset simultaneously for 250 msec. At this point, the cues were extinguished and a line was drawn from the fixation point out to the circle's circumference at a variable speed. 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 between where two randomly chosen cues had previously been. The line illusion was observed significantly more often in the former case than the latter, demonstrating that without the presence of a visual object, processing resources were not allocated.

The frequency of occurrence of the illusion was found to decrease as the number of cued locations in the display increased if the line was drawn to a cued location, whereas the opposite held true if the line was drawn between two cues. Looking at individual data (rather than averaging all subjects), there was a strong decline in the frequency of the illusion once a certain minimum number of cued locations were reached, providing support for the notion that there were a limited number of loci which were indexed and thus could be accessed for further processing — including the allocation of focused attention. The fact that processing could occur at any of a number of cued loci, coupled with the observation that it did not occur between cued loci presents problems for unitary accounts of attention, since they have no way to account for the allocation of the attentional resources within the display. A zoom lens explanation of these effects using a single large beam of attention is not plausible (Eriksen & St. James, 1986), nor are the suggestions of an annulus of attention (Egly & Homa, 1984), or of a moving attentional beam.

A similar study carried out by Schmidt & Pylyshyn (1993) used the line motion illusion to assess the limits of processing at multiple loci. In this experiment, a set of 12 cues appeared evenly around the circumference of a circle and then some randomly selected subset of these cues, ranging in number from 0 to 12 underwent a luminance increment. The cues were extinguished 250 msec later, and a line was drawn equally often from the centre of the display to the locations of a cue that had or had not undergone a change.

As we saw with the Burkell & Pylyshyn (1993) results, FINST theory predicts that as long as the number of changes do not outnumber the availability of indices, potential parallel access should be guaranteed for those items undergoing the change, but not for those items that are not as salient. Consistent with these predictions, the results demonstrated that if the line was drawn to a cue that had not undergone a change, there was no difference in the proportion of trials for which the illusion was reported regardless of the number of changes in the display. However, for items that underwent a change the illusion was reported more often in displays that had fewer changes than in displays that had a larger number of changes. If the number of changes was less than approximately seven, then report of the illusion occurred around 95% of the time. The proportion of illusions reported for displays with a greater number of changes and with the line drawn to a cue that had undergone a change was as poor as if the line were drawn to a cue that had not undergone a change.

The results from this second study also demonstrated a breakdown in the number of loci that can afford potential parallel access to higher order processes. They also demonstrated that not only a sudden onset but also a brightening, which we independently believe attracts indexes (Burkell, 1992) was responsible for the difference in processing observed. Again, a theory of attention restricting itself to a single focus requires a method to access information at disparate and unpredictable locations in the visual field if it is going to account for the current data. FINST indexes could provide a method by which attention can access multiple loci to account for the current results.

A number of additional studies were also carried out using the line-motion illusion. Although the results are relevant to such questions as when an index is assigned to an item, they are not as germane to the main assumptions of FINST theory. They did, however show that indexes may be assigned to items that become more distinct in relation to their environment even though these items themselves do not change. Thus we do get the illusion most strongly to items that remain bright when the other items briefly become dimmer. We also get the illusion most strongly to the feature that is numerically in a minority in the display consisting of two types of features (horizontal and vertical bars) and this effect reverses when the same feature becomes the most common one in the display. These and other studies currently under way will help pin down some of the underdetermined aspects of the theory.

Conclusions

The notion that has motivated this research is the following. Prior to the allocation of limited attentional resources by the visual system 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 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 this 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 the cued items 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) and 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 allow 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 are “popout” 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 line motion illusion of Hikosaka, Miyauchi & Shimojo, 1991), 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 Ophthalmology 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.

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

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.

Gawryszewski, L., Riggio, L., Rizzolatti, G., & Umilta, C. (1987). Movement of attention in the three spatial dimensions and the meaning of "neutral" cues. *Neuropsychologica, 25*(1a), 19-21.

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 Ophthalmology 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.

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

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. 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). Cuing and counting: Sometimes the spotlight isn't much help. Submitted.

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

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.

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.

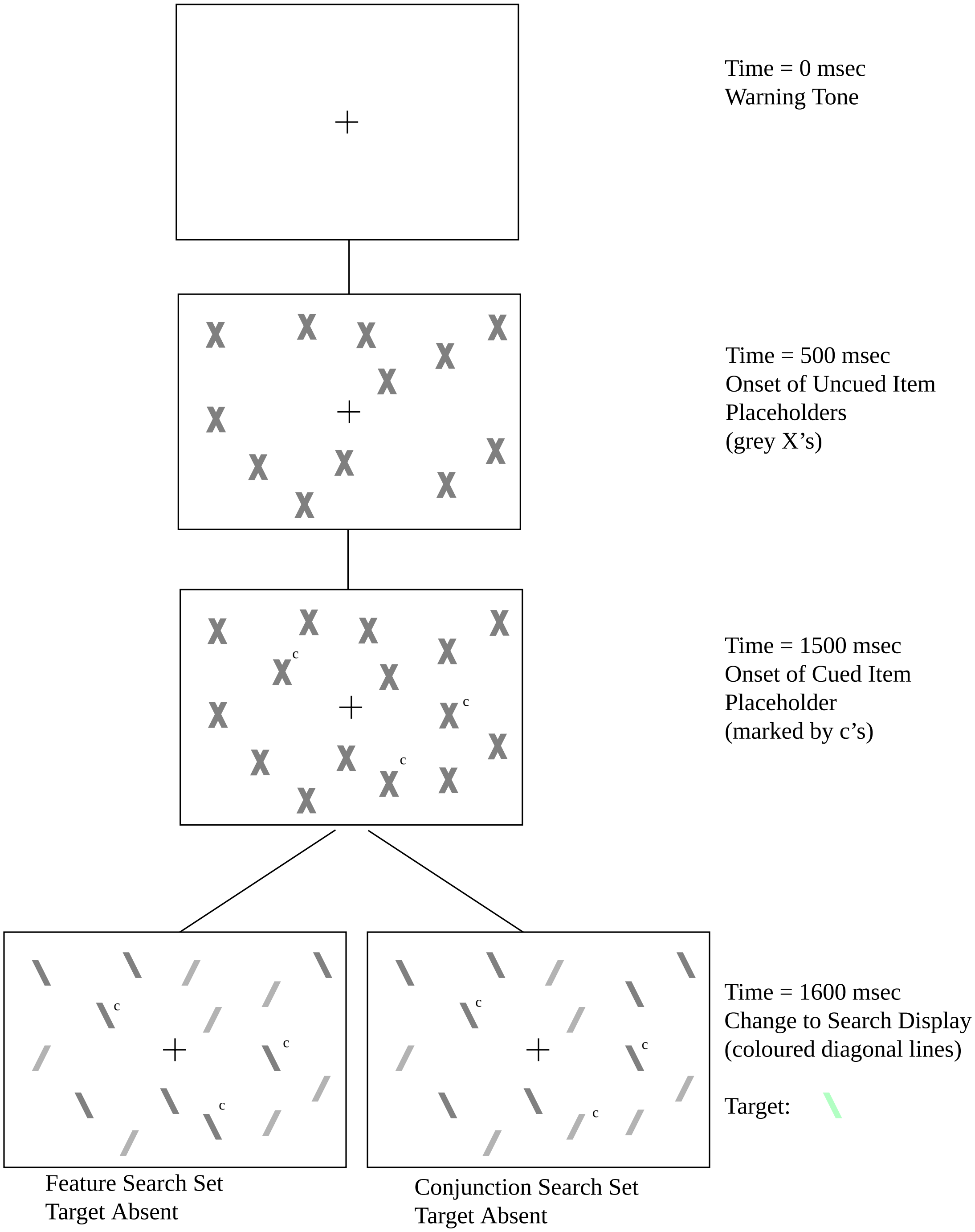
Yantis, S. & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception & Performance, 16*(1), 121-134.

— Notes —

2. To save having to say “features, feature-clusters or objects” we shall simply use the term “feature” as a stand-in for all these alternatives, with the understanding that it is a substantive empirical question whether FINSTs are assigned and remain with integral objects or with highly-local optical features. Current provisional evidence appears to favor an object-based view — see Kahneman, Treisman, Gibbs (1992).

3. We do not have a process model of the tracking task, but it is clear from such data as the dependence of performance to number of targets and distractors, to restrictions in target trajectories, to ancillary distracting visual tasks (Treisman, personal communication), that more is going on in tracking *tasks* than the mere invocation of an automatic tracking *mechanism*. Index maintenance is one such unspecified relevant process, and post-tracking processes such as *error recovery* and *response selection* constitute others.

4. It might be argued that subjects could not subitize conjunctions because there were two types of distractor in the conjunction condition and only one in the disjunction condition. For this reason an experiment was performed in which subjects enumerated letters, and there was always only one type of distractor. Subjects were given the task of enumerating O’s in either a background of X’s or a background of Q’s — a pair of tasks which Treisman (1985) showed corresponded to preattentive feature search and attentive conjunction search, respectively. Subitizing was evident for O’s in X’s, but not O’s in Q’s. Therefore, it seems unlikely that subjects could not subitize conjunctions simply because of the number of different types of distractors.


Time = 0 msec
Warning Tone
Time = 500 msec
Onset of Uncued Item
Placeholders
(grey X’s)
Time = 1500 msec
Onset of Cued Item
Placeholder
(marked by c’s)
Time = 1600 msec
Change to Search Display
(coloured diagonal lines)
Target:
Feature Search Set
Target Absent
Conjunction Search Set
Target Absent