The Role of Visual Indexes in Spatial Vision and Imagery*

Zenon Pylyshyn

Rutgers Center for Cognitive Science
Rutgers University
New Brunswick, NJ 08903
zenon@ruccs.rutgers.edu

Abstract

This paper describes a programmatic theory of a process in early vision called *indexing*. The theory hypothesizes that a small number of primitive indexes are available for individuating, tracking and providing direct access to salient visual objects. We discuss some empirical and theoretical arguments in favor of the proposed index as a resource-limited link between an internal visual representation and objects in the visual world. We argue that this link is needed to explain a large range of properties of vision, including the ability to simultaneously track multiple moving objects, to select a subset of visual items to process, as well as such larger issues as how mental images get their apparent metrical properties and why the world appears stable despite constant retinal motion.

Backgound

It is a truism that the visual system does some things in parallel and some things serially. The eye moves in rapid saccades several times each second. The center of the fovea is the center of our gaze and provides a focus of attention. In addition, attention can also move independent of eye movements. Many people (e.g., Posner 1980) have show that a region of most efficient processing, can be scanned across the visual field at about 100 degrees/second when the eye is fixated (published estimates run from 30 to 250 deg/sec). A large amount of research has been done on this so-called *covert* attention movement. The assumption has generally been made that such *focal attention* is confined to a contiguous region, possibly expandable in size, that must be scanned over the display from place to place.

There are two good reasons why the visual system should have such a locus of maximum resource allocation.

- One is that the visual system has a limited information capacity at some level so some selection must occur. Various people (e.g., Tsotsos, et.al. 1995) have argued for the importance of a serial component in visual processing simply from computational complexity considerations.

- The other reason is that there may be patterns which in principle require serial scanning for their evaluation— as suggested by the Perceptron Theorems of (Minsky & Papert 1969). Such a process of serial evaluation of patterns is referred to by Ullman as a "visual routine" (Ullman 1984).

Notice that a serial process that scans the display cannot by itself execute a visual routine, for example it is not possible to determine by scanning alone whether a set of points is collinear or whether a given point is inside a closed curve. That's because the process has to know which objects are being referred to. It also has to have a way to determine which objects it has already visited and which objects to visit next.. Like the Sesame Street lessons in counting, a critical part of the process is remembering which items have already been counted!

A number of people, including (Ullman 1984) and (Yantis & Jonides 1984), have referred to this process of keeping track of objects as "marking." While the metaphor of marking objects is tempting, it is also misleading since we can't literally mark the display. What this terminology suggests, instead, is that we have

* This paper was presented at the AAAI 1996 Spring Symposium Series on *Cognitive and Computational Models of Spatial Representation*, Stanford University, March 25-27, 1996. It also appears in: R Wright (Ed), *Visual Attention*, Oxford Univ Press, 1998, 215-231.

a geostable (or allostable) icon somewhere in our heads where we can place a marker. As I have argued elsewhere (Pylyshyn 1981), there are many reasons to refrain from hypothesizing such a metrical iconic display in the head. Consequently, over the past decade we have developed an alternative view of how places in the visual field are "marked" or, to use our terminology, "indexed." In the remainder of this paper I will briefly summarize the theory of *indexes* (sometimes referred to, for purely historical reasons, as FINST indexes) and will sketch a variety of empirical evidence to support the general idea (for a more detailed exposition, see Pylyshyn 1989, Pylyshyn 1994a, Pylyshyn, et. al. 1994). This theory is programmatic, however, and incomplete in many important respects — which is why we are continuing our research on it. On the other hand, it's hard to see how this growing body of evidence can be dealt with without something like indexes.

A Theory of Visual Indexes

According to the theory, an early preattentive stage in visual perception involves a resource-limited mechanism for *individuating* a small number (4-6) of visual tokens. We follow the precedent set by (Treisman 1988) and others and refer to these visual tokens as "objects", in order to emphasize that what are indexed are temporally enduring entities, identified in terms of their historical continuity.

Individuating is more primitive than encoding either the properties of the objects or their location in the visual field. Individuating implies that the objects are selected or distinguished from one another as discrete entities with a historical continuity. Once individuated, each object maintains its identity over time and continues to be identified as the "same" object despite changes in its location and possibly other properties as well. Objects are individuated by being indexed in the same sense that a book in a library or a datastructure in a computer might be indexed: the index serves as a mechanism by which subsequent operations can access the object. A small number of objects are selected (presumably by virtue of possessing certain salient properties) and indexed in order that the following further functions are possible.

1. Subsequent stages of the visual system are able to reference the indexed objects — say for purposes of determining their individual and relational properties. By hypothesis, only indexed objects can be bound to arguments of visual routines and consequently evaluated or otherwise operated on.

2. An index remains attached to its object as the object changes its retinal location or other properties, allowing it to be tracked, *qua* individual object.

3. Indexed objects can be interrogated (or "strobed") without the necessity of first locating them through some form of search. Consequently, once they have been indexed, any set of objects can be separated functionally from the rest of the display without reference to their properties or locations.

4. Motor programs, like visual routines, also require that their arguments be bound. Consequently, only indexed objects can be the targets of such visually-controlled motor actions as eye-movements or ballistic reaching.

As a result of having assigned indexes to salient visual objects, the cognitive system can *in effect* tag objects in the scene, and attach stored information to these objects. By linking stored information with indexes that point to the objects in the scene to which this information refers, the visual system can use access memory and visually present information with equal ease — and can use one to locate the other. In this way the scene can be thought of as an extension of memory. It can be interrogated for such things as the relative location of remembered objects and for other properties of the scene that have not yet been "noticed" or encoded. As others have noted, this provides a way to reduce the load on internal visual storage (O'Regan 1992, Ballard, Hayhoe & Pook 1995). Even more important, it may no longer be necessary for the perceptual representation itself to have metrical or pictorial properties (as assumed by Kosslyn 1995) since these can be extracted from the scene as needed. Rather the percept can be an incrementally evolving schema, with the following additional feature: by maintaining links to actual objects in the scene, indexes help to "situate" or "embody" vision. This allows the system to carry out a variety of operations on the scene, including scanning it and operating on it using what (Ballard et. al. 1995) call a "deictic strategy." This strategy minimizes memory load when operating on a display. It does so by relying on indexes to get back to where the information is located at the time it is actually needed for the task (in the Ballard et. al. case this is the task of copying a layout of colored boxes).

How Does Indexing Differ From Focal Attention?

Since *indexes* and *focal attention* both provide a way to allocate processing resources to an object in the visual

field, the question arises how they differ. These are the main points of difference, according to the theory.

- Several indexes can be assigned and are available in parallel. This does not mean that they will necessarily be accessed in parallel — only that they are simultaneously available for further processing. Moreover, whether they remain assigned indefinitely without being reactivated by being visited by focal attention is not known, though there is reason to believe that maintenance of indexes is not automatic and preattentive.

- According to the current assumptions of the theory of visual indexing, indexes can be assigned in one of two ways. One is autonomously, by the occurrence of a visual event such as the onset of a new object in the visual field (but perhaps also by other transients such as luminance changes). The other is by a deliberate cognitive act of assignment of an index to an object currently under scrutiny by focal attention. Unitary focal attention, by contrast, can be assigned either by scanning the display along a continuous path or by skipping to an object that has already been indexed.

- Indexes are *object-based*, therefore they stick with the object to which they are assigned as the object moves. This is what is meant by *individuating* an object: the continuity of the object over time is automatically maintained by the index, as well as it's distinctiveness from other objects.

- Indexes provide for a *direct* access to the indexed object, so these objects don't have to be *searched* for and located first. Consequently the relative distance between objects doesn't matter: a nearby object can be accessed just as quickly as one that is remove from the one currently being processed.

Summary: Why Do We Need indexes?

We have already suggested some general reasons why we have hypothesized a mechanism for assigning and maintaining visual indexes. Here we summarize the arguments presented in more detail elsewhere.

- We need indexes in order control where to move focal attention. Unitary focal attention need not be scanned around at random. It is usually directed to loci of interest or relevance to the task at hand. Consequently there must be some way to specify where such focal attention should move.

- We need indexes to execute *visual routines*, such as those which compute whether an item is inside a closed curve or whether *n* objects are collinear. By hypothesis, whenever we evaluate an *n*-place visual predicate we must first bind all *n* of its arguments to appropriate visual objects. Indexes provide just such a variable-binding mechanism.

- We need to situate vision in the world so we can act on the basis of what we see. The problem here is apparent when we note that one can point to or reach for objects without feedback as to where our hand is in relation to the object (Merton 1961). This means that there must be a cross-modality binding of places: The visual system must have a way to inform the motor system — which necessarily operates in a different coordinate system — where things are. Although this problem is far from solved, the role of indexes seems essential unless one is prepared to hypothesize a common global 3D coordinate frame to represent locations — which is far too strong a requirement.

- A great deal of evidence has recently been uncovered showing that our visual memory is much more limited that our phenomenology suggests. We retain very little detailed information from one fixation to another unless we have reason to notice particular features because they are relevant to our task. It appears, rather, that we typically use the world itself as a continuing source of information about the visual scene. This, however, requires a way to merge information in memory with information in the scene. Indexes are such a mechanism, since they provide a way to keep track of preattentively salient places so that memory information can be bound to them..

- Finally, we need indexes in order to avoid assuming a 2D metrical display in the head. Mental displays have often been hypothesized precisely in order to account for such things as visual stability and the ability to superimpose visual information in memory on visual information in the scene (see Kosslyn 1994; but see the review in Pylyshyn 1994b). We shall argue that these abilities can be accounted for to a large extend without the highly undesirable assumption that there is a 2D display in the brain.

Empirical Support for Indexes

The idea of a simple indexing mechanism has turned out to provide a rich source of empirical predictions. It also has far-reaching implications for explaining a wide range of phenomena. The following is a sample some of

our recent findings, summarized here to illustrate the range of empirical phenomena that can be handled by this simple theoretical apparatus. A large number of additional lines of investigation have also been conducted in our laboratory and elsewhere, but are not reported here. These include parametric investigations of multiple object tracking, studies relating indexing to multi-locus inhibition-of-return as well as to attentional enhancement (Sears 1995; Schmidt 1995; Wright 1994), studies relating indexing to such visual routines as those for detecting the inside-outside relation and collinearity, applications to the explanation of apparent motion (Dawson & Pylyshyn 1989), as well as preliminary studies of spatial indexing in the auditory modality. The question of how the indexing mechanism might be implemented—both computationally and neurophysiologically—also continues to be a major research pursuit (Acton 1993, Acton & Eagleson 1993, Eagleson & Pylyshyn 1991).

Multiple-Precuing of Locations in Search Tasks

Search tasks provide a nice demonstration of the use of indexes to control which items are examined or queried in visual search tasks. In a series of studies, Jacquie Burkell and I (Burkell & Pylyshyn 1996), showed that sudden-onset location cues could be used to control search so that only the precued locations are visited in the course of the search. This is what we would expect if onset cues draw indexes and indexes can be used to determine where to carry out processing. In these studies, a number of placeholders (12-24), consisting of black X's, appeared on the screen for some period of time (at least one second). Then an additional 3-5 placeholders (the late-onset items) were displayed. After 100 ms one of the segments of each X disappeared and the remaining segment changed color, producing a display of right-oblique and left-oblique lines in either green or red. The entire display had exemplars of all four combinations of color and orientation. The subject's task was to say whether the display contained a pre-specified item type (say a right oblique green line). In most studies there was only one of these "targets" in the display. As expected, the target was detected more rapidly when it was at a location precued by a late-onset cue. There were, however, two additional findings that are even more relevant to the indexing theory. These depend on the fact that we manipulated the nature of the precued subset in certain ways, to be explained below.

It is well known that when subjects try to find a target that differs from all non-targets by a single feature (e.g. it is the only red line or the only right oblique line in the display) then they are not only faster overall at locating the target, but the search rate is also very fast (i.e., response time increases very little as the number of non-targets is increased — about 10-15 ms per item). This is called a "simple feature" search. By contrast, when the target differs from some nontargets by one feature and from other nontargets by another feature — so that what makes the target distinctive is the combination of two or more features — then it takes longer to locate the target and the search rate is also slower (it takes more additional time for each added nontarget in the display — about 40-60 ms in our case). This is called a "conjunction feature" search. As mentioned above, the displays in the present experiments typically contained all four types of items, so the displays were always of the "conjunction feature" type. However, the subset that was precued by late onset placeholders could be either a simple or a conjunction feature search set. So the critical question here is whether the feature type of the subset is the determining factor in the search. We found clear evidence that it is. Here are the two most relevant findings.

1. When the precued subset consisted of elements that differed from the target by a single feature, search had the same pattern as with simple feature searches — i.e., we observed shorter response time and faster search rate. However, when the precued subset consisted of some elements that differed from the target in one of the features and some nontargets that different in the other feature, then search rate was much slower. This suggests that the precued subset was being selected and separated from the rest of the display. It provides evidence that indexes, assigned to sudden-onset placeholders, can be used to control which items are visited in a visual search task.

2. Even more relevant to the indexing thesis was the additional finding that when we systematically increased the distance between precued items (or their dispersion) there was *no* decrease in search rate. It seems that the spatial dispersion of the items does not affect the time it takes to examine them, even when the examination appears to be serial (e.g., the time increases linearly as the number of nontargets increases). This is precisely what one would expect if, as we predict, the cued items are indexed and indexes can be used to access the items without spatial scanning.

Parallel Tracking of Multiple targets

One basic assumption of the theory is that a small number of "sticky" index pointers can be assigned pre-attentively to primitive visual objects and will continue to be assigned to the same object as the object moves on the retina. This hypothesis has been tested directly using a multiple-target tracking paradigm.

In these studies a subjects were shown a screen containing 12-24 simple identical objects (plus signs, figure eights) which moved in unpredictable ways without colliding (because of a simulated barrier or a linear or distance-squared "force field" between them, depending on the study). A subset of these objects was briefly rendered distinct by flashing them on and off a few times. The subjects' task was to keep track of this subset of points. At some later time in the experiment an object was flashed on the screen. The subjects' task was to indicate whether the flash occurred on one of the tracked objects or one of the others (or in some cases, on neither).

The initial (Pylyshyn & Storm 1988) studies showed clearly that subjects can indeed track up to 5 independently moving identical objects. The parameters of these experiments were such that tracking could not have been accomplished using a serial strategy in which attention is scanned to each point in turn, updating its stored location each time it is visited, until the probe event occurs. Recently a large number of additional studies in our laboratory (Sears 1991, McKeever 1991) and elsewhere (Intriligator & Cavanagh 1992, Yantis 1992, as well as personal communications by Treisman, Julesz) have replicated these results, confirming that subjects can successfully track independently moving objects in parallel. Some of these studies carefully controlled for guessing strategies and also demonstrated qualitatively different patterns of performance than would be predicted by any reasonable serial-scanning algorithms we have considered. The results also showed that a zoom-lens model of attention (Erikson & St. James 1986) would not account for the data. Performance in detecting changes to elements located inside the convex hull outline of the set of targets was no better than performance on elements outside this region, as would be expected if the area of attention were simply widened or shaped to conform to an appropriate outline (Sears & Pylyshyn 1995). No spread of attention was also reported by (Intriligator & Cavanagh 1992) in their tracking study.

Subitizing

Other studies have shown the power of this framework to account for a large class of empirical phenomena in which simple visual objects are rapidly and pre-attentively individuated. One of these is subitizing, a phenomenon whereby the cardinality of sets of less than about 4 visual features can be ascertained very rapidly (about 60 ms per feature). We have shown (Trick & Pylyshyn 1993, 1994a, 1994b; Trick 1990; 1991) that subitizing does not occur when pre-attentive individuation is prevented (e.g., targets defined by conjunctions of features, or other properties that require focal attention, cannot be subitized). We have also shown that in determining the number of objects in a display, the spatial distribution of the objects is unimportant in the subitizing range but critical in the counting (N>4) range, and that precuing the locations of objects (with either valid or invalid location cues) makes little difference to subitizing performance (Trick & Pylyshyn 1993, 1994b). According to the indexing hypothesis, small numbers of salient locally-distinct points are indexed preattentively and in parallel. Hence, their cardinality can be determined by counting the number of active indexes without having to spatially scan focal attention over the display. In any case these studies show that a preattentive stage of item individuation is critical for subitizing. Such a stage is postulated for entirely independent reasons by the theory. Moreover, (Simon & Vaishnavi 1995) used an afterimage display to show that even when indefinite time is available for counting, the subitizing limit remains. They take this as supporting the contention that it is not the time available for counting that is the source of limitation, but the availability of resources for individuating items.

Attention-Dependent Line-Motion Illusion

Another interesting phenomenon, apparently dependent on focal attention, was demonstrated by (Hikosaka, Miyauchi and Shimojo 1993). They showed that when attention is focused on a particular point in the visual field, a line displayed between that point and another (unattended) point appears to "grow" away from the point of focal attention. This phenomenon provides another way to test the hypothesis of a pre-attentive multiple-locus indexing mechanism.. In a series of studies, (Schmidt, Fisher & Pylyshyn 1995) showed that the illusion could be induced by the late onset of up to 6-7 noncontiguous items among a set of 12. A line displayed between any of the other pairs of points, or between a point and an unfilled location, does not show

the illusion. So once again we have an attention-like phenomenon occurring simultaneously at several places in a visual scene. Whether this means that the line-motion illusion requires only a preattentive indexing or whether it means that attention can be rapidly shifted to each of the points in turn, is unknown. However, the result is consistent with our earlier reported subset-selection phenomenon. In the present case the onset objects are preselected and this in itself appears to be enough to cause the line-motion illusion to occur at any of the preselected objects.

Mental Images, Indexes and Visual Stability

The phenomenology of both mental imagery and of vision is extremely misleading from the perspective of constructing an information-processing theory of the underlying representation and process. I have already commented extensively on the problems raised by the intuitive assumption that mental imagery involves the examination of a two-dimensional display projected on some rigid internal surface (or at least on a surface that ensures that the display obeys local Euclidean properties). I will not rehearse this argument here. But I do want to make two comments. One is that studies in which mental images are "projected" onto visual perception involve a special case of the use of imagery, inasmuch as the physical stimulus can in this case provide some of the properties attributed to the image medium itself. So long as we have a way to link objects in our imaginal representation to objects in the visual field, the mental representations can inherit some of the properties of real rigid surfaces. The other comment is that, thought it may not always be recognized, some of the very same issues that arise in understanding the nature of mental images actually arise in vision itself. This is particularly true when vision is thought to involve the contruction of an iconic representation who extent goes beyond the fovea and which is constructed by superimposing information from individual fixations as the eye moves about.

Superposition of Mental Images and Perception

Many studies have purported to show that mental images involve information that is displayed in a two-dimensional format (I do not pretend that this is even close to being a well-defined notion — indeed this is part of the problem with the so-called "debate" — see Pylyshyn, 1994b — I simply use the term roughly the way it is used by its proponents). Some of the more robust findings come from experiments involving superimposing images onto visual perception. For example, (Hayes 1973) has shown that anticipating a figure by projecting an image of the correct one enhances its detection; (Shepard and Podgorny 1978) have shown that if a spot is displayed on a grid on which a subject imagines a figure, the response times for detecting whether the spot is on or off the figure shows exactly the same pattern as it does when the figure is actually displayed (e.g. faster times for *on-figure* versus *off-figure*, faster times when the spot is at a corner or vertex of the figure, etc); and (Farah 1989) has shown that detection sensitivity for light flashes is greater for locations on an imagined figure than off the figure. I have argued that all these results can be explained more simply by assuming that indexes can be assigned to relevant objects in the display, including regions such as columns and rows of the grid. This assignment would serve to indicate where various "imagined" objects would fall on the scene, and hence where to direct focal attention. Thus if we can think that *this* column (where the locative "this" refers to a particular indexed column) is where the vertical stroke of the imagined letter will be placed, then we can assign focal attention to those parts of the pattern. Indeed, in one condition of her experiment (Farah 1989) simply asked subjects to *focus their attention* on the appropriate parts of the grid rather than imagine a figure projected on them and obtained exactly the same results.

Another well-known result is also easily understood in terms of indexes assigned to objects in an actual display. This is the "mental scanning" result of (Kosslyn 1978), in which it was found that the time it takes to switch attention from one point in an image to another is a linear function of the physical distance in the imagined situation (typically a memorized map). We have shown (Pylyshyn 1981) that this phenomenon disappears when the experimental demands are changed (e.g., when subjects are not asked to imagine that they are in a real situation of looking at a display in which they have to scan from place to place). But the scanning result appears to be robust when done in image-projection mode (i.e., when the experiment is done in the light and subjects have to imagine the map while looking at some visual display). But in this case, I have argued, if subjects have the capability to index a number of objects (or even bits of texton features) in the real scene and to link their representations of the recalled objects to those locations, then they can of course scan their attention, and even their eyes, from place to place *on the display*. Having a real display with objects to which mental representations can be bound ensures that all relative distance and pattern properties entailed by the recalled information hold — including some inferred or previously unnoticed

properties — because of the physical and geometric properties of the display. For example, suppose we imagine that three places lie on a line (call them *A, B* and *C*). Suppose, further, that we do this while looking at some scene and that we choose three collinear objects in the scene, and associate or bind the imagined places to the scene objects. In that case when we scan from imagined object *A* to imagined object *C* we are bound to pass over the real location of imagined object *B — because of geometrical properties of the physical scene to which A, B, and C are bound.* In the real scene B *does* lie on the path from A to C. In a purely mental representation we would have to make use of knowledge that might be stated as the constraint, "If three points *A, B,* and *C* are collinear, in that order, then in travelling from *A* to *C* we must pass by *B*". The only alternative is the assumption that there is a brain property that realizes this constraint and that it is used when we imagine collinear points. Although this is a logical possibility, the facts of cognitive penetrability of geometrical reasoning argue against it in general (but for more on this argument see Pylyshyn 1981).

Finally there are a number of other results that are equally amenable to being explained in terms of the indexing mechanisms, including the adaptation of perceptual-motor coordination to imagined locations (Finke 1979). So long as we can think of an object being located at an indexed point, we can act towards it in ways that may resemble our actions towards real objects located at those places.

Saccadic Integration and Visual Stability

The phenomenology of visual perception suggests that we have access to a large vista of a spatially stable scene, extending far beyond the foveal region. It seems as though the percept fills in blind spots and other scotomas, corrects for the rapid drop in resolution and color sensitivity with retinal eccentricity, and combines pictorial information from successive glances into a single extended internal image. Yet none of this is objectively true. It can easily been shown that the visual system does not have access to nearly the kind of information we feel is there. Consider just two brief relevant examples to illustrate this point.

- There is good evidence that far less information is extracted from individual glances than we feel is the case, based on our phenomenal experience — unless the information is relevant to some task at hand. (Irwin 1993, McKonkie & Currie 1995, and Irwin, McConkie, Carlson-Radvansky, & Currie 1994) and others have shown that surprisingly little qualitative information is retained between saccades. Moreover, the idea that pictorial information from successive glances is superimposed onto a master-image has been pretty generally discredited (O'Regan 1992; Bridgeman, van der Heijden & Velichkovsky 1994; Irwin 1993; Intraub, Mangels & Bender 1992).

- There is good reason to believe that information that is currently not on the fovea is stored in a different form from foveal information. Many of the signature properties of early vision — such as spontaneous perception of line drawings as three-dimensional and spontaneous reversal of figures — do not arise when part of the figure is off the fovea (Hochberg 1968). Indeed the non-foveal portion of the display differs in many ways from the foveal information and is much more like an abstract visual memory than a continuation of the foveal display. For example, the ability to construct a phenomenal percept from sequentially presented information — such as occurs in an anorthoscope (Rock 1983) or in the sequential presentation of segments taken from different parts of a picture — depends on the memory load of the information not foveally present (Hochberg, 1968).

Many people have recognized the similarity between theories of mental imagery and theories of visual stability. Indeed, Kosslyn (1994) has explicitly argued that one reason for positing a pictorial display for mental imagery is that it is also required in order to explain the stability and temporal continuity of vision. Thus it is not surprising that the same issues arise. We have already suggested ways in which indexes may play a role in endowing representations underlying mental images with geometrical and metrical properties. Under certain conditions, indexes can also play a crucial role in explaining visual stability and saccadic integration without requiring that we posit an extended pictorial representation. Before discussing how indexes might play a role, consider what the functional requirements are for realizing what we call visual stability. What we need at the very minimum is the following:

- The visual system must be able to keep track of the individuality of objects independent of their location in the visual field. There are several ways in which this could be done.

  ◦ The correspondence of items from one fixation to another might be computed by locating distinctive properties or distinctive objects in the successive views and establishing a mapping between them.

  ◦ The correspondence might arise from a global

recalibration that makes use of efferent information. This is the "corollary discharge" view. Although the idea that some sort of extraretinal signals are involved in visual stability is usually associated with the global image assumption, it is in fact independent of this assumption and it is premature to dismiss it.

- Correspondence maintenance for a small number of objects may simply be a result of a primitive preattentive mechanism of the visual system, such as the FINST indexing mechanism. For this to be possible, the indexing mechanism must be capable of operating at very high speed since saccades are very fast. We shall return to this issue below.

- The visual system must be able to connect seen properties with recalled properties in order to integrate them into some sort of global representation. As we have already noted, however, this global representation will not just be an image consisting of superimposed glances. Indeed the representation will contain abstract visual and also nonvisual information.

- The visual system must be able to establish that an object is the *same* object that was previously viewed, even when eye movements cause the object to leave the visual field briefly and then return. This is one of the most challenging requirements. It is the reason why we have had to posit another type of index, called an *anchor*, which operates in the proprioceptive and motor modalities but which can be cross-linked to visual indexes. Studies of this mechanism are in its infancy and properties of this type of index are still unknown (see, however, Pylyshyn 1989, Table 1b).

As mentioned earlier, the amount of information that is carried over from one glance to another is extremely limited. (Irwin 1995) has shown that the position of at most 3-4 objects is retained from one fixation to another, and this is most likely to include the position to which the eye is about to saccade. Based on such observations (McKonkie & Currie 1995, and Irwin, McConkie, Carlson-Radvansky, & Currie 1994) have argued that on each fixation only one significant benchmark is encoded and on the next fixation a fast parallel search attempts to identify that benchmark, which is then used to calibrate the location in space of other items in that fixation. However the relocation-by-features idea seems implausible, even if it could be accomplished in the short time available, since it ought to lead to frequent and significant errors when the scene is uniformly textured or otherwise free of unique features. For another, it is not clear how the information from a pair of benchmark objects could be used to calibrate locations of the other items in the scene unless it does so by establishing a mapping between two 2D displays — a proposal which the authors themselves eschew. The process transsaccadic integration would be simpler and more reliable if a small number of significant features could actually be *tracked* through the saccade. In doing this they would provide the anchors by which schematically encoded perceptual information could be integrated from fixation to fixation, in a manner suggested by (Pylyshyn 1989) and others (e.g., Intraub, Mangels & Bender 1992).

This proposal, however, raises the important question of whether indexes can remain assigned during a saccadic eye movement. If index maintenance were based on purely local retinal processes, such as those proposed in the network models of (Koch & Ullman 1985) or (Acton 1993), it seems implausible that an index could keep tracking a object moving across the retina at up to about 800 degrees/second — even putting aside the problem of saccadic suppression and smearing. The fastest covert attention movement reported in the literature—and even this has been questioned as being too high—is 250 degrees/sec (Posner, Nissen & Ogden 1978). However, if index maintenance were able to make use of predictive information, such rapid tracking might be possible. There are two main sources of such predictive information. One is extrapolation from portions of the current trajectory, using some sort of adaptive filtering with local data support, as proposed by (Eagleson & Pylyshyn 1988, 1991). The other is extraretinal information such as efferent and afferent signals associated with the eye movement. The appeal to extraretinal signals has usually been associated with metrical superposition theories of visual stability. As we have already noted, the superposition view has been discredited in recent years. Yet the role of extraretinal information in some aspect of transsaccadic integration has continued to be accepted.

If it can be shown that indexing survives saccadic shifts it would provide an important mechanism for transsaccadic integration compatible with current evidence on the subject. This question continues to be an open empirical issue that we are currently planning to examine in our laboratory. In particular, we are planning to ask whether the phenomena that we believe demonstrate indexing — such as subset search, multiple object tracking, and subitizing — survive a saccadic eye movement during the part of the process when indexes are keeping track of the critical visual elements.

Acknowledgments

This research was supported by a grant from the Institute for Robotics and Intelligent Systems (Project HMI-1, through the University of Western Ontario) and the Natural Science and Engineering Research Council of Canada. The work was carried out over the past several years by various members of the UWO research group: B. Acton, J. Burkell, M. Dawson, R. Eagleson, B. Fisher, P. McKeever, W. Schmidt, C. Sears, R. Storm, L. Trick, R Wright.

References

Acton, B. 1993. A network model of indexing and attention. M.A.Sc. Dissertation, Dept of Electrical Engineering, University of Western Ontario.

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

Ballard, D.H., Hayhoe, M.M., & Pook, P.K. 1995. Deictic codes for the embodiment of cognition. Under review by *Behavioral and Brain Sciences*.

Bridgeman, B., Van der Heijden, A.H.C. & Velichkovsky, B.M. 1994. A theory of visual stability across saccadic eye movements. *Behavioral and Brain Sciences,* 17(2), 247-292.

Burkell, J.A., and Pylyshyn, Z.W. 1996. Searching through selected subsets of visual displays: A test of the FINST indexing hypothesis. Submitted.

Cavanagh, P. 1990 Pursuing moving objects with attention. *Proc 12th Annual Meeting of the Cognitive Science Society,* Boston (P1046-1047). Hillsdale, NJ Erlbaum.

Dawson, M.R.W., and Pylyshyn, Z.W. 1989. Natural constraints in apparent motion. In Z.W. Pylyshyn (Ed.), *Computational Processes in Human Vision: An interdisciplinary perspective.* Norwood: Ablex Publishers.

Eagleson, R, & Pylyshyn, Z.W. 1991. The Role of Indexing and Tracking in Visual Motion Perception. Conference on Spatial Vision in Humans and Robots, York University, June 19-22, 1991.

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.*

Farah, M.J. 1989. Mechanisms of imagery-perception interaction. *Journal of Experimental Psychology: Human Perception & Performance, 15(2),* 203-211.

Finke, R.A. 1979. The functional equivalence of mental images and errors of movement. *Cognitive Psychology, 11,* 235-264.

Hikosaka, O., Miyauchi, S. & Shimojo, S. 1993. Focal visual attention produces illusory temporal order and motion sensation. *Vision Research, 33. 1219-1240.*

Hochberg, J. 1968. In the mind's eye. In R.N. Haber (Ed), *Contemporary Theory and Research in Visual Perception.* New York: Holt, Rinehart & Winston.

Intraub, H. Mangels, J., and Bender, R. 1992. Looking at pictures but remembering scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 180-191.

Intriligator, J. & Cavanagh, P. 1992. An object-specific spatial attentional facilitation that does not travel to adjacent spatial locations. *Investigative Ophthalmology and Visual Science, 33,* 2849 (abstract*).*

Irwin, D.E. 1993. Perceiving an integrated visual world. In D.E. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV,* (pp 121-143). Cambridge, MA: MIT Press.

Irwin, D.E. 1995. Properties of Transsaccadic Memory: Implications for Scene Perception. Talk presented at Cambridge Basic Research. June 26, 1995.

Irwin, D.E., McConkie, G.W., Carlson-Radvansky, L.A., & Currie, C. 1994, A localist evaluation solution for visual stability across saccades. *Behavioral and Brain Sciences, 17,* 265-266.

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

Kosslyn, S.M., Ball, T.M., & Reiser, B.J. 1978. Visual images preserve metrical spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 46-60.

Kosslyn, S.M. 1994. *Image and Brain.* Cambridge, MA: MIT Press/Bradford.

McConkie, G.W. & Currie, C. 1995. Coordinating Perception Across Saccades: The Saccade Target Theory of Visual Stability. Talk presented at Cambridge Basic Research. June 26, 1995 (abstract).

McKeever, P. 1991. Nontarget numerosity and identity maintenance with FINSTs: A two component account of multiple target tracking. MA Thesis. Department of Psychology, University of Western Ontario.

Merton, P.A. 1961. The accuracy of directing the eyes and the hand in the dark. *Journal of Physiology, 156,* 555-577.

Minsky, M.L. & Papert, S. 1969. *Perceptrons.* Cambridge, MA: MIT Press.

O'Regan, J.K.. 1992. Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46,* 461-488.

Posner, M.I., Nissen, M.J., & Ogden, W.C. 1978. Attended and unattended processing modes: The role of set for spatial location. In H.L. Pick & I.J. Saltzman (Eds.), *Modes of perceiving and processing information.* Hillsdale: Erlbaum.

Posner, M.I. 1980. Orienting of attention. *Quarterly Journal of Experimental Psychology, 32,* 3-25.

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

Pylyshyn, Z.W. 1994b. Mental Pictures on the brain: Review of *Image and Brain*, by S Kosslyn. *Nature, 372*(6503), 289-290.

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

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

Pylyshyn, Z., Burkell, J., Fisher, B., Sears, C., Schmidt, W. & Trick, L. 1994. Multiple parallel access in visual attention. *Canadian Journal of Experimental Psychology, 48*(2), 260-283.

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

Rock, I. 1983. *The logic of perception.* Cambridge, MA: MIT Press.

Schmidt, W. C., Fisher, B. D., & Pylyshyn, Z. W. 1995. Multiple onset stimuli elicit illusory line motion. Submitted to *the Journal of Experimental Psychology: Human Perception and Performance.*

Sears, C. 1991. Spatial Indexing and Information Processing at Multiple Locations in the Visual Field. MA Dissertation, Dept of Psychology, University of Western Ontario, London, Canada.

Sears, C. 1995. Inhibition of return of visual attention and visual indexing. PhD Dissertation, Department of Psychology, University of Western Ontario, London, Canada.

Sears, C.R.& Pylyshyn, Z.W. 1995. Multiple object tracking and attentional processing. Accepted for publication in the *Canadian Journal of Experimental Psychology.*

Shepard, R.N., & Podgorny, P. 1978. Cognitive Processes that resemble perceptual processes. In W.K. Estes (Ed.), *Handbook of Learning and Cognitive Processes (Vol 5),* Hillsdale, NJ: Erlbaum.

Simon, T. & Vaishnavi, S. 1995. In afterimages, five is too many to count: Implications for the role of object individuation in visual enumeration. In press, *Perception and Psychophysics.*

Treisman, A. 1988. Features and objects. *Quarterly Journal of Experimental Psychology, 40A, 201-237.*

Trick, L. 1990. Subitizing and counting. Paper presented at the annual meeting of the Canadian Psychological Association, June, 1990, Ottawa.

Trick, L. 1991. Three theories of enumeration that won't work and why, and then one that will: Subitizing, counting and spatial attention. In *Nature and Origins of Mathematical Abilities*, J. Campbell (Ed.). Elsevier Press.

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

Trick, L. M., and Pylyshyn, Z.W. 1994a. Why are small and large numbers enumerated differently? A limited capacity preattentive stage in vision. *Psychological Review, 101*(1), 1-23.

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

Tsotsos, J.K., Culhane, S., Wai, W., Lai, Y., Davis, N., Nuflo, F., 1995. "Modeling visual attention via selective tuning", *Artifical Intelligence 78*(1-2), 507 - 547,.

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

Wright, R.D. 1994. Shifts of visual attention to multiple simultaneous location cues. Canadian Journal of Experimental *Psychology, 48,* 205-217.

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

Yantis, S. 1992 Multielement visual tracking: Attention and *perceptual organization. Cog*nitive Psychology, 24, 295-340.