Spatial Vision, Vol. 3, No. 3, pp. 179–197 (1988)

Tracking multiple independent targets: Evidence for a parallel tracking mechanism*

ZENON W. PYLYSHYN and RON W. STORM
Center for Cognitive Science, University of Western Ontario, London, Ontario, Canada N6A 5C2

Received 23 February 1988; revised 8 July 1988; accepted 13 July 1988

Abstract—There is considerable evidence that visual attention is concentrated at a single locus in the visual field, and that this locus can be moved independent of eye movements. Two studies are reported which suggest that, while certain aspects of attention require that locations be scanned serially, at least one operation may be carried out in parallel across several independent loci in the visual field. That is the operation of indexing features and tracking their identity. The studies show that: (a) subjects are able to track a subset of up to 5 objects in a field of 10 identical randomly-moving objects in order to distinguish a change in a target from a change in a distractor; and (b) when the speed and distance parameters of the display are designed so that, on the basis of some very conservative assumptions about the speed of attention movement and encoding times, the predicted performance of a serial scanning and updating algorithm would not exceed about 40% accuracy, subjects still manage to do the task with 87% accuracy.

These findings are discussed in relation to an earlier, and independently motivated model of feature-binding—called the FINST model—which posits a primitive identity maintenance mechanism that indexes and tracks a limited number of visual objects in parallel. These indexes are hypothesized to serve the function of binding visual features prior to subsequent pattern recognition.

INTRODUCTION

It is generally accepted that humans can attend selectively to only one region in the visual field at any one time. Evidence for this belief comes from both psychophysical and neurophysiological studies (e.g., Posner *et al.* 1978; Hoffman, 1979; Schulman *et al.* 1979; Posner, 1980; Jonides, 1983; Prinzmetal and Banks, 1983; Tsal, 1983; Eriksen and Yeh, 1985; Jolicouer *et al.* 1985). There has been some discussion in the literature concerning how broad a region can be covered within focal attention. Some studies suggest that the region might be quite large under certain conditions (e.g., Eriksen and Spencer, 1969; Shiffrin and Gardner, 1972; Shiffrin and Geisler, 1973; Kinchla, 1974; Shiffrin *et al.* 1976). There is even some support for a variable-size attention hypothesis (the 'zoom lens' metaphor of Eriksen and St. James (1986)). Nonetheless, there is general agreement that there is only *one* region of focal attention, as opposed to several independent and noncontiguous regions (see, also, the discussions in Hoffman and Nelson, 1981; Kahneman and Henik, 1981; Laberge, 1983).

In addition to the evidence that there is a single locus of attention, there are also a number of studies showing that this locus can be moved within the visual field independently of any eye movements. The rate of attention movement is very rapid. Estimates range from a low of 33.3 ms/deg (30 deg/s) to a high of 4 ms/deg (250 deg/s).

*Send reprint requests to Prof. Zenon Pylyshyn, Centre for Cognitive Science, Social Science Building, University of Western Ontario, London, Ontario, Canada N6A 5C2.

For example, Posner *et al.* (1978) report attention movement at 4 ms/deg, Tsal (1983) reports 8.5 ms/deg, Shulman *et al.* (1979) report 19 ms/deg, Jolicour *et al.* (1983), using an automatic contour tracing paradigm, found the rate to be 24 and 26 ms/deg, while Eriksen and Schultz (1977) found that attention moved at 33.3 ms/deg. In studies of mental image scanning, Pinker (1980) reports 13 and 18.6 ms/deg, Kosslyn (1978) reports 17 ms/deg, while Finke and Pinker (1982) found the rate to be 19.5 ms/deg. The latter may or may not be relevant, although Pylyshyn (1981) has argued that when subjects scan mental images superimposed on perception they use the same scanning mechanism as in perception proper. Although a few researchers have questioned the evidence on which the continuous-movement assumption rests (Remington and Pierce, 1984; Eriksen and Murphy, 1987; Yantis, in press), the assumption nevertheless seems to be generally accepted.

The view that there is a single region of maximal discriminability, and the view that people can scan something that might roughly be described as "where they are focusing their cognitive effort" seems reasonable in the light of the accumulated evidence. The present paper does not quarrel with this general picture. However, viewing this region of maximal processing as the *locus of visual attention* may be misleading, insofar as it suggests that at any point in time there is no access whatsoever to any place other than the one region in the visual field which is being 'attended' to. We will argue that if one distinguishes several operations that are involved in what is commonly called 'attending', perhaps defining several stages of processing, at least one of these stages can be shown to have more than one independent locus and may thus actually be a 'pre-attentive' stage according to the usual use of this term. The particular stage that we shall be concerned with is the one that maintains the identity of a visual feature as it moves about in the visual field, i.e. a primitive (pre-attentive) indexing and tracking mechanism.

The hypothesis that there is a primitive visual mechanism capable of indexing and tracking features or feature-clusters was first introduced by Pylyshyn *et al.* (1972) and has been further elaborated in Pylyshyn (in press). This work is part of a preliminary exploration of how mechanisms of early vision might allow us to maintain a stable, distally-anchored representative of the world, despite constantly changing retinal stimulation, and how the motor system could be commanded to move in relation to locations sensed by the visual system (the so-called cross-modality binding problem). Although a great many of the details of the model are not germane to the present paper, the following brief summary of some of the reference-binding or indexing assumptions of the model will help motivate the studies to be described herein. These assumptions were initially introduced because they are independently required in order to explain a variety of phenomena, including aspects of perceptual-motor coordination, the interface between top-down and bottom-up visual processing through the invocation of what Ullman (1984) calls 'visual routines', and certain phenomena involving the registration of 'mental' images projected onto actual displays (see Pylyshyn, 1984, pp. 247–251).

The model hypothesizes a certain primitive operation which is preattentively performed upon certain features in the visual display and which is a prerequisite to detecting various relational properties involving those features. This operation consists of assigning an index or internal reference, called a FINST, to the features in question. For example, prior to recognizing that a certain feature is 'inside' or 'above' another one, both features have to be picked out or individuated in some way so that the *inside* or *above* predicates can refer to them. According to the FINST model, this process of individuating the features is carried out by binding one of a small number of available internal reference tokens or pointers, called 'FINSTs', to these features. The mechanism that carries out this binding does so in a way that results in the FINST continuing to point to the same retinal feature as the feature moves about on the retina (where 'same feature' here refers to the sequence of retinal features that correspond to a single distal feature). A FINST is thus a 'sticky' reference token (the name derives from an early observation that these reference tokens act like 'instantiation fingers' that realize reference bindings by, in effect, 'pointing to' distal places). This automatic tracking aspect of the FINST hypothesis suggests that the mechanism for assigning and maintaining these references may be closely related to the mechanism which solves the 'correspondence' problem for apparent motion. Indeed, in the case where motion is not strictly continuous, the tracking aspect of the FINST hypothesis presupposes the solution of the correspondence problem (Dawson and Pylyshyn, 1988).

The hypothesized mechanism for binding FINSTs to features provides a stage of processing which is prior to, and a prerequisite for, such further processing as recognition and discrimination. In contrast with the common view of visual attention focusing (as it appears, for example, in discussions of 'attentional scanning'), FINSTing can occur independently and in parallel at several places in the visual field. In this sense it is a preattentive operation, although the selection of some subset of these automatically indexed places for further processing or tracking may involve deliberate cognitive intervention.

The present study examines this proposal by asking whether subjects can track the movements of multiple independent targets that are visually indistinguishable, as suggested by the FINST hypothesis. If they can, they should be able to detect the occurrence of a distinct visual event located on an object being tracked, and discriminate it from an event located on some other, visually identical object. Note that if subjects *can* successfully carry out this task then this logically entails that they *somehow* have kept track of the targets over time, since the information as to whether an object is a target or nontarget cannot be determined from the current state of the display, nor from some memory trace of an earlier state, but only from the identity over time of the individual targets (i.e. from their historical continuity). An additional study will be described that specifically addresses the question whether this sort of multiple-target tracking can be carried out without serially sampling the set of targets one at a time (i.e. without using a serial time-sharing algorithm), once again as predicted by the FINST hypothesis.

EXPERIMENT I

This study constitutes both a concrete illustration of what the hypothesis claims and a simple test of its plausibility. In itself it is not intended to test all aspects of the hypothesis. In particular it does not address the question of whether FINSTs can be maintained in parallel. We shall see that there exists a serial strategy that could allow tracking of multiple independent targets. This alternative will be discussed in a subsequent section.

If the hypothesis that we can assign a limited number of 'sticky' indexes (FINSTs) to features in a visual display is correct, subjects should be able to track a subset of visually identical and randomly moving objects, providing the target subset is somehow identified at the start of a trial. This, indeed, is explicitly the claim made by the FINST hypothesis which posits a mechanism that can maintain the identity of features independent of their position. The maintenance of identity despite changes in location is precisely what we mean by 'tracking'. Thus if the FINST hypothesis is correct we ought at least to be able to show that subjects can track multiple targets in an animated display containing a number of randomly moving objects, under conditions where targets and distractors cannot be distinguished by local visual properties (i.e. where the targets and nontargets are identical except for their histories).

In this study the number of targets that had to be tracked was varied from one to five. There were always ten objects in all so the number of targets never exceeded the number of distractors (in order to eliminate the more efficient strategy of tracking only distractors). Tracking had to be done without eye movements. To ensure that eye movements were not being used in the tracking task, movement of the eyes was monitored.

Subjects were shown a display consisting of 10 stationary plus (+) signs and were told to note the subset of from 1 to 5 that were flashing. After 10 s the flashing stopped and all 10 objects began moving. The subjects' task was to track the subset that had been flashing, without moving their eyes, and to indicate whenever one of those target objects 'flashed', i.e. briefly changed into a solid square shape. Whenever this happened, subjects were instructed to press a response key. A detailed description of the display and the timing is given below. Both accuracy and reaction times were recorded.

Method

Subjects. Seven University of Western Ontario students, six males and one female, volunteered to participate in the experiment without pay. All subjects had normal or corrected-to-normal vision and had no previous experience in psychophysical experiments of this kind.

Materials. An animated sequence of frames (generated in advance) was displayed by an Apply II + microcomputer on a 50 cm Sony monochrome monitor at a viewing distance of 120 cm. The display consisted of a white fixation square subtending a visual angle of 0.42 deg. This fixation square was presented in the center of the screen with a black background, the background subtending a visual angle of 21.5 deg. The animated stimuli always consisted of ten moving white crosses. A randomly chosen subset of from one to five of the total field of ten objects was designated as targets. The remainder were designated as distractors. Each object subtended a visual angle of approximately 0.42 deg and moved with a velocity and direction that was changed at random every few hundred milliseconds. The velocities of the objects ranged from 1.25 to 9.4 deg/s. The directions were chosen from among 8 equal divisions of the compass. The random motion of the objects was subject to the restriction that no two objects could be closer than 0.75 deg apart, so that the continuity of their identity was never ambiguous (as it would be if they were allowed to collide). In generating the animation sequence to meet this restriction, trajectories were generated for each object. After each frame, the location of objects was tested to determine whether any two objects were too close together. If they were, the last few frames of the sequence were rejected and the generation was restarted at that point with a new random choice of directions and/or velocities. When an object was about to go off the edge of the screen its motion was reflected ('bounced') off that edge.

Each trial consisted of a variable number of animation frames with the duration of the animated portion of the display ranging from 7 to 15 s. After a predetermined time (at least 3 s after the start of the animation), a solid white square, subtending 0.6 deg of visual angle, was flashed over (i.e. replaced) one of the moving objects for about 83 ms. Sometimes the 'flash' occurred over a target and sometimes over a distractor. The target-flash condition (to which subjects had to respond) was sometimes the first flash in a trial and at other times was preceded by from 1 to 3 distractor-flash conditions (each of these 4 conditions occurring equally often), in order to ensure that subjects did not merely respond to any flash. There was only one target-flash condition in each trial. The trial ended after the subject responded to that flash, or at the end of the animation sequence which occurred at least 4 s after the target-flash.

A telegraph key was interfaced to the microcomputer through a CCS 7440A programmable timer which was used to measure the subjects' response latencies. Eye movements were monitored with a Reading Eye Trac II also interfaced to the microcomputer. A trial on which a detectable eye movement occurred (i.e. a movement that was greater than 2 deg off the fixation point) was terminated and the subject informed. An additional trial was then added with the same target size, delay, and type of flash condition, though with new randomly chosen target and distractor placements.

Procedure. The task was explained to the subjects. They were told that after initiating a trial by pressing the space-bar on the microcomputer keyboard, a square would appear in the center of the screen. It was stressed that they must keep their eyes fixated on that square throughout the trial. To help prevent eye movements a chin rest was used and subjects were given feedback whenever an eye movement was detected. Subjects were told that some time while they were tracking the objects a square would flash on the screen and their task was to press a response key as quickly and as accurately as possible if, and only if, the flash occurred at the location of a target. Subjects were provided with 50 practice trials after which they proceeded to initiate the trials at their own speed. There were 10 blocks of 50 trials. All variables were randomized, including the 5 different target set sizes.

Results

Inspection of the response latency data revealed a number of trials with unusually long response latencies. Outliers that were more than two times the interquartile range away from the median were replaced with the median of the observations for that cell (there were 35 such cells, one for each of 7 subjects and 5 target numbers). This procedure is a more conservative correction than that recommended by Tukey (1977) and resulted in 4.9% of the observations being replaced.

In order to determine whether there was a speed–accuracy tradeoff across the 5 conditions, the correlation between response latency and response failure rate was computed for each subject. The lowest such correlation was 0.74, indicating that the pattern of response latency did not reflect a shifting response criterion.

Figure 1 shows the accuracy of responding as a function of number of targets. Overall, subjects failed to respond correctly on only 9.3% of the trials on which a target flashed. The pattern of errors shows that the trials with a larger number of targets had poorer performance. A repeated-measures analysis of variance revealed that the frequency with which subjects failed to respond to target flashes varied with the number of targets ($F(4, 24) = 26.05$, $P < 0.05$). False positive responses (i.e. responses that

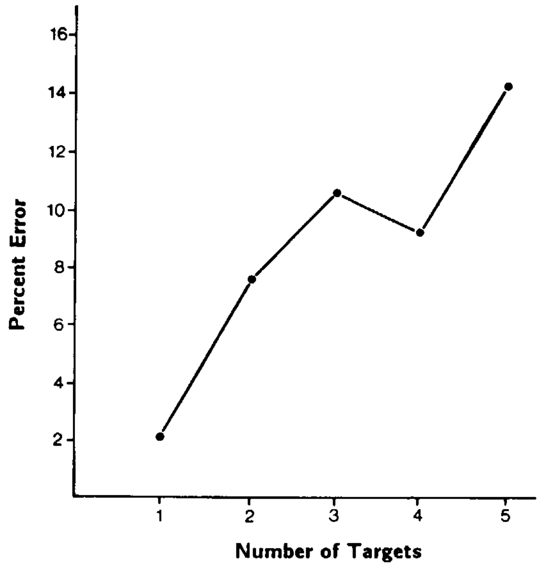

Figure 1. Percent of trials on which subjects failed to respond to a probe flashed on a target, as a function of number of targets.

followed the flashing of a distractor) occurred on only 4.9% of the trials and an analysis of variance showed that their frequency was not significantly affected by the number of targets ($F(4, 24) = 1.96$, $P > 0.05$).

The pattern of response latency scores parallels that for accuracy scores. The latency

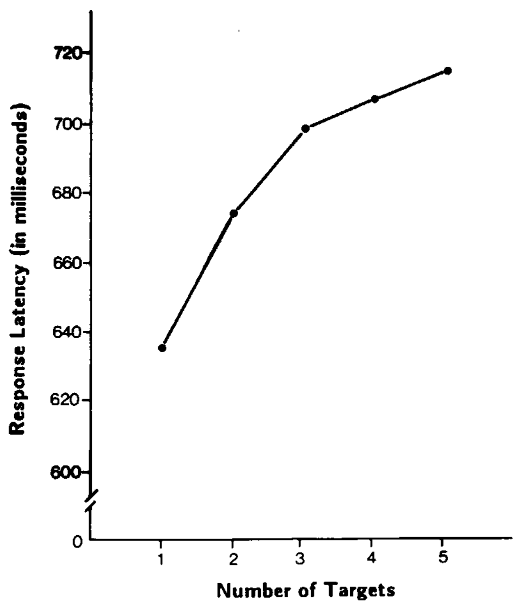

Figure 2. Mean response latencies (correct responses only) to a probe flash occurring on a target, as a function of number of targets.

results are shown in Fig. 2. An analysis of variance showed that the response latencies varied with the number of targets being tracked ($F(4, 24) = 26.11$, $P < 0.05$).

Discussion

The results show that subjects are able to perform the task of tracking multiple independently-moving targets quite well under the conditions of this study, even when there were five such targets (in which case the performance level was still at 85.6%). Individual subjects' error scores ranged from 0 to 8% when tracking one target, and from 6 to 28% when tracking 5 targets.

In appears that, although subjects are able to do this task well, it takes them longer and they make more errors when they are required to track a larger number of targets. This pattern of results is not consistent with the view that the tracking is being performed entirely by a purely parallel process, unless it is a 'limited resource' parallel process, such as one in which the number of targets affects the quality or precision of information about each target (as, for example, in the models developed by Anderson (1973), Townsend (1974) or Townsend & Ashby (1983)). On the other hand, the fact that the task can be done as well as it was, does suggest that *some* parallel processing may be implicated, even if this processing falls in the category of a 'limited resource' parallel process. In order to examine this question in detail, it is first necessary to consider whether the ability to track multiple targets, demonstrated in this study, can be explained entirely in terms of a serial scanning process which, for example, 'time-shares' among the separate targets. This question is discussed in the next section, prior to introducing a second study designed to address the question more directly and quantitatively.

THE SERIAL TRACKING ALGORITHM

In the tracking task described earlier, each object in the display can be distinguished from other objects by one of only two ways. The first is by the historical or temporal continuity of its identity, which can be traced back to its initial location or other distinctive property (e.g., whether it was flashing) at the start of the trial. The second is by its current properties. But the only currently distinguishing properties of any object are its location or its motion, since the objects are identical with respect to all other properties. An algorithm which uses the first method of tracking objects (i.e. keeping track of them in terms of their history) must be able to trace the identity of all target objects simultaneously, i.e. it must be a parallel algorithm. An algorithm which uses the second logically possible characteristic to keep track of objects, i.e. their current location or motion, must, on the other hand, continually update these changing characteristics by sequentially sampling the objects. This is the basis for the serial tracking strategy.

In order to design a test which rules out the serial tracking algorithm it is necessary to show first how the maximum performance attainable by such an algorithm depends upon a number of temporal and spatial factors, such as the distances between targets, the velocities at which the targets move, the velocity at which attention can move, the time spent processing information about the targets, and the type of information the algorithm uses to identify targets (e.g., location and/or motion). The serial tracking algorithm described below uses target location to keep track of targets, although we will modify it shortly so that the target location that is stored is one that is predicted by extrapolating the current motion vector. Motion information could also be incorporated into the algorithm in other ways, and this will be discussed later.

The steps that a serial mechanism would need to go through in order to keep track of more than one target in a tracking task are as follows.

Serial tracking algorithm

(1) The visually distinct targets (e.g., ones that are flashing at the start of the trial) are initially sampled by moving attention to each target in turn. When attention is focused on a target, its position is encoded (according to some location code such as its Cartesian coordinates) and stored in a 'target location table' in memory. All targets must be so encoded during the initial setup period.

(2) During scanning, the location of each target is retrieved in turn from the target location table in some prescribed order.

(3) In accordance with the most widely accepted version of the serial scanning view, we assume that in going from its current location to the specified location of the next target in the sampling sequence, attention moves in a continuous fashion, passing through all intermediate points in the display. Because of this the time for this step increases monotonically with increasing inter-target distance.

(4) After attention is moved to the next location retrieved from the target location table, the object nearest to this selected location is taken to be the next target in the sampling sequence (locating the nearest object at this step may involve some local searching as well).

(5) The current location of this object is encoded and stored as the updated location of that assumed target.

(6) Steps 2 through 5 are repeated until the tracking task is terminated by the detection of a flashing object. The detection of a flash is assumed to be a separate process which occurs in parallel with this algorithm (i.e. it is 'interrupt driven'). It is also assumed that when a flash is detected, attention is moved to the location of the flash. If that location is one that corresponds to a location in the target location table, a 'yes' response is emitted. Otherwise a 'no' response occurs.

The tracking ability of a system using the above algorithm will be poor if a target cannot be correctly located in step 4 of the procedure. A number of factors influence the accuracy with which a target can be located. For example, if the targets move quickly, location information in memory is likely to be outdated by the time the target is sampled again at some later time. Similarly, if the distance between objects is decreased, the likelihood of mistaking some nearby object for the intended target in step 4 of the algorithm will increase. Thus the performance of the serial model is adversely affected by increasing the length of the path joining all the targets, increasing the number of targets being tracked, increasing the density of objects in the display, and increasing the velocity of the objects. These factors affect the performance of a serial model by influencing the speed with which information about the display is outdated relative to the sampling rate of the serial mechanism.

Quantitative predictions of the serial model

In the next experiment, these factors were taken into account in order to design a display that would result in poor expected performance for the serial scan algorithm. The initial parameters for the display used in Experiment II were selected based on the experience with Experiment I and based on some rough estimates of the performance that might be expected if the speed of attention movement was in the order of 20 ms/deg (a figure that is roughly around the median of the speeds reported in the studies cited in the Introduction). It appears that our provisional adoption of a display consisting of 4 targets and 4 distractors, a mean velocity of 8 degree/sec, and the restriction that there always be an object within 1.5 deg of each target does indeed tax the serial scan algorithm. These parameters turned out to lead to predicted tracking performance for the serial scan algorithm that did not exceed 20% accuracy, as determined by our more detail calculations described below. The predicted performance was calculated by examining the actually sequence of animation frames generated for Experiment II (to be described in the next section), based on this provisional design. The method of calculating the expected performance is described below. It should be emphasized, however, that the actual figures that are used in these calculations are ones that apply to the experiment that will be described in detail in the next section.

The tracking task carried out by the serial algorithm can be divided into three stages. Once the objects are set into motion there is a probability that a target will be lost the first time it is sampled. This depends on how soon the target is sampled after the display is set into motion, which is turn depends on its serial position in the sampling order. Next, once all targets have been sampled at least once there is a probability associated with losing a target during each subsequent sampling. Finally there is a probability that the target will be lost after its last sampling when a flash is detected and attention moved to check whether or not the flash occurred on a target. The probability of losing a target during this last phase also depends on the serial position of the target in the sampling order. The probability of keeping track of a target $P(K_{i,f})$ over an entire trial can be expressed in terms of its serial position in the scan sequence i and the length of the interval between samplings f as follows (for convenience, we express all intervals initially in terms of 'number of animation frames'):

$$P(K_{i,f}) = [1 - P(C_b)]\,[1 - P(C_f)]^n[1 - P(C_e)] \tag{1}$$

where $P(C_j)$ is the probability of confusing the intended target with another object after j frames have gone by (i.e. when the stored location is j frames old), f is the number of frames between successive samples of the same target, i is the serial position of the target in the sampling order (in Experiment II we will have, $1 < i < 4$), b is the number of frames prior to the first sampling of target i (in Experiment II we will have $b = (i-1)*f/4$), n is an integer representing the number of times target i is sampled (in Experiment II we will have $n = \text{int}((59 - b)/f)$, since there are an average of 59 frames in the animated sequence), and e is the number of frames between the last time target i was sampled and the termination of the trial by a flash (in Experiment II we will have $e = 59 - n*f - b$).

Once we know f (which, in Experiment II was a mean of 59 frames, each frame being 50 ms), Equation (1) gives us the probability of successfully tracking a target through an entire trial, expressed in terms the serial position that the target could occupy and the function $P(C_j)$. If we were given a particular value of f and had a way of computing $P(C_j)$, we could calculate a value of $P(K_{i,f})$ for each of 4 serial positions i that the target could occupy. These 4 values could then be averaged to give $P(K_f)$, the probability of tracking a single target through the given f frame sampling cycle. Assume for now that we have $P(C_j)$ and hence can compute $P(K_f)$. (We will return to the question of how we obtain $P(C_j)$ shortly.)

In order to compute expected performance we also need $P(T_{j,f})$, the probability that exactly j objects are successfully tracked by the serial algorithm when the inter-sample interval is f frames. Since the targets are independent, this is obtained from,

$$P(T_{j,f}) = P(K_f)^j \tag{2}$$

From which, finally, we can express the expected performance $E(f)$ of the serial algorithm, on 'target' trials, in terms of the intersample interval f as follows:

$$E(f) = \sum_{j=1}^{4} P(T_{j,f}) \times \frac{j}{40} \tag{3}$$

where j is the number of targets successfully tracked over a trial, and $P(T_{j,f})$ is the probability that j targets are successfully tracked throughout an f frame inter-sample separation.

From this we can go on to express the expected performance as a function of scanning rate, which was our original goal. To complete this computation we need to known two final quantities: (a) the distances (in degrees) that must be scanned in the f frames between samples; and (b) the probability $P(C_j)$ that a target will be lost if it is not sampled during some interval of j frames. Both these quantities were determined empirically by examining the actual animation sequences used in the experiment.

The first quantity—the distance attention must travel in j frames—was measured by making the conservative assumption that subjects scan the targets using the shortest of the 3 possible path lengths connecting all 4 targets. The database that generated the experimental trials was then consulted and actual distances (converted to degrees) measured. The mean of the shortest paths was used in computing the mapping from number of frames f to the scanning velocity (or more precisely, the reciprocal of the velocity, expressed in milliseconds per degree).

The second quantity—the probability of losing a target as a function of length of interval—was also found by using the actual database of animation sequences for each trial. The procedure consisted of stepping through each frame of the animation database and counting the number of times that selecting an object in step 4 of the serial algorithm (by taking the object closest to the stored previous location of the assumed target) would result in picking the wrong object. These probabilities were stored in a table and used in computing the estimated probability of correctly tracking all 4 targets, stated in terms of f and hence in terms of scanning velocity. The result of these calculations is expressed graphically in Fig. 3 (note that Fig. 3 also contains some information that will be discussed in the Discussion of Experiment II section of this paper, but is included in the same diagram for comparison purposes).

There are a couple of things to note here. First is that the graphs in Fig. 3 are not strictly predictions of the performance expected in the experiment to be described, since subjects could possibly still use a guessing strategy in cases where the tracked targets are lost. The expected performance would then depend on the number of different responses that could be made and on any response biases that exist. We shall return to this possibility in discussing the results of Experiment II. In the meantime, a useful way to view the graphs generated from the equations discussed above, is as the predicted tracking ability of the serial algorithm as a function of scanning velocity.

The second thing to note is that so far we have assumed that the serial tracking algorithm uses only an encoding of the last target location to relocate the target in step 4 of the algorithm. Perhaps if we took into account the speed and direction of a target, and if we extrapolated that motion over time to predict where the target might be on the

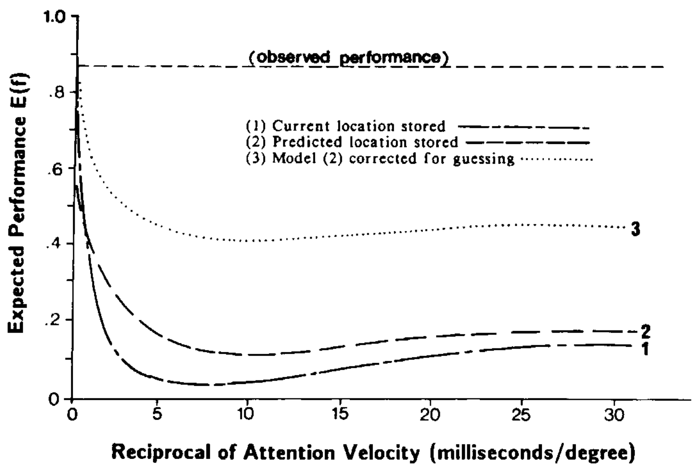

Figure 3. Expected performance of the Serial Tracking Algorithm (described in the text), as a function of the (reciprocal) of the speed of attention scanning. The data were calculated by running simulations of the serial tracking algorithm over the actual stimulus sequences used in the experiment. Predictions use the standard model (curve 1), a model which stores the predicted future location of the sampled point (based on its velocity and direct) (curve 2), and a model which guesses randomly among the three available responses when it detects any probe and also detects that it has failed to correctly track all targets (curve 3).

next sample, this might provide a more precise way of relocating the target than just using the target's current location. To see whether this additional source of information would improve the expected performance of the serial algorithm, predictions were computed for this modification of the algorithm. The method used is the same as before only instead of recovering a target by its old location, the velocity vector was also extracted from the display and extrapolated to predict where the target would appear on a later sampling. The object nearest the extrapolated location was then taken to be the target.

Figure 3 shows that using both location and velocity vectors to compute a predicted location which is then stored in the target location table, leads to only marginally better performance than storing only the current location information. (The fact that performance based on location information alone surpasses performance based on motion and location when the velocity of attention is less than 1 ms/deg, is attributable to the fact that it takes two 50 ms-frames to compute the velocity vector, so that the motion-based procedure in effect has a longer 'dwell time'.) The reason is that velocity vectors do not serve as good predictors of the target's global trajectory in our current design because the objects' velocity changed often. Indeed, an examination of the object trajectories revealed that the difference between the direction that an object was moving at time t and the direction of its actual location after time $t + d$ (with d chosen to correspond roughly to the time it would take to scan 4 targets at 20 ms/deg) shows the relation to be quite poor. The distribution of direction differences was very nearly uniform in the range from 1 to 180 deg and the correlation between these two directions is only 0.11. In other words the direction in which an object is moving now is a poor predictor of the direction it ends up in by the time the algorithm gets around to sampling it next.

A final point about Fig. 3 is that the probability of correctly tracking all targets is not precisely a monotonically increasing function of scanning speed, as one might have expected. It appears that there is an anomalous dip in predicted tracking accuracy at scanning velocities of around 5–10 ms/deg. This is very likely an artifact of the particular distributions of velocities and trajectories actually used in this experiment. Recall that the trajectories were generated algorithmically subject to a variety of constraints, and that records of actual trajectories were used in computing Fig. 3. The procedure for generating trajectories consisted of attempting a random Brownian motion of all points (targets and non-targets) and looking for violations of such restrictions as those on the minimum and maximum distances allowed between points. When a violation was detected, the last few frames were deleted and the generation resumed. As a result of such a procedure, the resulting distribution of such properties as direction, length, and velocity of the set of straight trajectories-segments was skewed. For example, there appear to be more parallel segments than expected from a random ensemble, since these are more stable under the restricting criteria. Although such statistical distortions are inherent in the methodology adopted, they could easily have resulted in distributions that penalized predicted performance at certain scanning speeds. This would be true, for instance, if the distribution of distances between targets and nontargets had a peak around the distance that the average target moved in the time to scan a complete path at scan velocity of 5–19 ms/deg. Such distributional anomalies are known to exist in our trajectories and appear to be the most likely explanation for the departure from monotonicity of the curves in Fig. 3.

EXPERIMENT II

Experiment II was similar to Experiment I except that, as discussed in the previous section, the display was designed so that the predicted performance of the serial scan algorithm would be poor. A display consisting of 4 targets and 4 distractors was adopted. The separation among the 4 targets was such that the mean path length required to scan all 4 targets was 34 deg of visual angle (computed as the mean, over trials, of the shortest connecting path possible on each frame). By adopting a rather fast mean velocity (8 deg of visual angle per second) and by arranging for each target to be at all times close to another object (i.e. no more than 1.5 deg away) in order to increase the probability that a target would be misidentified in step 4 of the serial algorithm, it was possible to decrease the expected performance of the serial model to well below that observed in Experiment I, as discussed in the previous section and shown in Fig. 3.

Method

Subjects. Eleven University of Western Ontario graduate students, seven male and four female, served as paid subjects. All subjects had normal or corrected to normal vision and none had participated in the previous experiment.

Materials. The display was the same as that used in Experiment I with the following exceptions. There were a total of 8 objects, four of which were designated as targets. The speed of movement of the objects, as well as the frequency with which they changed speed and direction, was somewhat higher in this experiment (the animation frame rate was higher in this study than in the previous one). The speed and direction of motion of the objects was changed every 100–150 ms. The velocities of the objects ranged from 2 to 15 deg/s, with a mean of about 7.6 deg/s. Unlike the first study, the maximum distance between any two objects, as well as the minimum distance, was constrained. The distance between objects was kept between 0.75 and 1.5 deg.

As in Experiment I, each trial consisted of a variable number of animation frames. In the present experiment, however, the maximum duration of the animated portion of the display was only 4 s. After a variable number of frames, a solid white square appeared on the screen for about 50 ms. In contrast with Experiment I, there were three types of flash conditions classified in terms of where the flash could occur. It could occur at the location of a target, a distractor, or some other randomly chosen location on the screen, at least 0.75 deg away from any object.

To allow the subject to indicate which of these possible locations were occupied by the flash, three telegraph keys were interfaced to the microcomputer in the same manner as in the previous experiment. Once again eye movements were monitored and trials on which eyes moved were terminated and a new trial added to the block of trials.

Procedure. The instructions and experimental procedure were the same as in Experiment I, except for the use of 3 response keys (in addition, the pre-animation time during which subjects picked out the 4 targets was reduced to 5 s). Subjects were given 60 practice trials, 20 with each type of flash, and were then instructed to proceed by initiating the trials at their own pace. They were then given 3 blocks of 90 trials and 2 blocks of 30 trials (a total of 110 trials under each of the 3 conditions).

Results

Outliers were replaced using the same procedure employed in Experiment I. This resulted in the replacement of 4.7% of the observations. In order to determine whether there was a speed-accuracy tradeoff, a correlation was computed between standardized response latencies and the frequency of correct responses to flashes. The correlation was based on 33 data points obtained across 11 subjects and 3 stimulus flash types (z-scores were used in order to remove between-subjects variance in this analysis). The resulting correlation was 0.52, indicating that the difference in response choices and response latencies did not result from a speed–accuracy tradeoff.

Table 1 presents the confusion matrix for the various types of stimulus conditions, with the rows denoting the three stimulus flash conditions, and the columns denoting

Table 1.
Confusion matrix of stimulus and response types (percents are based on all responses given for that stimulus type).

Stimulus type	No response	Target	Distractor	Other	Total
Target	14[a]	1042 87.1%[d]	147 123%[g]	7 0.6%[j]	1196
Distractor	21[b]	149 12.5%[e]	988 83.1%[h]	52 4.4%[k]	1189
Other	5[c]	51 4.4%[f]	59 5.0%[i]	1059 90.6%[m]	1169
Total	40	1242	1194	1118	

Selected comparisons (*a priori* t-tests), significant ($P < 0.05$): a-c, b-c, d-h, h-m, j-k. Non-significant ($P > 0.05$): a-b, d-m, g-c, f-i the responses made to each flash type, including failure to respond. The matrix shows data collapsed over 11 subjects. *A priori t*-tests were performed between various cells in this table and are presented at the bottom of the table.

The first column shows subjects' failure to detect the three different types of flashes. The pairwise comparisons indicate that randomly generated flashes are detected more frequently than either flashing targets or flashing distractors. There was no significant difference between detecting flashing targets and flashing distractors. These comparisons also indicate that correct responses to random flashes and target flashes are more frequent than correct responses to distractor flashes, and that target flashes are less likely to be confused with random flashes than are distractor flashes. The reverse should occur if there were a bias towards distractor responses. The comparison between the frequency of the 'target', 'distractor' categories when the probe was on neither also reveals that there is no significant response bias towards either distractor or target responses.

The figures show clearly that a high level of performance (83–88%) was attained with these displays. A comparison with Fig. 3 shows that the performance was much higher than predicted from the serial model, even if the attention scanning speed were equal to the fastest rate reported in the literature. For example, at the fastest scanning rate of 4 ms/deg, reported by Posner *et al.* (1979), performance predicted by the serial scanning algorithm is less than 8% if only location information is encoded and used to re-identify targets, and less than 20% if velocity information is used to compute a predicted location, and thus help in relocating targets in step 4 of the serial algorithm. Clearly, predictions of the serial scan algorithm are inconsistent with the obtained results. Some possible reasons for this discrepancy are considered in the next section.

Discussion of Experiment II

Although the results of Experiment II provide strong evidence that the serial algorithm discussed earlier is not capable of yielding the observed performance, there are a number of other considerations that need to be addressed. One set of considerations concerns a possible modification of the serial algorithm, specifically in the use of a different method to relocate targets in step 4. Other questions concern the possibility that the target might be accidentally recovered or that there might be a trade-off between speed of scanning and time for encoding. These two questions were explored by simulating the serial algorithm and recording performance. Another issue concerns the possibility that a guessing strategy, used in conjunction with the serial scan algorithm, might lead to high performance. A final issue concerns the possibility that tracking was accomplished by a mixture of parallel and serial methods, specifically by a process of tracking some small subset of targets in parallel and then of serially attending to the remaining targets. These questions will be addressed in this section.

Modification of the serial algorithm. Several possible modifications of the serial algorithm were examined. The first is the question of whether there is other information that can be used to help locate targets in step 4 of the serial algorithm. In the calculations so far velocity information has only been used in order to aid in predicting target location. But the velocity of a target might also be used more directly in helping to distinguish targets from nontargets. Instead of just recovering the nearest object, the nearest object with a motion similar to the target sampled earlier could be recovered. Such a method would only work, however, if the velocity of a target was a reliable discriminative property of that target. Since targets change their velocity every 50–150 ms, this is not a reliable method. Indeed, the correlation between the direction of velocity vectors that are separated by approximately the time needed to scan attention between all targets at 20 ms/deg, was 0.084.

Additional questions explored by simulations. Two additional questions were explored by simulating the serial algorithm and applying it to the database of animation sequences used in Experiment II. The first was the possible trade-off between scanning speed and time taken to encode the location of each target. An extreme case was simulated, in which it was assumed that attention could be shifted instantaneously and that a dwell time of 100 ms was required to compute and encode the location of each object and its velocity vector. In Experiment II this is in fact the *minimum* time that a subject could take since determining the velocity requires that at least two frames be sampled and this takes 100 ms. In this simulation, it was found that a target flash was correctly identified on only 30.6% of the target trials.

The second question explored by simulating the serial tracking process was the possibility of accidental target recovery during the execution of the serial algorithm. It should be noted that the expected performance levels shown in Fig. 3 were computed assuming that a lost target is never recovered. Although there is no reliable way of recovering a lost target, such a target may nonetheless still be recoverable by chance. For example, a distractor may be selected on one sampling rather than the real target but on a later sampling a target might be recovered by chance instead of the distractor being erroneously tracked. This possibility was not taken into account in Equation (3). It is possible, however, to determine the likelihood of such chance recovery by simulating the serial model and running it on the actual animation sequence. Because of the large amount of processing involved, the stimulation was run only for a scanning speed of 20 ms/deg, a figure that represents close to the median of the scan rates reported in the studies cited in the Introduction. The simulation used both the location and velocity vector of the target to predict where the target would be on a later sampling, since that had yielded slightly better performance in our earlier calculations. In this simulation the serial algorithm correctly identified target flashes in 398 out of 1000 target trials, yielding a performance or 39.8%. Although this is about twice as good as the prediction based on Equation (3), it still falls far short of the observed performance level.

Use of a guessing strategy. Although formula 3 provides a method of computing the probability that the targets are successfully tracked, subjects' performance may also reflect a strategy of guessing in cases where the target has been lost. Two relevant points should be noted from the data in Table 1: (1) There appears to be no systematic bias in the use of the three response categories; (2) subjects are able to discern whether or not a flash had occurred, even when they were wrong about whether it occurred on a target, distractor, or neither (since the number of cases of failure to respond was extremely low). It is natural to assume that subjects might guess whenever they were uncertain as to where the detected flash occurred, thus raising their performance levels. If that is so, then we may assume that on those occasions where a subject detects a flash and yet the target has not been successfully tracked, the subject will select one of the three possible responses with equal probability—and will thus select the correct 'target' response on 1/3 of those trials. Thus the expected performance, allowing for guessing, will be:

$$E_g(f) = E(f) + [1 - E(f)]/3 \tag{4}$$

Where $E_g(f)$ is the expected performance on a 'target' trial, corrected for guessing, and $E(f)$ is the probability that the target in question has been successfully tracked.

The resulting values of $E_g(f)$ are shown in Fig. 3, plotted as a function of scanning speed. They show that even with guessing, the performance of the serial algorithm is clearly inferior to that of our subjects. Note that comparing observed performance against this measure represents an extremely conservative test of the serial algorithm, since this estimate is sure to be an inflated one. The reason it is inflated is that it assumes: (1) that the subject can detect the occurrence of a flash event when its location is not known; and (2) the subject is prepared to guess on just those trials where a mistracking has occurred. Assumption (2) is unrealistic since the subject cannot in general tell when the serial algorithm has gone astray. However, if the subject guesses on other (i.e. correct) trials the performance will decrease, eventually regressing towards a mean value of 0.33.

A possible hybrid tracking strategy. A final possibility that might be considered is that, although parallel tracking may be occurring, it may involve only a subset of the 4 targets. Perhaps subjects can track, say, two targets and guess whenever a flash occurs that is not on one of the targets being tracked. Since we found in Experiment I that tracking performance on a small number of targets is higher than on a larger number of targets, perhaps it is possible to optimize performance when four targets are presented by tracking only a subset. When tracking fewer targets, the scan path is shortened and the travel time of attention is shorter, permitting targets to be sampled more quickly and thus reducing the probability of an error in step 4 of the serial algorithm. The improved performance on the tracked subset would, of course, be offset by a higher error rate on the non-tracked targets. Nonetheless, the question arises whether there might not be a net benefit from such a strategy, which would require that the subject guess on the cases where the flash was not on the subset being tracked. The expected performance $E_s(t)$ for tracking subsets t of 1, 2 and 3 targets was computed as follows:

$$E_s(t) = (0.979)^t \times (t/4) + (0.333) \times [(4 - t)/4] \tag{5}$$

This equation contains two terms, the first term reflects expected performance if the flash occurs on one of the targets being tracked, and the second term reflects expected correct performance when a flash occurs on a target not being tracked. These are weighted according to the probability of the occurrence of these two types of events. The proportion of trials in which a tracked target flashes is equal to the size of the subset t being tracked divided by the number of targets in the display (i.e. 4). The estimated probability of a correct response to a flashing target when the corresponding target is actually being tracked is based on the results of Experiment I, where it was found that when only one target was being tracked, a flash was correctly detected 97.9% of the time. For lack of better evidence this is taken as our best estimate of performance in recognizing targets actually being tracked. For purposes of this calculation, we assume that when a flash occurs on a target not being tracked, the subject simply guesses. Since there are three possible responses (which occur approximately equally often), we assume that the guess is correct on 1/3 of the target trials.

Using this estimator, the expected number of correct responses on target trials was computed when tracking subsets of different sizes. The resulting predictions were for a performance of 49.4% when tracking only one target, 64.6% when tracking two targets, and 78.6% when tracking 3 targets. Comparisons of these performance scores with the observed performance (87.1%) revealed that the former were all significantly lower than the observed score ($P < 0.05$). Thus the strategy of tracking a subset of the targets and guessing on the remainder would not in general lead to better performance. (Note, by the way, that if Equation (4) is used to calculate the expected performance for tracking all four of the targets, the resulting performance level, 91.9%, does not differ significantly from the observed score ($P > 0.05$). This at least shows that extrapolating the prediction of Equation (5) to the full set of 4 targets does not lead to a contradiction.)

GENERAL DISCUSSION AND CONCLUSIONS

The results of these experiments show that subjects are able to accurately track up to 5 independently moving targets, and that this ability cannot be explained entirely in terms of a serial attention scanning process—at least not of the kind that has been investigated in experiments dealing with the movement of a locus of maximum discrimination. On the other hand, the increase in reaction time and drop in accuracy that occurs when the number of targets being tracked is increased, shows that either *some* serial processing is occurring, or else there is an interaction between global attentional load and the rate at which information associated with individual targets is processed. The latter approach (advocated in certain applications by Townsend, 1974; Townsend & Ashby, 1983) amounts to hypothesizing a resource-limited parallel process.

The present data are compatible with a parallel but limited-resource process, though not with a strictly serial process of the sort that has become widely accepted. However, even if one adopts a limited-resource parallel process explanation of the present results, it is still of interest to ask what sort of mechanisms may be responsible for the trade-offs that have to be assumed in this kind of model. For example, it would be desirable to develop a detailed computational model that specifies which computational resources are limited and shows how certain resource-allocation strategies lead to a slowdown in the individual processing speeds of parallel processes when the overall load is increased on the system (i.e. when more objects must be tracked).

The FINST model discussed in the introduction, which was developed for independent reasons, suggests an alternative analysis which does not need to assume that an increasing load on a parallel process results in a deterioration in performance. The alternative is to factor the tracking task into a limited number of parallel components, each with unlimited capacity within the bounds of this task, and a more conventional serial component. The model hypothesizes a primitive parallel mechanism which can 'track' visual features in a special sense, namely it provides a set of pointers or indexes (called FINSTs) that can be bound to visual features in such a way that they remain bound as the features move about.

The notion of a pointer (or FINST) plays the same role here as it does in computer systems. In neither case does a pointer directly provide information about the objects being pointed to; it merely provides an access path to such information for any process which can follow the path. For example, in our case the pointer does not directly encode such information as the type of feature it is pointing to, or the number of such features, or whether one or more of the features has changed. To determine those properties it may be necessary to applying a serial process to the set of objects pointed to. For instance, this is clearly the case for the property 'number of objects'. To determine how many objects are being FINSTed, it appears to be necessary to scan the FINSTed objects one at a time and count them at least that is the natural interpretation of the results of subitizing studies such as those reported by Klahr (1973) or Klahr and Chi (1975). Similarly, it may be that determining whether one of the FINSTed objects has just flashed may require that each of the objects be checked serially, following the access path provided by the FINST indexes.

This view has the advantage that it accounts for both the phenomena reported in this study. It accounts for the fact that the ability to track multiple targets does not appear to depend upon serially scanning attention to all the objects being tracked. Thus maintaining the identity and location of the subset being tracked appears to be carried out by a parallel process. On the other hand, the FINST hypothesis also explains why it takes longer to detect a change in one of the targets when there are more targets. This, presumably, is due to the fact that when the flash occurs, each of the FINSTed objects must be checked serially to determine whether *it* was the one that flashed, a process that takes more time the more such targets there are.

Whether the two stage approach advocated here, or some limited-resource parallel process provides the best explanation of the present data remains an open research question. From our perspective it is interesting that the FINST model, initially developed to serve rather different purposes, suggests that it may be useful to distinguish at least two stages in multiple-target visual tracking; one a parallel preattentive indexing stage and the other a serial checking stage invoked in selecting a response.

Acknowledgements

This research was supported by Grant A2600 from the Natural Science and Engineering Research Council of Canada, awarded to the first author.

REFERENCES

Anderson, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychol. Rev.* **80**, 417–438.

Dawson, M. R. and Pylyshyn, Z. W. (1988). Natural constraints in apparent motion. In: *Computational Problem in Human Vision: An Interdisciplinary Perspective*, Z. W. Pylyshyn (Ed.), Ablex, Newark, N. J.

Eriksen, C. W. and Murphy, T. D. (1987). Movement of attentional focus across the visual field: A critical look at the evidence. *Percept. Psychophys.* **42**, 299–305.

Eriksen, C. W. and Schultz, D. W. (1977). Retinal locus and acuity in visual information processing. *Bull. Psychonomic Soc.* **9**, 81–84.

Eriksen, C. W. and Spencer (1969). Rate of information processing in visual search: Some results and methodological considerations. *J. Exp. Psychol. Monograph* **79**, (2, part 2), 1–16.

Eriksen, C. W. and St. James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Percept. Psychophys.* **40**, 225–240.

Eriksen, C. W. and Yeh, Y. (1985). Allocation of attention in the visual field. *J. Exp. Psychol.: Human Percept. Perf.* **11**, 583–597.

Finke, R. A. and Pinker, S. (1982). Spontaneous imagery scanning in mental extrapolation. *J. Exp. Psychol.: Learning, Memory Cognition* **2**, 142–147.

Hoffman, J. E. (1979). A two-stage model of visual search. *Percept. Psychophys.* **25**, 319–327.

Hoffman, J. E. and Nelson, B. (1981). Spatial selectivity in visual search. *Percept. Psychophys.* **30**, 283–290.

Jolicouer, P., Ullman, S. and Mackay, M. E. (1985). Boundary tracing: An elementary visual process. *Memory Cognition* **4**, 219–227.

Jonides, J. (1983). Toward a model of the mind's eye. *Can. J. Psychol.* **34**, 103–112.
Kahneman, D. and Henik, A. (1981). Perceptual organization and attention. In: *Perceptual Organization*, M. Kubovy and J. R. Pomerantz (Eds). Lawrence Erlbaum, Hillsdale, N. J.
Kinchla, R. A. (1974). Detecting target elements in multielement displays: a confusability model. *Percept. Psychophys.* **15**, 149–158.
Klahr, D. (1973). Quantification processes. In: *Visual Information Processing*, W. G. Chase (Ed.). Academic Press, New York.
Klahr, D. and Chi, M. T. H. (1975). Span and rate of apprehension in children and adults. *Bull. Psychonomic Soc.* **7**, 434–439.
Laberge, D. (1983). Spatial extent of attention to letters and words. *J. Exp. Psychol.: Human Percept. Perf.* **9**, 371–379.
Pinker, S. (1980), Mental imagery and the third dimension. *J. Exp. Psychol.: Gen.* **109**, 354–371.
Posner, M. L. (1980). Orienting attention. *Q. J. Exp. Psychol.* **32**, 3–25.
Posner, M. I., Nissen, M. J. and Ogden, W. C. (1978). Attention and unattended processing modes: The role of set for spatial location. In: *Modes of Perceiving and Processing Information*, H. L. Pick and I. J. Saltzman (Eds.). Lawrence Erlbaum, Hillsdale, N. J.
Prinzmetal, W. and Banks, W. P. (1983). Perceptual capacity limits in visual detection and search. *Bull. Psychonomic Soc.* **21**, 263–266.
Pylyshyn, Z. W., Elcock, E. W., Marmor, M. M. and Sander, P. T. (1978). *Proceedings of the Second International Conference of the Canadian Society for Computational Studies of Intelligence.* University of Toronto. (Available from The Report Store.)
Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychol. Rev.* **88**, 16–45.
Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science.* MIT Press/Bradford Books, Cambridge, MA.
Pylyshyn, Z. W. (in press). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, in press.
Remington, R. and Pierce, L. (1984). Moving attention: Evidence for time-invariant shifts of visual selective attention. *Percept. Psychophys.* **35**, 393–399.
Shiffrin, R. M. and Gardner, G. T. (1972). Visual processing capacity and attentional control. *J. Exp. Psychol.* **93**, 72–82.
Shiffrin, R. M. and Geisler, W. S. (1973). Visual recognition in a theory of information processing. In: *The Loyola Symposium: Contemporary Issues in Cognitive Psychology*, R. Soslo (Ed.). Halsted Press, Washington, D. C.
Shiffrin, R. M., McKay, D. P. and Shaffer, W. D. (1976). Attending to forty-nine spacial positions at once. *J. Exp. Psychol.: Human Percept. Perf.* **2**, 14–22.
Shulman, G. L., Remington, R. W. and McLean, J. P. (1979), Moving attention through visual space. *J. Exp. Psychol.: Human Percept. Perf.* **5**, 522–526.
Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In: *Human Information Processing: Tutorials in Performance and Cognition*, B. H. Kantowitz (Ed). Lawrence Erlbaum, Hillsdale, N. J.
Townsend, J. T. and Ashby F. G. (1983). *Stochastic Models of Elementary Psychological Processes.* Cambridge University Press, Cambridge, U.K.
Tsal, Y. (1983). Movements of attention across the visual field. *J. Exp. Psychol.: Human Percept. Perf.* **9**, 523–530.
Tukey, J. W. (1977). *Exploratory Data Analysis.* Addison-Wesley, Reading, MA.
Ullman, S. (1984). Visual routines. *Cognition* **18**, 97–159.
Yantis, S. (in press). On analog movements of visual attention. *Percept. Psychophys.*