VISUAL COGNITION, 2004, *11* (7), 801–822

Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities

Zenon W. Pylyshyn

Rutgers Center for Cognitive Science, Rutgers University, Piscataway, New Jersey, USA

The task of tracking a small number (about four or five) visual targets within a larger set of identical items, each of which moves randomly and independently, has been used extensively to study object-based attention. Analysis of this multiple object tracking (MOT) task shows that it logically entails solving the correspondence problem for each target over time, and thus that the individuality of each of the targets must be tracked. This suggests that when successfully tracking objects, observers must also keep track of them as unique individuals. Yet in the present studies we show that observers are poor at recalling the identity of successfully tracked objects (as specified by a unique identifier associated with each target, such as a number or starting location). Studies also show that the identity of targets tends to be lost when they come close together and that this tendency is greater between pairs of targets than between targets and nontargets. The significance of this finding in relation to the multiple object tracking paradigm is discussed.

The multiple object tracking (MOT) paradigm has provided a number of interesting and often counterintuitive findings and has become one of our principle paradigms for studying object-based visual attention, and the nature of the connection between percepts and individual objects in a scene (Blaser & Pylyshyn, 1999; Blaser, Pylyshyn, & Domini, 1999; Blaser, Pylyshyn, & Holcombe, 2000; Pylyshyn, 1989, 1994, 1998; Pylyshyn, Burkell, Fisher, Sears, Schmidt, & Trick, 1994; Pylyshyn & Storm, 1988; Scholl & Pylyshyn, 1999; Scholl, Pylyshyn, & Feldman, 2001; Scholl, Pylyshyn, & Franconeri, 2004; Sears & Pylyshyn, 2000). This methodology has also been used by a number of laboratories (Bahrami, 2003; Cavanagh, 1999; Culham, Brandt, Cavanagh,

Please address all correspondence to Center for Cognitive Science, Rutgers The State University of New Jersey, Psychology Building Addition, A115, Busch Campus, 152 Frelinghuysen Road, Piscataway, NJ 08854-8020, USA. Email: zenon@ruccs.rutgers.edu.

I wish to thank Amir Amirrezvani, John Dennis, and Ashley Black for their help in organizing and carrying out the experiments reported here, and especially Carly Leonard who contributed not only to the data analysis but also to the design of several of these experiments. This research was supported by grant 1R01-MH60924 from the National Institutes of Health.

http://www.tandf.co.uk/journals/pp/13506285.html DOI: 10.1080/13506280344000518

Kanwisher, Dale, & Tootell, 1998; He, Cavanagh, & Intriligator, 1997; Intriligator & Cavanagh, 2001; Viswanathan & Mingolla, 2002; Yantis, 1992) to study aspects of visual attention. In this experimental paradigm, observers track about four or five objects that move randomly among a larger set of identical, independently moving objects. Perhaps the most important and surprising finding (often even to subjects themselves) is the fact that observers are able easily to track four or even five such objects, using only the individual objects' continuing identity and the fact that they had been designated as targets at the start of each trial. They can do this when the movement parameters (speed, distances) precludes their doing so by scanning their attention sequentially to each item and updating its location as encoded in memory (as shown by Pylyshyn & Storm, 1988). Despite the simplicity of the paradigm there are a number of puzzles raised by this methodology that could usefully be clarified, both by a more formal analysis of the task and by further empirical explorations. The purpose of this paper is to contribute to this clarification.

While there are many variants of the MOT task, a typical experiment is illustrated in Figure 1. A number of simple objects (typically about eight circles or squares) are displayed on a screen. A subset of these elements (typically about four) are briefly made visually distinct, often by flashing them a few times. Then all the objects move randomly and independently. Sometimes the motion of the objects is constrained so they do not collide, but in recent work they more often travel independently and are allowed to occlude one another (in which case they may provide occlusion cues such as T-junctions). After some period of time (typically 5 or 10 s) the motion stops and observers are required to indicate which objects were the ''targets''. This is done either by flashing one of the objects and having the observer indicate by a key press whether that object was a target, or by requiring the observer to pick out all the targets using a mouse pointing device. The experiment (and its many variants) has repeatedly shown that observers can track at least

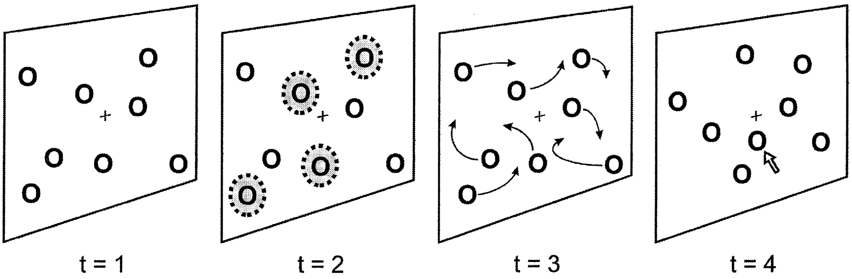

Figure 1. The sequence of events in a typical MOT experiment, in which the observer indicates whether the item flashed at the end of the trial was one that was being tracked (shaded circles indicate items being flashed).

four items in a field of eight identical items over a period of 10 s with an accuracy of 85–90%.

This result raises the question: How is it possible for these items to be tracked, given that they move independently and that no fixed visual properties distinguish the target subset from the nontarget subset? In the original study, where this paradigm was first introduced (Pylyshyn & Storm, 1988), we noted that at any instant the only distinct property that distinguishes targets from nontargets was their (changing) locations, but that it was unlikely that this property could have served as the basis for tracking in that study. Given certain assumptions, we showed that the tracking performance that might be expected (which we obtained by simulating a location-updating algorithm on the actual displays used in that experiment) would not exceed 30%, even with various additional favourable variants such as recording the object's velocity vector along with its location, or the use of a more sophisticated guessing strategy (e.g., based on extrapolation of its recorded motion). This was far below the 87% tracking performance observed in that experiment. Thus we concluded that a location-updating process could not have been responsible for the observed tracking performance.[1]

The theoretical framework that motivated the work on MOT is called visual indexing or FINST theory (which had been developed years before in an entirely different context—Pylyshyn, Elcock, Marmor, & Sander, 1978). Empirical predictions derived from this theory have also been tested using a variety of different methods, including subitizing (Trick & Pylyshyn, 1993, 1994a, 1994b), subset search (Burkell & Pylyshyn, 1997), and illusory line-motion (Schmidt, Fisher, & Pylyshyn, 1998). The theory claims that indexes pick out and track objects as token individuals, and not as elements that possess certain specific properties (since an indexed individual will keep being tracked even if its properties change). Another way to put this is that the indexing mechanism defines the equivalence relation ‘‘perceived as the same persisting individual’’, and thus functions as a form of preconceptual or ‘‘demonstrative’’ reference (Pylyshyn, 2000, 2001a, 2001b). It is not the purpose of the present paper to

[1] This conclusion was based on the following assumptions: (1) encoding the location of targets requires focal attention, and (2) focal attention is unitary and has to be moved smoothly from one target object to another at a finite speed, and (3) the locations have to be updated based on locating the nearest object at the specified location on each scanned cycle. Of course any of these assumptions could be questioned. For example, location encoding could conceivably be done in parallel—although the evidence from search experiments is that encoding location, like the encoding of other localized features, requires focal attention (an assumption that is explicit in the account of conjunction-search results within the feature integration model of Treisman & Gelade, 1980). Also the assumption that only one location can be encoded at once has recently been confirmed by (Hess, Barnes, Dumoulin, & Dakin, 2003). Nonetheless, it must be recognized that the argument in (Pylyshyn & Storm, 1988) only rules out one particularly plausible proposal that uses serially updated locations for tracking.

discuss this particular interpretation of visual indexing. The purpose, rather, is to examine some of the logical requisites of MOT, and to report some experiments designed to cast further light on the nature of the tracking process.

THE LOGIC OF MULTIPLE OBJECT TRACKING

The critical aspect of the MOT task is that targets are visually distinct *only* at the start of the trial, after which nothing distinguishes a target from a nontarget except its historical provenance, which is traced back to the start of the trial. Thus *targethood* is defined by historical continuity: A particular object is a target if *that very token object* was identified as a target at the start of the trial. Consequently, in order to track a particular target, its individual identity has to be tracked through the entire trial. We refer to this requirement as the Discrete Reference Principle (DRP). The following discussion of DRP is presented to clarify some misunderstandings that occur frequently.

The critical aspects of this task are: (1) The objects to be tracked are visually distinct *only* at the start of the trial, (2) what is being tracked is the individuality of the objects—the fact that they are *the same object*. Consequently the following defines *being a target* in the MOT task: A particular object X_n is a target if, and only if, it is the *same individual object* as a particular object that had initially been a designated target. More formally,

For any individual object X_n at time t, $X_n(t)$ is a target if, and only if,
(1) $X_n(t - \Delta t)$ is a target, where Δt may approach zero in the limit, and
(2) $X_n(0)$ is *visibly* a target.

The definition is put in this form, as a recursion over time, because this form reflects how a viewer must determine whether some particular object $X_n(0)$ is a target. In applying this definition a viewer must be able to determine that clause (1) holds which, in turn, requires that the token identity (or, as it is sometimes called ‘‘numerical identity’’) of individuals $X_n(t)$ and $X_n(t - \Delta t)$ must be determined. Thus a critical part of determining whether some object is a target is being able to trace its individuality (or individual identity) back over time to the start of each trial and thus ascertaining that it was one of the objects that had been visibly distinguished as a target at that time. Another way to put this is that in order to track a particular object, the index n of that object $X_n(t)$, must be determined. This equivalent way of putting the logical requirement of tracking is called the discrete reference principle (DRP). This principle says that in order to track a set of objects, (1) each individual object in that set must be kept distinct from every other object in the display and (2) each individual target object must be identified with a particular individual target object in the immediately preceding instant in time. The problem of meeting the second of these requirements is sometimes referred to as the correspondence problem. A solution to this problem (for each of n individual objects) establishes the distinction between each object and every other object, as well as the equivalence of each particular object token at time t with a particular object token at time $t - \Delta t$. This, In turn, is equivalent to assigning a distinct reference or index to each target that is successfully tracked.

Under special circumstances it might be possible to determine that a particular individual object had been a member of the target set at some previous time (other than at the beginning of the trial) without the benefit of DRP. For example, among the possible special circumstances in which we could compute set membership without DRP would be cases in which the set of targets was distinguishable as a group in some way, say by its colour, its location, its spatial pattern, or its form of movement. In such cases, an object might be classed correctly as a target by virtue of its recognizable group membership, rather than by virtue of having tracked the individual object. But in general, the only way to determine that a particular individual object belonged to the target set in the previous instant is by knowing *which particular individual* in the target set it had been.

It has sometimes been suggested that an object's membership in a target set might be determined without tracing the object's individuality, simply by treating the target set as a whole—for example, by labelling the items in the target set with the same label (say, T for Target) and then checking for objects with the label T at the end of the trial. But this is a misleading reification of the notion of labelling. Since the label is not physically attached to the object or its representation, it will only move with the object if the object is tracked to ensure that the label continues to be associated with (i.e., ''attached to'') the same object. In general you cannot determine whether some perceived object has a particular label assigned to it without first determining which object it is—i.e., by tracking it. The same is true of any proposal which suggests that the set of targets might be treated in a unitary way, say as the vertices of a deforming polygon (Yantis, 1992). Yantis showed that instructing observers to use this ''polygon strategy'' improved their tracking performance, and that restricting the motion of objects so that the polygon never ''collapsed'' into a concave polygon (by constraining vertices from travelling through one of the sides of the imaginary polygon), improved performance. The Yantis results do indeed show that thinking of the targets as vertices of a polygon helps tracking performance. However using the ''polygon'' strategy does not change the logical requirements of the task, unless it allows some redundancy in the motion of the objects to be exploited (as may be the case when the motion is restricted so the covering polygon remains convex). So long as each target moves independently of the other targets there is no redundancy to exploit. When objects move independently, the polygon strategy does not offer an alternative explanation of tracking (any more than thinking of the objects as birds in flight or ants on a beach) since each target object still has to be tracked independently in order to determine the moment-to-moment location of the vertices of the polygon. Unless the set as a whole has a perceptible distinguishing property, an object's membership in the set can only be determined by tracing the object's history back to the start of the trial when it was visually distinct.

THE DISCRETE REFERENCE PRINCIPLE AND TRACKING THE IDENTITY OF OBJECTS

If, as we claim, a distinct internal reference token (say αi) is assigned to each successfully tracked target, it should then be possible to associate a given overt label with each target. All that an observer needs to do in order to correctly identify each target is to learn a list of pairs $<\alpha i, \beta i>$, where βi are external labels or correct overt responses. Thus, if a target is initially provided with a unique overt label by the experimenter, and if that target is successfully tracked, an observer should be able to report that overt label—simply as a consequence of having tracked it, together with having memorized a short list of paired associates linking discrete internal references with overt labels. This leads to the prediction that tracked objects should be identifiable by overt labels assigned to them during the initial target-identification stage, so long as the paired associates can be recalled under those conditions. If there are only four targets, this requires that a list of only four paired-associates ($\alpha 1$–$\beta 1$, $\alpha 2$–$\beta 2$, $\alpha 3$–$\beta 3$, $\alpha 4$–$\beta 4$) be recalled in order that the identification of each tracked object is correctly reported. We shall provide evidence that such a list of pairs is easily recalled under conditions of the present experiments.

EXPERIMENT 1

Method

Materials. The first experiment was a typical MOT tracking study except that observers were required to identify each target as well as to pick out the set of targets. The targets were given discrete identities at the start of each trial: Either a distinct name (one of the numbers 1, 2, 3, or 4) or a distinct starting location (one of the four corners of the screen). The experiment used four circular targets and four identical nontargets, each 47 pixels or 2.7° of visual angle. Each circle was surrounded by a 2 pixel (approx 0.12°) white border and the interior of the circle was blue. The screen background was black. Initial item positions were generated randomly, with the constraint that each had to be at least 5° from the edges of the display and at least 4° from each other. When two objects overlap, one of them (chosen at random) is always depicted as occluding the other. Because such objects provide T-junction occlusion cues, they may freely self-occlude without a significant decrement in performance (Viswanathan & Mingolla, 2002), and hence their trajectories can be computed entirely independently of one another, unlike some previous studies (e.g., Pylyshyn, 1988; Scholl & Pylyshyn, 1999) that used a repulsion or fence around each object to prevent collision.

The motion algorithm is the same as that used in other recent MOT experiments. Items were each assigned random horizontal and vertical velocity components varying between −2 and +2 units (representing the number of pixels that the object could move in each 17.1 ms frame). These could be incremented or decremented on each video frame by a single step, with a probability referred to as the ''inertia'' of the object motion. In the present experiment, this probability was set at .10, which keeps the objects from changing velocity too suddenly. Since the position of each item was determined independently, this results in independent and unpredictable trajectories (within the permitted range of the change). In the resulting motion, items could move a maximum of 0.12° vertically or horizontally per frame buffer. Since frame buffers were displayed for 17.1 ms each (corresponding to two screen scans of 8.55 ms for the iMac's 117 Hz monitor), the resulting item velocities were in the range from 0 to 7.02°/s, with an average velocity across all items and trials of 2.37°/s.

Design and procedure. Observers were instructed to make two responses on each trial after the objects stopped moving: First, they were to move the cursor to each object they believed was a target, and then to press the mouse button to indicate (or guess if they were unsure) their selection, and then immediately after selecting a target in this way, they were to use their nondominant hand to press one of four keys on the keyboard to indicate the identity of the each target they selected. After four such pairs of responses, the trial ended. The observer then pressed the spacebar to initiate the next trial. Observers were asked to keep looking at the fixation cross because that would make their tracking task easier. Eye movements were not monitored because different fixation strategies have been found not to affect performance on this task. (For example, Pylyshyn & Storm, 1988 monitored fixation and discarded trials on which observers made eye movements; Scholl & Pylyshyn, 1999 instructed observers to maintain fixation but did not monitor eye movements; while Intriligator & Cavanagh, 2001, and Yantis, 1992 employed no special constraints or instructions concerning fixation. Yet all these studies yielded qualitatively similar results.)

There were 240 trials in this experiment, organized into five blocks of 48 trials. The first two blocks were baseline conditions (see bellow), which were followed by three blocks constituting the main studies. After each block, observers were invited to take a short break. Trials were randomly assigned to a duration of 2 s, 5 s, or 10 s (with an equal number of each in every block). In the main experiment (blocks 3–5) each trial began with an initial 2 s target-identification phase during which the target items were flashed on and off and were also given an identifying label (called its ID) in one of two ways. In the name condition, the targets were identified with one of the numbers 1, 2, 3, or 4 displayed inside the circles. In the location condition, the four targets each initially appeared in a different corner of the screen (about 2 diameters, or 5° of visual angle, from each screen edge). These two conditions provided two different forms of overt labels or response IDs, which observers had to report at the end of each trial. Observers had to press numbered keys in the name condition, or one of the keys that formed a square arrangement on the numerical keypad (keys 7, 8, 4, 5, which were labelled with arrows on the keys) in the location condition. The name condition and the location condition were each run on a different group of observers and therefore that difference was analysed as a between-subjects factor. To obtain the ID score for each observer in each trial-duration condition we simply counted the number of ID responses that were correct in each trial (expressed as a percentage of the number of objects, which was always 4), and averaged over trials. The tracking score was similarly computed as the percent of targets that were correctly classified as targets.

In addition to the combined ID and tracking trials described above, observers also took part in two baseline conditions: a static ID-only memory-control task that did not involve tracking, and a track-only baseline task in which they tracked four targets but did not have to recall the ID of the objects. These two tasks were used in order to provide baseline performance measures for the two component skills involved in the experimental manipulations. They were presented first, thus providing observers with extra practice in both tasks and providing us with conservative baselines. In the ID-only task, observers simply viewed a display consisting of the initial 2 s of a tracking trial—where the target circles were identified as targets and also given the labels 1, 2, 3, or 4 (there was no static condition corresponding to the location labels since in the static case this ID would be available trivially at the end of each trial from the layout of the objects). Once the 2 s static presentation was completed, the trial continued for one of the designated durations (2, 5, or 10 s) with all objects remaining at fixed positions on the screen (but without their numerical identifiers). The track-only task was identical to the main ID and track task used in each of the two experiments except observers did not have to indicate the identity of each target.

Two different groups of 10 observers were run in the two experimental conditions. Each condition involved 240 trials and took approximately 1 hour. One observer in the name condition was lost due to equipment failure; consequently, there were only 9 observers in the name condition and 10 observers in the location condition.

Results

The results of this experiment are shown in Figures 2–4.

Comparison of ID and tracking performance. Recall that the ID type (names vs. locations) was a between-subjects factor in this design. Consequently a mixed within-subjects and between-subjects analysis of variance was carried out and it revealed that the overall effect of task type (tracking vs. ID) was statistically significant, $F = 202.0$; $df = 1,17$; $p < .001$, the effect of trial duration was significant, $F = 130.6$; $df = 2,17$; $p < .001$, and the interaction of these two measures was also significant, $F = 27.9$; $df = 2,17$; $p < .001$, but the effect of ID type (names vs. locations) was not statistically significant, $F = 0.18$; $df = 1,17$; $p > .9$. As can be seen from Figure 2, tracking performance decreased with trial duration, and ID performance deteriorated even more rapidly as trial duration increased, reaching a value of less than 30% after 10 s of tracking. (Notice that in all experiments involving location IDs, the score for the 2 s trials is higher for location ID than it is for the name ID. This is due to the fact that whereas the name displayed inside the circular targets disappears after the 2 s inspection time, the object remains relatively close to its starting location for some time after it begins to move. After 2 s, it has moved an average of only 1 diameter from where it began, so the location ID is relatively easy to guess by inspection.)

Baseline performance. The design of Experiment 1 included a ‘‘static ID-only’’ condition (in which the objects to be identified remained in fixed positions for the duration of the trial) and also a ‘‘track-only’’ condition in

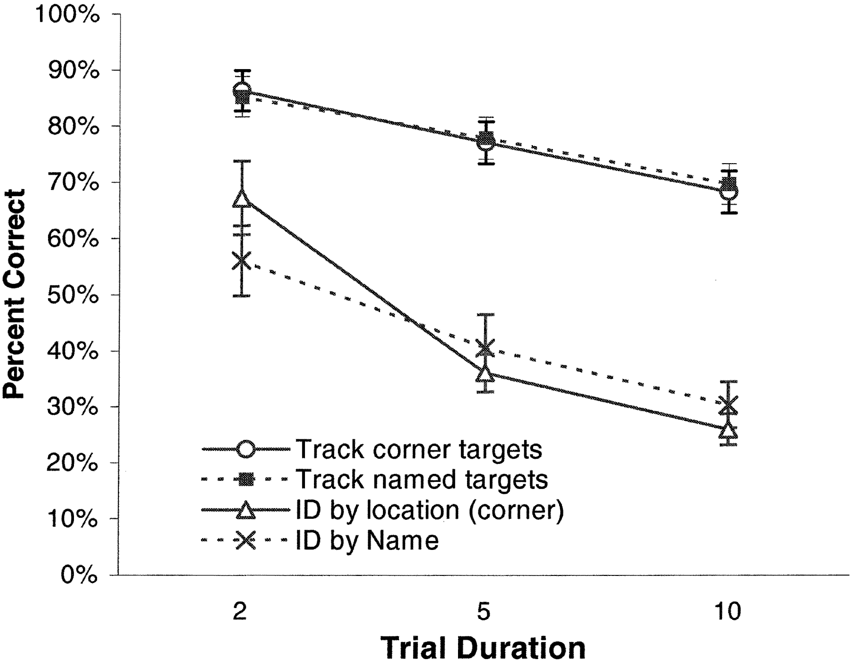

Figure 2. Performance in tracking and in recalling the identity labels assigned to targets in MOT as a function of trial duration. ID performance is poorer and decays more rapidly than tracking performance.

which observers tracked the targets but did not have to identify them. The static ID-only and the tracking-only conditions used the identical displays and timings as were used in the combined tacking-ID conditions. (Note that in the baseline ID case only the name ID was used since the location ID is available trivially by just inspecting the display, hence only the nine observers who were in the names condition provided data for the static ID analysis.) These two conditions allowed us to measure how well each could be done without the simultaneous requirement of the other.

Figure 3 shows baseline performance in the tracking task and in the ID tasks (both based on the names condition) when these are carried out separately (the performance when both tracking and ID are done together is also reproduced from the names condition of Figure 2 for convenience in comparing these with the baselines conditions). These data were analysed using a within-subjects ANOVA which showed a significant effect of task, $F = 27.1$; $df = 1,8$; $p < .001$, of trial duration, $F = 29.9$; $df = 2,8$; $p < .001$, as well as a significant Task × Duration interaction, $F = 19.3$; $df = 2,16$; $p < .001$. As can be seen from Figure 3, the ability to report ID names is nearly perfect (remaining at over 93%) and is

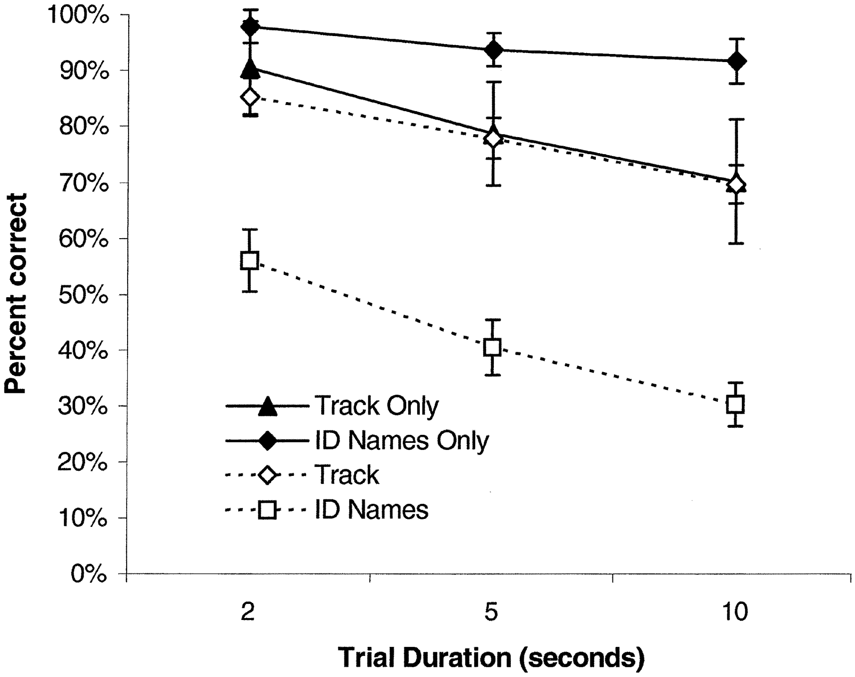

Figure 3. Baseline performance for tracking and for recall of ID when carried out independently. For comparison, dotted lines show the performance when both are done together, reproducing the names condition of Figure 2.

significantly *higher* than the tracking performance. Thus, it does not appear to be the case that tracking is generally easier than reporting the names of objects when the latter are recalled after 10 s.

ID performance on trials in which all targets were correctly tracked. In considering possible reasons for the difference between tracking and ID scores, it might be noted that tracking performance places an upper bound on ID performance in all conditions: It is not possible to identify objects that one has not tracked. Consequently the poorer ID performance and the more rapid decline in that performance with trial length could be an artifact of the dependence of these two measures. To control for this possibility, we analysed the ID performance on only those trials on which all the objects had been successfully tracked. This reduced the number of data points available for analysis from a maximum of 48 trials per subject per condition to a mean per subject of 28.12, 19.75, and 11.00 for the name condition and 30.2, 18.4, and 11.2 for the location condition for trial durations of 2 s, 5 s, and 10 s respectively. In addition, two observers had to be discarded because they did not have two or more trials with perfect tracking for durations of 5 or 10 s. As a result, the power of the test is reduced. Nonetheless, the results are clear. The resulting ID performance, shown in Figure 4, exhibits the same significant drop with increasing trial length as was observed with the overall mean ID score (shown in Figure 2). The effect of trial duration (a within-subjects effect) was again significant, $F = 37.6$; $df = 2,15$; $p < .001$. The difference between the two forms of ID response—the location and name conditions (a between-subjects effect)—was not significant, $F = 0.25$; $df = 1,15$; $p > .62$, nor was the interaction of trial duration and the two forms of ID response, $F = 2.6$; $df = 2,15$; $p > .05$. Although ID performance on the perfectly tracked trials was higher that the overall ID performance (shown in Figure 2), the performance was still below 52% at the 5 s and 10 s trial durations, whereas the tracking performance on these trials was selected to be 100% (in fact on these selected trials, the ID performance at the 10 s duration was exactly the same as the mean ID for the unselected 10 s trials). Consequently the data on correctly tracked objects provides no reason to reject the conclusion that ID performance is poorer and falls off more rapidly than tracking performance with increasing trial duration.

EXPERIMENT 2

Another possible reason why ID performance is worse than tracking performance, and becomes increasingly so as the length of the trial increases, is that the tracking task itself may interfere with recall of the paired associates consisting of internal references and external labels, or $<\alpha i, \beta i>$. The ID baseline condition of Experiment 1 only measured the recall of IDs when no tracking task intervened between presentation and recall of objects and IDs. Experiment 2 was

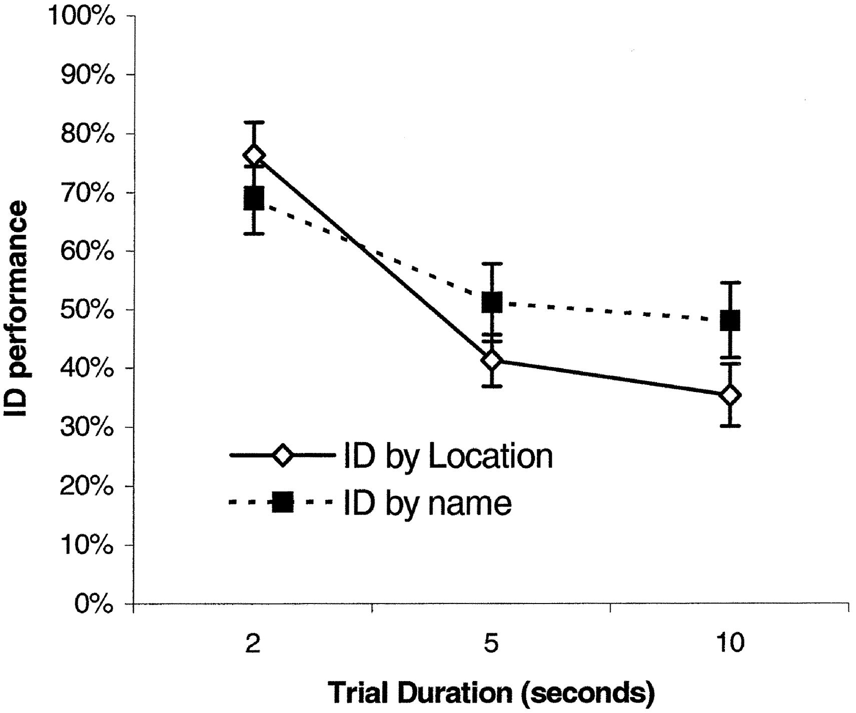

Figure 4. Mean ID performance on only those trials in which all targets were tracked correctly.

designed to examine the possibility that tracking actually interferes with ID recall by measuring ID and tracking performance under conditions in which the two tasks could interfere with one another, but in which the ID task was independent of the specific items being tracked. Experiment 2 measured recall performance on the ID task when the intervening time was occupied by an *unrelated* tracking task, called the *embedded tracking task*, that involved tracking *different* items from those whose ID had to be recalled.

Method

Design. In Experiment 2 observers participated in a static ID task, in which the initially numbered objects did not move, as well as in an unrelated tracking task that occurred while the IDs of the static circles was being held in memory. In the ID task the target circles were presented with ID numbers displayed inside, as in the name ID condition of Experiment 1. After 2 s, the circles and numbers disappeared and a *different* set of eight circles appeared (without numbers), four of which flashed. The items in this embedded tracking task then began to move and observers had to track the subset of targets that had flashed, just as they did in Experiment 1. After 2, 5, or 10 s (randomly assigned to trials), the objects stopped moving and observers had to pick out the targets using a mouse. When they had completed this tracking task, the tracked objects disappeared and the original static display, this time without the numbered IDs in the circles, appeared on the screen. Observers then had to select each of these circles in turn (in any order) and to indicate which number had been in each (using the same keypad method as used in Experiment 1). A record was kept of each observer's performance both in the embedded tracking task and in recalling the IDs of the static target objects after a delay interval of 2, 5, or 10 s that was filled with the unrelated tracking task.

Procedure. The procedure was the same as in Experiment 1, except that observers had to recall the ID of one set of target objects and then to track an unrelated set of objects in the same way they had tracked them in other MOT experiments. As in Experiment 1, trials were randomly assigned to a duration of 2 s, 5 s or 10 s with an equal number of each in each of three 48-trial blocks. Each trial consisted of an ID task and an embedded (unrelated) tracking task as described above. The experiment lasted for about 1 hour. Eleven observers were paid for their participation in this experiment.

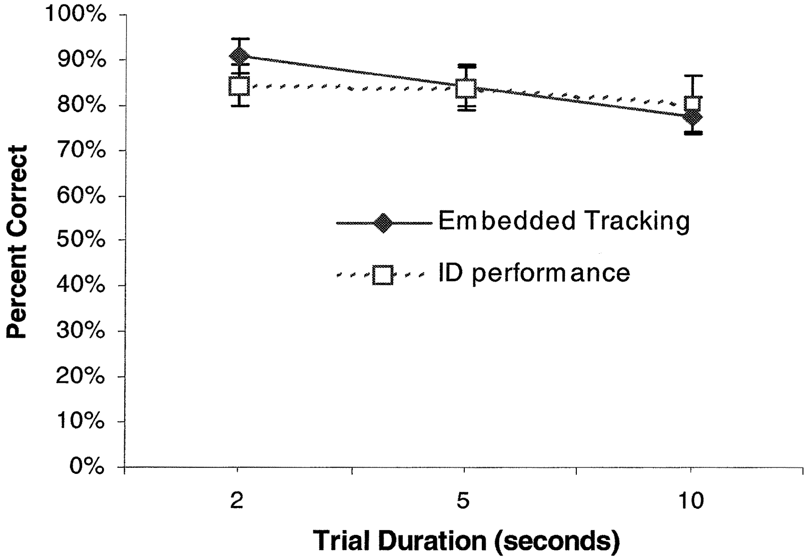

Figure 5. Performance in recalling ID names in a static display, when the recall is delayed by an interpolated tracking task identical to that which occurred in Experiment 1, except involving an unrelated set of items. The high performance on the ID task in this case shows that the poor recall of IDs in Experiment 1 was not due solely to the interfering effect of the tracking task.

Results

The results of Experiment 2 are shown in Figure 5. Both tracking and ID performance remained high (above 80%), thus providing no support for the hypothesis that the poor ID performance is due merely to the disruptive effect of the concurrent tracking task. There was no statistical difference between the ID and tracking score, although the tracking did decrease more rapidly with time than the ID score (the interaction was significant, $F = 14.7$; $df = 2{,}20$; $p < .001$, which is very different from the pattern found for ID performance when it is an integral part of tracking (illustrated in Figures 2–4).

EXPERIMENT 3

The data so far suggest that the higher frequency of ID errors does not arise from interference due to the tracking task itself—i.e., to the necessity of maintaining the distinction between targets and nontargets. As a further test of this hypothesis we examined whether making the tracking task easier would affect the pattern of ID errors. One way in which the problem of distinguishing targets and nontargets can be reduced (or even or eliminated) is by making targets and nontargets easy to distinguish or, in the extreme, by eliminating nontargets entirely. If ID errors are not due to the problem of maintaining a distinction between targets and nontargets, but rather to the task of tracking the targets themselves, one might expect that ID errors would persist even with this reduced tracking task.

Method and procedure

Experiment 3 was similar to Experiment 1, except that the location ID condition was omitted throughout and two additional conditions were added. In one condition, called the easy track condition, the nontargets were exactly the same as in the previous experiments except that the inside of the nontarget circles were a different colour (targets were blue and nontargets were green). In the other condition nontargets were eliminated and observers merely had to keep track of the ID of four moving objects (this involved the identical displays that were used in the earlier experiments except that the nontargets were rendered invisible). Trials were arranged into six blocks, two each for the normal tracking (MOT), easy track (EZT), and targets only (TO) conditions. Blocks were alternated between these three conditions (in the order: MOT, TO, EZT, MOT, TO, EZT) and trial duration was randomized as before. Nine undergraduate volunteers served as subjects in exchange for course credit.

Results

Figure 6 shows the results of Experiment 3. Even though the ID scores were somewhat higher than in the regular ID-while-tracking condition of Experiment 1, they were nonetheless significantly lower than tracking scores and dropped to

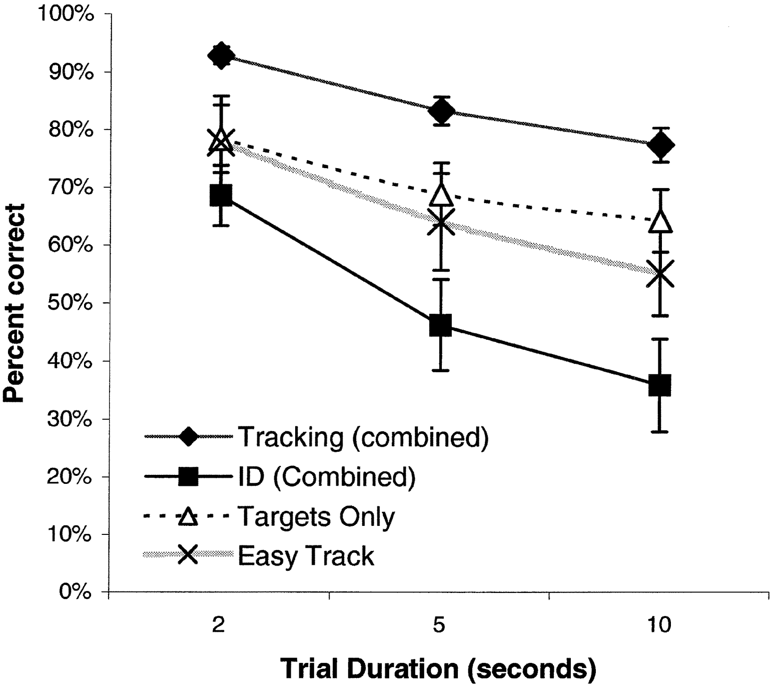

Figure 6. ID performance while tracking normal MOT displays (labelled as ''combined'' and shown along with tracking performance), displays in which the nontargets are easily distinguished from targets by their colour, and displays that contained no nontargets (these displays consisted of only four targets that initially had numbers in them). The solid lines replicate part of Experiment 1.

64% for the targets-only condition and 55% for the easy track condition, as compared to 35% for the combined ID-while-tracking condition. A within-subjects analysis of variance of the ID scores confirmed the statistical reliability of the effects of duration, $F = 42.2$; $df = 2{,}16$; $p < .000$, and type of ID, $F = 15.3$; $df = 2{,}16$; $p < .000$. No other effect was statistically reliable.

DISCUSSION OF RESULTS SO FAR AND MOTIVATION FOR EXPERIMENT 4

These experiments leave us with the following puzzle: Why does performance in maintaining the identity of individual tracked objects not match the corresponding performance in tracking the objects. Earlier we claimed that correct tracking assumes the Discrete Reference Principle and this principle, together with the ability to recall the correspondence between given labels and internal references, entails perfect recall of the ID of successfully tracked objects. The results of Experiments 1–3 are not consistent with this prediction. Experiment 2 showed that the discrepancy between tracking and ID performance was not attributable to interference between the tracking task and the ID recall task, since such interference was not observed when the tasks were independent, and Experiment 3 showed that ID errors persist even when there were no nontargets.

One possible reason for these results might lie in the nature of the errors that observers make, and in the consequence that different types of errors have on the two measures (ID and tracking). With certain kinds of errors it is possible to violate the discrete reference principle (i.e., to fail to maintain the individuality of targets) and still correctly assign individuals to the target–nontarget category (i.e., still appear correctly to track the targets). For example confusing one individual object with another object represents a failure to correctly track that object, yet this failure does not show up as an error in tracking performance if the objects involved in this exchange are both targets: If you switch target 1 with target 2 and still classify both as targets, tracking performance is not diminished. In contrast, switching the identity of a target with that of a nontarget does show up as a tracking error. So it becomes relevant to ask what kinds of errors observers tend to make, and in particular to determine whether there are circumstances under which they tend to make target–target confusions more frequently than target–nontarget confusions. To examine this question, we repeated Experiment 1, using software that enabled us to record the complete trajectory of each object, as well as the actual ID response that observers made to each object they classed as a target. We could thus measure the tendency to swap the identity of targets with other targets and compare this with the tendency to swap targets with nontargets under comparable conditions.

EXPERIMENT 4

The purpose of this experiment was to examine whether there is a greater tendency to swap target–target (TT) pairs than target–nontarget (TN) pairs and to determine whether this tendency is associated with how close objects came to one another during a trial. The measure of ID performance used in Experiments 1 and 2 was the proportion of targets that were given the correct ID on each trial. This score does not translate directly into the number of swapped IDs for a number of reasons. For example, two ID errors may or may not represent two TT swaps since they could be due to a combination of two TN swaps. Similarly, three ID errors could arise from any one of a number of different TN and TT swap combinations. Because of this the following changes were made in scoring the outcomes of Experiment 4. By keeping track of which ID responses were given to which objects we could determined the actual pairwise swaps, defined in the case of TT swaps as the assignment of IDs to two targets in such a way that target X was given target Y's correct ID and vice versa, or in the case of TN swaps as the assignment of target X's ID to a nontarget and the loss (or null ID assignment) of an ID for target X, which was treated as a nontarget.

Scoring

In order to compare the tendency to make TT swaps with the tendency to make TN swaps we used two modified scoring procedures. (1) In order to make the analysis less dependent on how swaps were counted (especially in cases of three swaps that may have involved intermediate objects) we classified and counted individual *trials* that had swaps of several different kinds, expressed as a percent of the total trials for each observer. Trials were categorized as TT swaps if they had one or more pairs of targets whose IDs were exchanged, and as TN swaps if they had one or more pairs of target–nontargets whose IDs were swapped, as defined above. (2) In addition to counting trials we also counted the number of pairs of objects whose IDs were swapped (by each observer), using a scoring scheme that allows us to examine the frequency of swaps among objects at different minimum distances apart.

For the second measure, all the pairs of objects that had a target as one member of the pair were first scored in terms of the closest distance they had come to one another in that trial. Then the pairs were marked as having been correctly tracked or as having been involved in one of the two kinds of swaps (TT or TN). We used the minimum distance between pairs because there is reason to think that this is an important factor in determining tracking errors and very likely for ID errors as well. For example, He et al. (1997) and Intriligator and Cavanagh (2001) showed that when MOT is carried out from a distance that places objects within the limits of what they call ''attentional resolution'' tracking is impaired. We scored all pairs of objects according to the minimum distance between them in each trial. We then selected values of interpair distances that would partition the pairs of objects into segments with roughly equal numbers of pairs. We found that by dividing the distances between objects into the range from 0 to 36 pixels (roughly 2.16° of visual angle or 75% of an object's diameter), from 37 pixels to 76 pixels (roughly 2.22–4.56° of visual angle), and from 77 pixels to the maximum video buffer size (roughly 4.62–28.8° of visual angle), we partitioned the TT pairs into sets of 2999, 2468, and 2633 pairs and the TN pairs into sets of 7912, 6702, and 6986 pairs, respectively. Notice that there were about 2.6 times as many TN pairs as TT pairs, which is close to the expected ratio of 2.5, based on the fact that there are six distinguishable target–target pairs that could be swapped ($^4C_2$) and 15 target–nontarget pairs that could be swapped in each trial.[2] Because observers

[2] Notice that although there are actually 24 (4!) target–nontarget pairs, they are not all distinguishable because nontargets were not identified by an ID. Consequently swapping a target T with a nontarget X is indistinguishable from swapping target T with any other nontarget Y ($Y \neq X$). Thus in determining how many distinguishable ways there are of swapping targets and nontargets we have only to consider how many ways there are of choosing target candidates to swap. There are $^4C_1$ or four ways in which one target could be swapped with a nontarget, $^4C_2$ or six ways for choosing two targets to be swapped, $^4C_3$ or four ways for choosing three targets to be swapped and one way for all four targets to be swapped with nontargets, giving a total of 15 distinguishable ways of swapping targets with nontargets.

were required to provide four IDs (even if they had to guess), a target that takes part in a TN swap automatically incurs a TT swap, since a TN swap means that there is a lost ID that has to be replaced by another ID. For this reason only the number of TT swaps in excess of the TN swaps on any trial were scored as TT swaps for purposes of this second analysis.

Materials and method

Experiment 3 was essentially a replication of the name condition of the ID and track task of Experiment 1, using software that enabled us to measure the distance between all pairs of objects. The software also recorded the actual identities of objects selected by observers. Because we were particularly interested in ID swaps, we used only 5 s and 10 s trials in order to increase the number of useable data points. Fifteen student volunteers, drawn from the Rutgers psychology subject pool, were tested.

Results

An initial analysis of the results showed that there was no significant difference in the pattern of ID scores between the 5 s and 10 s trials in this experiment. In order to increase the number of useable data points (i.e., swapped pairs) we combined the 5 s and 10 s trials. Results in terms of the first measure, the proportion of trials with TT and TN swaps (as well as both types of swaps and no swaps), are shown in Figure 7. Analysis of these data showed a main effect of swap type, measured in terms of the proportion of trials containing different each of the four swap types, $F = 25.1$; $df = 3{,}45$; $p < .001$. A paired comparison of effects (with Bonferroni correction for multiple comparisons) also confirmed that the ID of a target was much more likely to be swapped with that of another target than with that of a nontarget. All pairs were reliably different from one another except for the comparison between number of trials with TN pairs and the number of trials with no swaps.

The second set of analyses involved scoring *pairs of objects* in which one member was a target (as described under ''Scoring'', above). The overall number of TT and TN swaps, expressed as a percentage of the total number of TT and TN pairs at each distance range, is shown in Figure 8 (the measure is expressed as a percentage of all pairs of each type at each distance range, and is very low because of the large number of possible pairs—e.g., for each observer there were 176 possible TN pairs at the shortest distance, of which 4.1 were swapped, and 67 possible TT pairs, of which 5.6 were swapped). The results show that (1) there are more swaps of either kind when the distance between objects is small, giving a significant distance effect, $F = 24.7$; $df = 2{,}14$; $p < .000$, (2) there are more TT swaps than TN swaps resulting in a significant swap type effect, $F = 38.6$; $df = 1{,}7$; $p < .000$, (3) there is an interaction between these two factors, $F = 7.4$; $df = 2{,}14$; $p < .01$, such that the difference between TT and TN swaps decreases the larger the

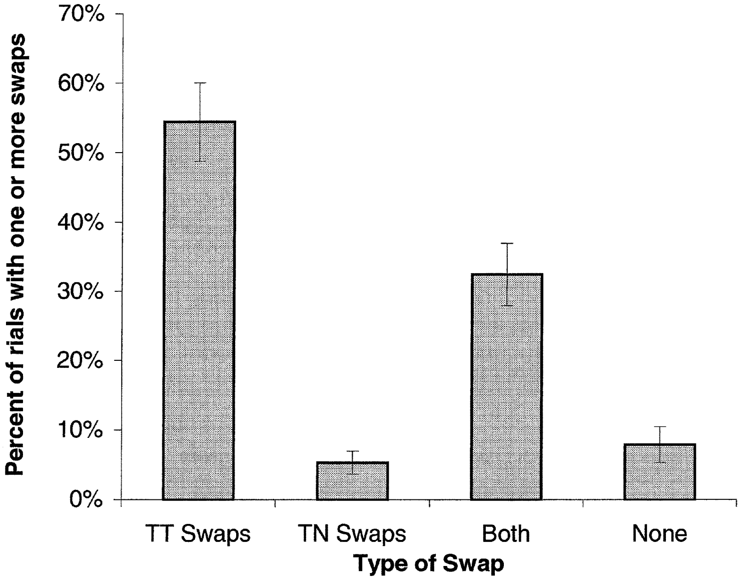

Figure 7. Chart showing the percentage of trials containing swaps of different kinds (including trials with both types of swaps and with no swaps). Only the difference between the proportion of trials with TN swaps and with no swaps failed to reach statistical significance.

distance between the pairs—or to put in another way, the number of swaps drops off with distance faster for TT pairs than for TN pairs. These results confirm our hypothesis that in MOT there is a much greater tendency for targets to be exchanged with other targets than with nontargets and that this tendency is exacerbated when objects come close together during a trial.

SUMMARY AND CONCLUSIONS

The studies reported here suggest that observers are better at tracking four independently moving identical objects than they are at keeping track of which one was which (i.e., than keeping track of their initially assigned names or their distinct starting locations). Although this appears to be inconsistent with the Discrete Reference Principle, or the need to keep track of each target as a distinct individual while tracking, it might be explained by the further hypothesis that errors are not randomly distributed among the objects being tracked, but rather target–target pairs are more readily confused (especially when they pass close to one another) than are target–nontarget pairs. Experiment 4 provides direct evidence for that hypothesis.

Although an asymmetry between target–target confusions and target–nontarget confusions arising from near-collisions may account for the divergence

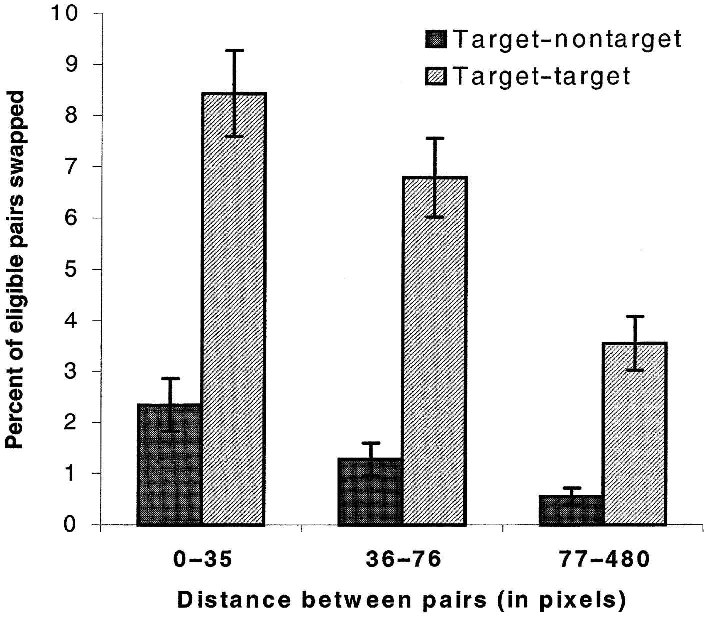

Figure 8. Chart showing that the proportion of both TT and TN swaps decreases as the distance between pairs increases (1 pixel corresponds to about 0.06° of visual angle). Although the difference between the tendency for TT swaps and TN swaps decreases with larger distances, it remains highly significant even for the largest distances.

between tracking and ID performance, it does not illuminate the question of what mechanism is responsible for this asymmetry. One possibility that we are currently investigating involves the notion of nontarget inhibition. Using a search task involving a split presentation of the search set Watson and Humphreys (1997) showed that when a subset of items is attentionally selected, the unselected items may actually be inhibited. If this were true in the MOT task, then we might expect that targets would be more often confused with other targets than with the inhibited nontargets because inhibition keeps the nontargets somehow out of reach of the imperfect tracking of targets. Of course the Watson and Humphreys tasks differ from the present ones in a number of critical ways. In particular, the inhibited items are always grouped, either temporally or by motion, whereas in the present studies the only common property that the nontargets have is that they are the items that are not being tracked. Despite this difference the possibility remains that nontarget objects are inhibited and that this, in turn, explains the asymmetry in the distribution of errors. This possibility is explored in a companion paper (Pylyshyn, 2004).

Other possibilities can also be considered. For example, we have assumed that the combination of having a unique internal name for tracked targets (as claimed by DRP), together with having a memorized set of pairs of internal names and external labels, ought to allow correct ID responses. But this also assumes that what we have called the internal name or discrete reference is available for use outside the tracking task, which may not be the case. It is possible that the internal name is available only for the purpose of tracking and is not reported outside that process. This would be like a local variable in a computer subroutine, which is not available to the program that calls the subroutine. Such encapsulation of information among processes is common in cognitive processes, especially in early vision (see, e.g., Pylyshyn, 1999). Whatever the ultimate answer to this question turns out to be, loss of ID in tracking does seem to be a robust phenomenon that needs to be clarified with further experiments and analyses.

REFERENCES

Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition*, *10*(8), 949–963.

Blaser, E., & Pylyshyn, Z. W. (1999). Measuring the independence of attention to multiple features [Abstract]. *Perception*, *28*, 56.

Blaser, E., Pylyshyn, Z. W., & Domini, F. (1999). Measuring attention during 3D multielement tracking [Abstract]. *Investigative Ophthalmology and Visual Science*, *40*(4), 552.

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature-space. *Nature*, *408*, 196–199.

Burkell, J., & Pylyshyn, Z. W. (1997). Searching through subsets: A test of the visual indexing hypothesis. *Spatial Vision*, *11*(2), 225–258.

Cavanagh, P. (1999). Attention: Exporting vision to the mind. In C. Taddei-Ferretti & C. Musio (Eds.), *Neuronal basis and psychological aspects of consciousness* (pp. 129–143). Singapore: World Scientific.

Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. H. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*, *80*(5), 2657–2670.

He, S., Cavanagh, P., & Intriligator, J. (1997). Attentional resolution. *Trends in Cognitive Sciences*, *1*(3), 115–121.

Hess, R. F., Barnes, G., Dumoulin, S. O., & Dakin, S. C. (2003). How many positions can we perceptually encode, one or many? *Vision Research*, *43*, 1575–1587.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of attention. *Cognitive Psychology*, *4*(3), 171–216.

Pylyshyn, Z. W. (Ed.). (1988). *Computational processes in human vision: An interdisciplinary perspective*. Stamford, CT: Ablex.

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

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

Pylyshyn, Z. W. (1998). Visual indexes in spatial vision and imagery. In R. D. Wright (Ed.), *Visual attention* (pp. 215–231). New York: Oxford University Press.

Pylyshyn, Z. W. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(3), 341–423.

Pylyshyn, Z. W. (2000). Situating vision in the world. *Trends in Cognitive Sciences*, *4*(5), 197–207.

Pylyshyn, Z. W. (2001a). Connecting vision and the world: Tracking the missing link. In J. Branquinho (Ed.), *The foundations of cognitive science* (pp. 183–195). Oxford, UK: Clarendon Press.

Pylyshyn, Z. W. (2001b). Visual indexes, preconceptual objects, and situated vision. *Cognition*, *80*(1/2), 127–158.

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking (MOT): II. Inhibition of moving nontargets. *Manuscript submitted for publication*.

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

Pylyshyn, Z. W., Elcock, E. W., Marmor, M., & Sander, P. (1978). *Explorations in visual-motor spaces*. Paper presented at the second international conference of the Canadian Society for Computational Studies of Intelligence, University of Toronto.

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

Schmidt, W. C., Fisher, B. D., & Pylyshyn, Z. W. (1998). Multiple-location access in vision: Evidence from illusory line motion. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 505–525.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*(2), 259–290.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object: Evidence from target-merging in multiple-object tracking. *Cognition*, *80*, 159–177.

Scholl, B. J., Pylyshyn, Z. W., & Franconeri, S. L. (2004). The relationship between property-encoding and object-based attention: Evidence from multiple-object tracking. *Manuscript submitted for publication*.

Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processes. *Canadian Journal of Experimental Psychology*, *54*(1), 1–14.

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

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

Trick, L. M., & Pylyshyn, Z. W. (1994a). Cuing and counting: Does the position of the attentional focus affect enumeration? *Visual Cognition*, *1*(1), 67–100.

Trick, L. M., & Pylyshyn, Z. W. (1994b). Why are small and large numbers enumerated differently? A limited capacity preattentive stage in vision. *Psychological Review*, *101*(1), 80–102.

Viswanathan, L., & Mingolla, E. (2002). Dynamics of attention in depth: Evidence from multi-element tracking. *Perception*, *31*(12), 1415–1437.

Watson, D. G., & Humphreys, G. W. (1997). Visual marking: Prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychological Review*, *104*(1), 90–122.

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

Original manuscript received September 2002
Revised manuscript received December 2003