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

Zenon W. Pylyshyn[1]
Rutgers Center for Cognitive Science

Abstract

The task of tracking a small number (about 4 or 5) 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 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.

Background: Multiple Object Tracking

The multiple object tracking (MOT) paradigm has provided a number of interesting and often counter-intuitive 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, submitted; Sears & Pylyshyn, 2000). This methodology has also been used by a number of laboratories (Bahrami, 2003; Cavanagh, 1999; Culham, Brandt, Cavanagh, 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 4 or 5 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 4 or even 5 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 8 circles or squares) are displayed on a screen. A subset of these elements (typically about 4) 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 seconds) 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 4 items in a field of 8 identical items over a period of 10 seconds with an accuracy of 85% - 90%.

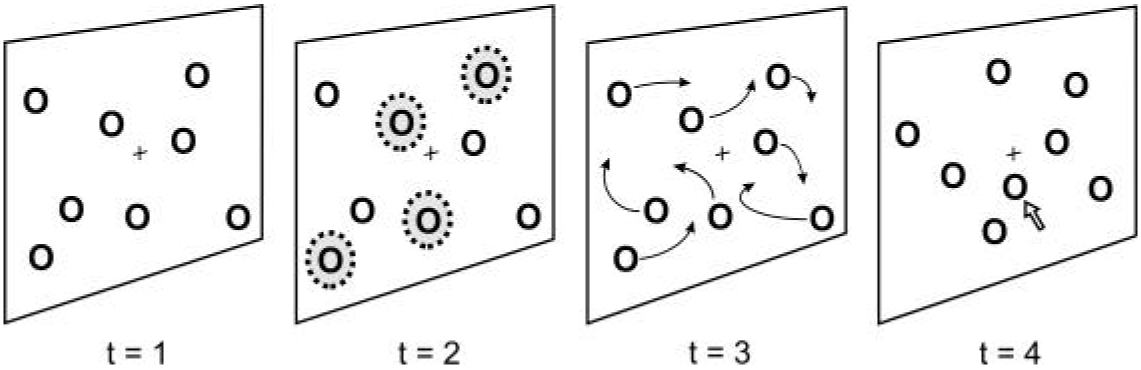

Figure 1. The sequence of events in a typical MOT experiment, in which the observer indicates whether the item flashed at the end of the trial was one that was being tracked (shaded circles indicate items being flashed).

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 favorable 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.[2]

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 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 (a) each individual object in that set must be kept distinct from every other object in the display and (b) 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-Δt. 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 color, 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 labeling 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 labeling. 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 traveling 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 (ID) of objects

If, as we claim, a distinct internal reference token (say a_i) 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 $<a_i, b_i>$, where b_i 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 4 targets, this requires that a list of only 4 paired-associates (a_1–b_1, a_2–b_2, a_3–b_3, a_4–b_4) be recalled in order that the identification of each tracked object be correctly reported. We shall provide evidence that such a list of pairs is easily recalled under conditions of the present experiments.

Experiment 1

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 4 circular targets and 4 identical nontargets, each 47 pixels or 2.7 degrees 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 degrees from the edges of the display and at least 4 degrees 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 0.10, 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 deg/s, with an average velocity across all items and trials of 2.37 deg/s.

Design and Procedure. Observers were instructed to make two responses on each trial after the objects stopped moving: (a) 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 (b) immediately after selecting a target in this way, they were to use their non-dominant hand to press one of four keys on the keyboard to indicate the identity of the each target they selected. After 4 such pairs of responses, the trial ended. The observer then pressed the space bar 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, 1997, 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 5 blocks of 48 trials. The first two blocks were baseline conditions (see bellow) which were followed by 3 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 two-second 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 4 targets each initially appeared in a different corner of the screen (about 2 diameters, or 5 degrees 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 which formed a square arrangement on the numerical keypad (keys 7, 8, 4, 5, which were labeled 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 analyzed 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 percent 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 4 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 seconds 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 two-second 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 & 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 one 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 Figure 2, Figure 3, and Figure 4.

(a) 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<0.001), 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<0.001), but the effect of ID type (*Names* vs *Locations*) was not statistically significant (F=0.18; df =1,17;p>0.9). 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 seconds 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. The is due to the fact that the 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 not moved far from where it began, so the *Location* ID is relatively easy to guess by inspection.)

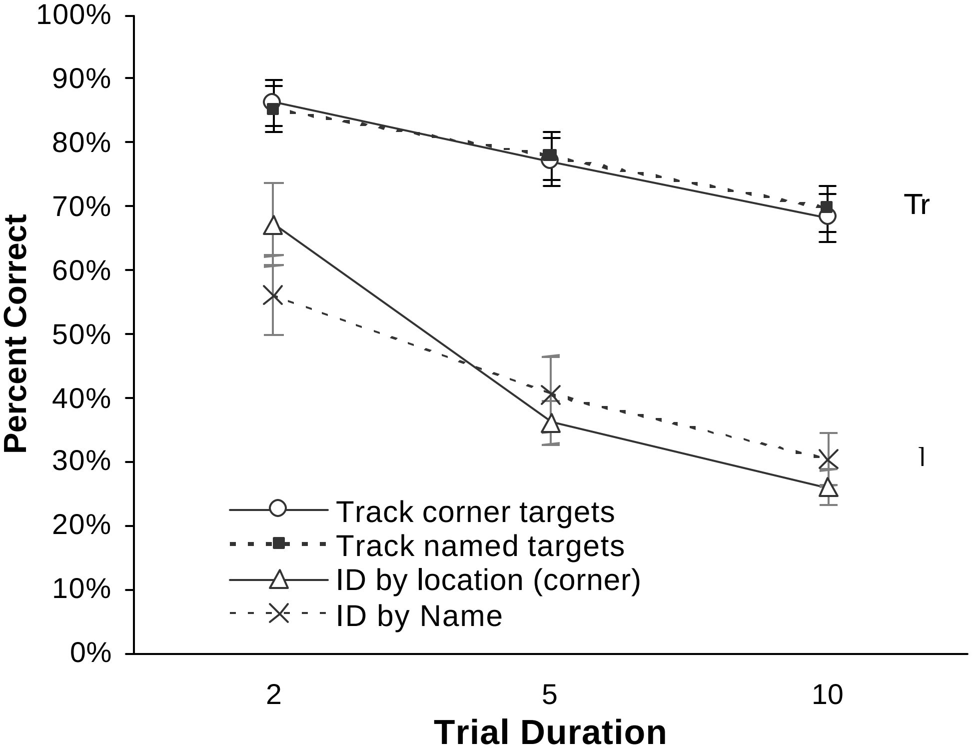

Figure 2. Performance in tracking and in recalling the identity labels assigned to targets in MOT as a function of trial duration. ID performance is poorer and decays more rapidly than tracking performance.

(b) 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 “tracking-only” condition in 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 9 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 analyzed 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 x duration interaction (F=19.3*; df*=2,16; $p < 0.001$). As can be seen from Figure 3, the ability to report ID *Names* is nearly perfect (remaining at over 93%) and is 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 seconds.

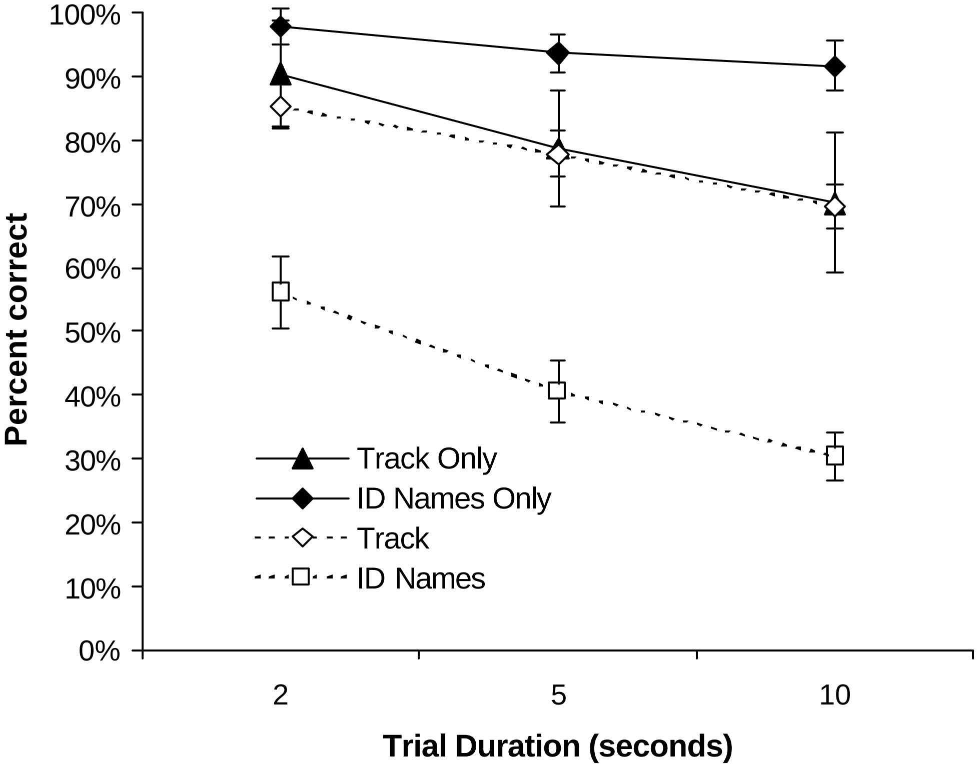

Figure 3 Baseline performance for tracking and for recall of ID when carried out independently. For comparison, dotted lines show the performance when both are done together, reproducing the *Names* condition of Figure 2.

(c) 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 analyzed 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, 2 observers had to be discarded because they did not have 2 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 second and 10 second 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.

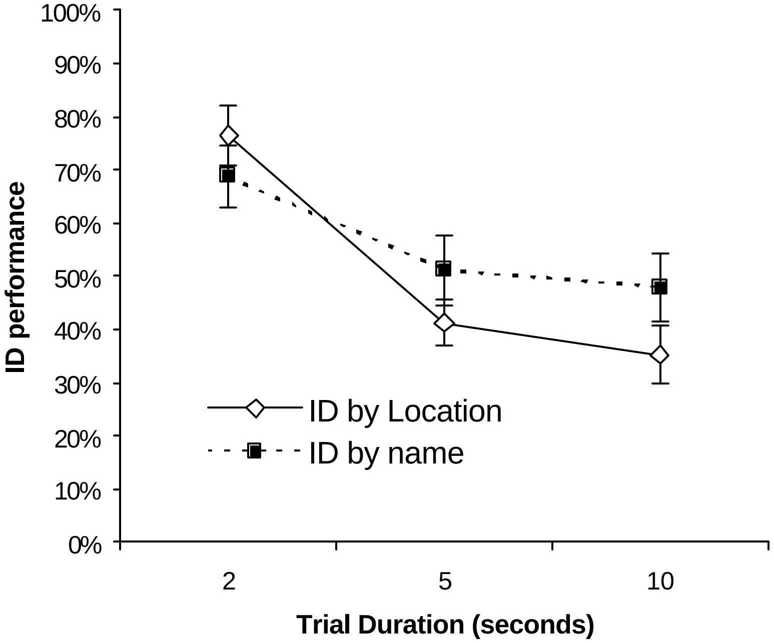

Figure 4. Mean ID performance on only those trials in which all targets were tracked correctly.

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 $<a_i, ß_i>$. 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 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.

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 seconds, the circles and numbers disappeared and a *different* set of 8 circles appeared (without numbers) and 4 of these were 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 seconds (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 seconds 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 a about one hour. Eleven observers were paid for their participation in this experiment.

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 Figure 2, Figure 3, and Figure 4).

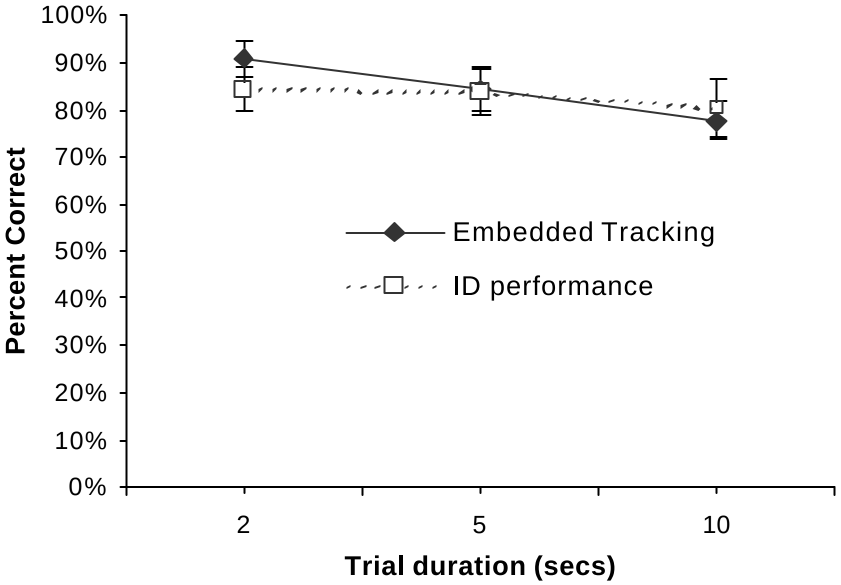

Figure 5. Performance in recalling ID *Names* in a static display, when the recall is delayed by an interpolated tracking task identical to that which occurred in experiment 1, except involving an unrelated set of items. The high performance on the ID task in this case shows that the poor recall of IDs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

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 color (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 6 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 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.

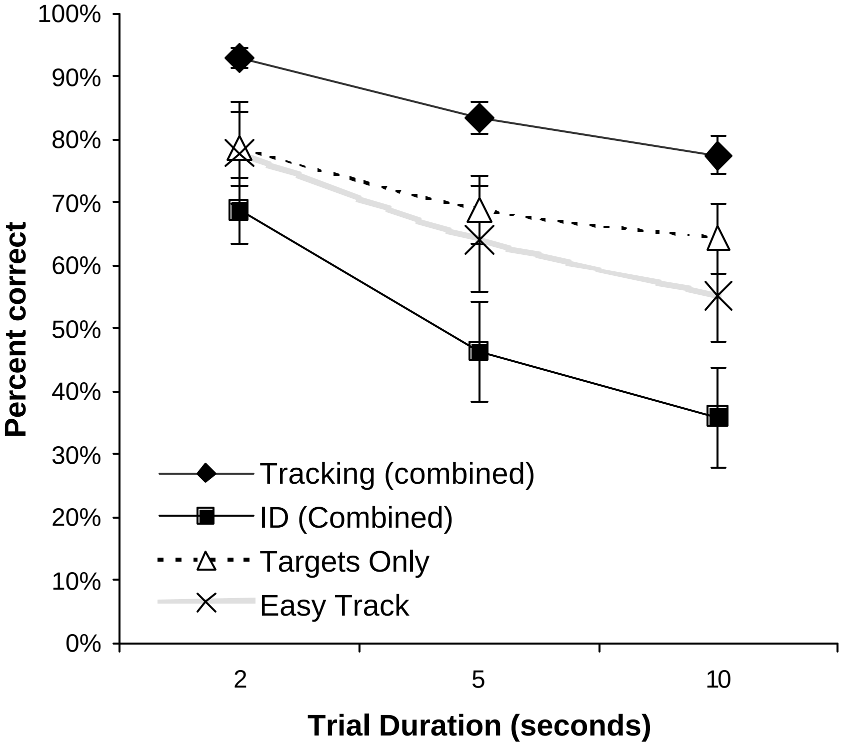

Figure 6. ID performance while tracking normal MOT displays (labeled as “combined” and shown along with tracking performance), displays in which the nontargets are easily distinguished from targets by their color, and displays that contained no nontargets (these displays consisted of only 4 targets that initially had numbers in them). The solid lines replicate part of Experiment 1.

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. Results of experiments 1 and 2 and 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. By 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, 2 ID errors may or may not represent 2 TT swaps since they could be due to a combination of 2 TN swaps. Similarly, 3 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, Cavanagh, & Intriligator, 1997; Intriligator & 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 inter-pair 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 degrees of visual angle or 75% of an object's diameter), from 37 pixels to 76 pixels (roughly 2.22 to 4.56 degrees of visual angle), and from 77 pixels to the maximum video buffer size (roughly 4.62 degrees to 28.8 degrees 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 6 distinguishable target-target pairs that could be swapped ($^4C_2$) and 15 target-nontarget pairs that could be swapped in each trial.[3] Because observers were required to provide 4 ID's (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 & 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 and 10 second 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 second and 10 second trials in this experiment. In order to increase the number of useable data points (i.e., swapped pairs) we combined the 5 second and 10 second 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 4 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.

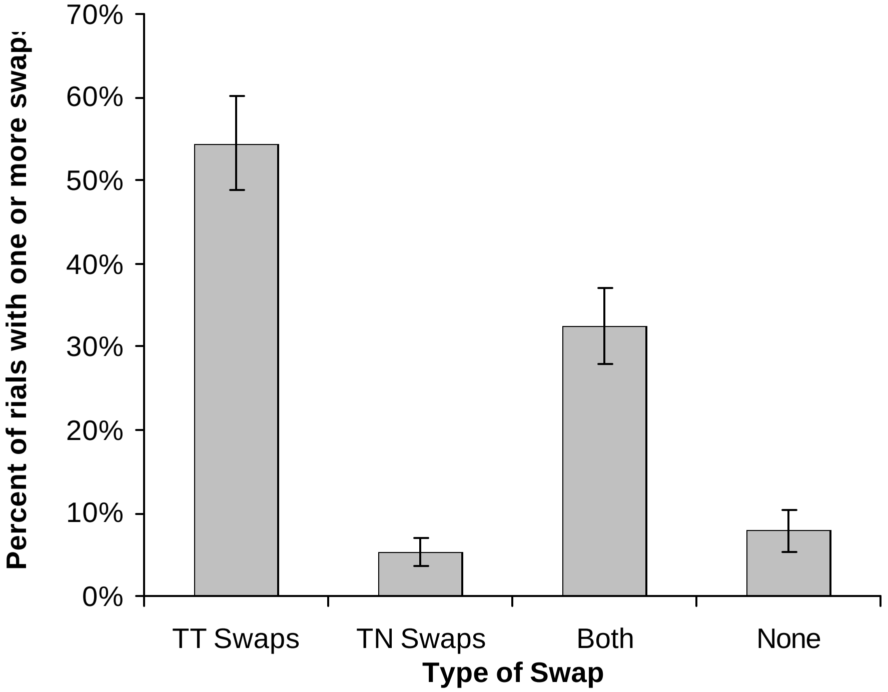

Figure 7. Chart showing the percent of trials containing swaps of different kinds (including trials with both types of swaps and with no swaps). Only the difference between the proportion of trials with TN swaps and with no swaps failed to reach statistical significance.

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 percent of the total number of TT and TN pairs at each distance range, is shown in Figure 8 (the measure is expressed as a percent 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 (a) 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), (b) there are more TT swaps than TN swaps resulting in a significant Swap Type effect (F=38.6*; df*=1,7; p< .000), (c) 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 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.

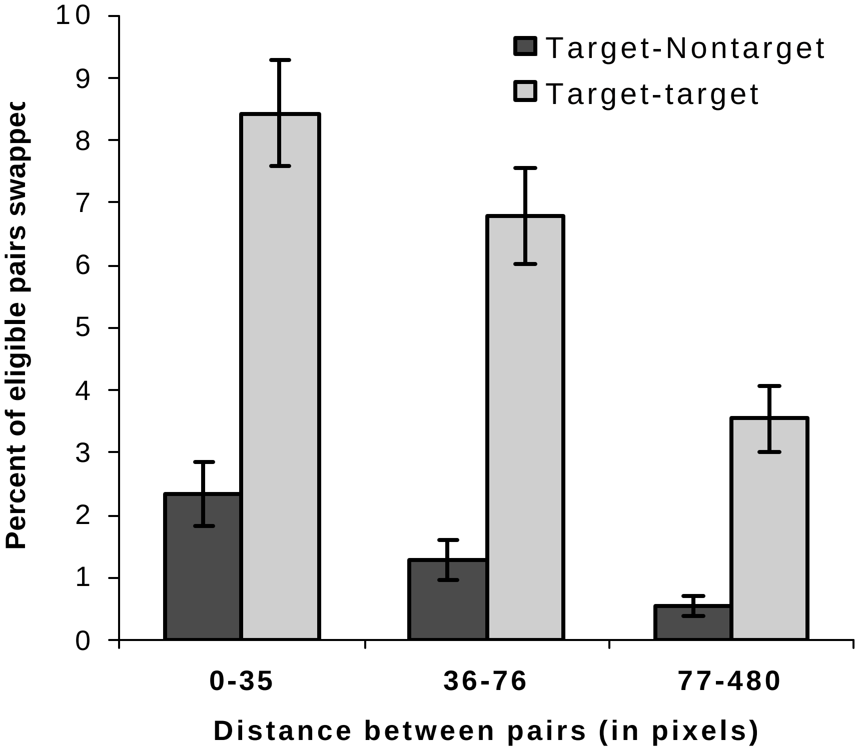

Figure 8. Graph showing that the proportion of both TT and TN swaps decreases as the distance between pairs increases (1 pixel corresponds to about 0.06 degrees of visual angle). Although the difference between the tendency for TT swaps and TN swaps decreases with larger distances, it remains highly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

Summary and Conclusions

The studies reported here suggest that observers are better at tracking 4 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 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 & 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, submitted).

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*, p56.

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

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature-space. *Nature, 408*(Nov 9), 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. J *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. (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. (submitted). Some puzzling findings in multiple object tracking (MOT): II. Inhibition of moving nontargets.

Pylyshyn, Z. W. (Ed.). (1988). Computational Processes in Human Vision: An interdisciplinary perspective. Stamford, CT: Ablex.

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 Proceedings of 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. (submitted). The relationship between property-encoding and object-based attention: Evidence from multiple-object tracking.

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.

Notes

[1] Email to: 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.

[2] 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.

[3] 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 ? 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 4 ways in which 1 target could be swapped with a nontarget, ${}^4C_2$ or 6 ways for choosing 2 targets to be swapped, ${}^4C_3$ or 4 ways for choosing 3 targets to be swapped and 1 way for all 4 targets to be swapped with nontargets, giving a total of 15 distinguishable ways of swapping targets with nontargets.

Some puzzling findings in multiple object tracking (MOT): II. Inhibition of moving nontargets

Zenon W. Pylyshyn

Abstract

We present three studies that examine the question whether multiple-object tracking (MOT) benefits from the active inhibition of nontargets, as proposed in an earlier study that reported target-target identity switching is more common that target-nontarget switching. Using a probe-dot technique, the first study showed poorer probe detection on nontargets than on either the targets being tracked or in the empty space between objects. The second study used a nontracking task to control for the possibility that this result was confounded by the placement of probes relative to the moving objects, independent of the fact that targets were being tracked. The third study examined the question of how localized the inhibition is and whether it spreads to nearby locations. The result of these three studies led to the conclusion that nontargets are subject to a highly localized object-based inhibition. Implications of this finding for the FINST visual index theory are discussed and it is concluded that we need to distinguish between a stage at which tokens representing enduring objects are differentiated and a stage at which these objects are made accessible through indexes, with only the second being limited to 4 or 5 objects.

Introduction

The idea of attention-related inhibition has been around for some time and has played a role in accounting for a wide range of phenomena, from memory to perceptual selection (Dagenbach & Kubat-Silman, 2003). The construct of inhibition has played a wide roll in vision science and has been an essential postulate in neuroscience theorizing, especially since the addition of inhibition as one of the basic processes in the formation of neural circuits (Houghton & Tipper, 1996; Milner, 1957). Yet the idea that the visual system might use inhibition to keep irrelevant (distractor) items from interfering with a primary task is not as well-studied. Recently (Watson & Humphreys, 1997) argued that items could be inhibited by a top-down process, called “visual marking,” based on the need to keep items with some particular properties out of reach of a primary search task. Many researchers have now replicated this finding and have also confirmed the goal-directed nature of the inhibition (Atchley, Jones, & Hoffman, 2003; Baylis, Tipper, & Houghton, 1997; Braithwaite & Humphreys, 2003) (although there is a question of whether the effect is purely top-down or whether it must be mediated by such visual events as abrupt offsets, Donk & Theeuwes, 2001).

In (Pylyshyn, 2004) we suggested that inhibition of nontarget items might help us to understand what goes on in the experimental paradigm known as Multiple Object Tracking (MOT). MOT has been used by a number of laboratories to study aspects of visual attention (see the review in Pylyshyn, 2001). In this experimental paradigm, observers track 4 or 5 objects (the “targets”) that move randomly among a set of identical, independently-moving objects (the “distractors”). While there are many variants of the MOT task, a typical experiment is illustrated in Figure 1. A number of simple items (typically about 8 circles or squares) are displayed on a screen. About half of these elements are briefly made visibly distinct, often by flashing them on and off a few times. Then all 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. After some period of time the motion stops and observers are required to indicate which objects are the targets. The experiment (and its many variants) has repeatedly shown that observers can track up to 4 and 5 items in a field containing the same number of identical distractor items over a period of up to 10 seconds with an accuracy of 85% - 95%.

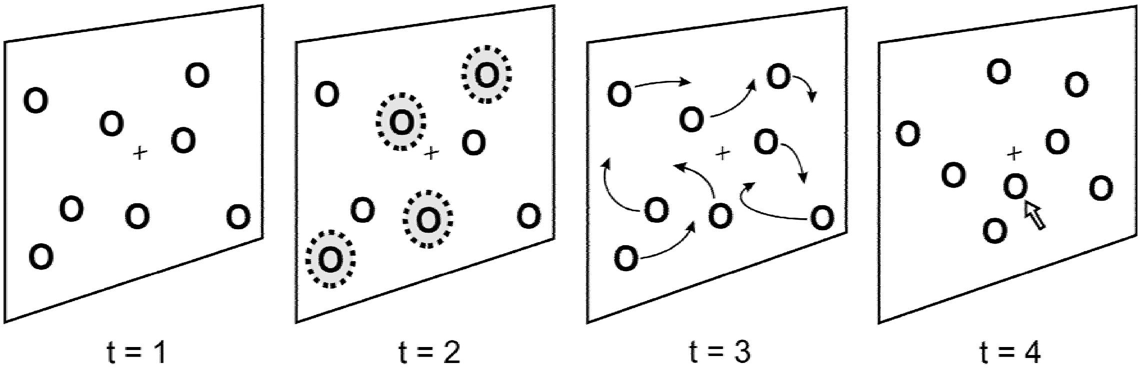

Figure 1. The sequence of events in a typical MOT experiment, in which the observer indicates whether the item flashed at the end of the trial was one that was being tracked (shaded circles indicate items being flashed).

The reason that we suggested that nontargets may be inhibited in this paradigm is that it would help account for the following puzzling finding. If we provide a unique identifier for each target (e.g., a number appearing inside the circle or a unique starting location such as one of the corners of the screen) observers are poor at recalling which identifier goes with which target, even when they have correctly track the targets in question. We showed that this arises because observers confuse (and switch identities between) target-target pairs more often than target-nontarget pairs. If the nontargets were inhibited this result would make sense since nontargets would effectively be taken out of the set of contending stimuli. This, in turn, entails that either everything that is not tracked is inhibited, or else that the individual moving nontargets alone are inhibited. Without some independent baseline measure of enhancement or inhibition, the first option (everything except targets is inhibited) is indistinguishable from the more natural view that tracked objects are attentionally enhanced.

The apparent enhancement of tracked targets relative to nontargets is well established and is implicit in MOT studies that required observers either to judge whether a selected item is a target or to detect/discriminate a feature on an item (Pylyshyn & Storm, 1988; Scholl & Pylyshyn, 1999; Sears & Pylyshyn, 2000). The object-based nature of this apparent enhancement has also been demonstrated in studies that measured either detection (Intriligator & Cavanagh, 1992) or discrimination of events on or off targets (Sears, 1991). The possibility that the inhibition applies only to individual moving nontargets, as opposed to applying to the entire region outside the targets themselves, has not been studied directly except for certain special cases. For example, there is evidence that moving items can be inhibited if they can be treated as a group, either because they share a common feature or because they maintained a rigid configuration (e.g., Kunar, Humphreys, & Smith, 2003; Watson, 2001; Watson & Humphreys, 1998a).

The possibility of object-based inhibition of moving objects is suggested by studies of Inhibition-of Return (IOR) where it was found that IOR tends to move with the inhibited object (Christ, McCrae, & Abrams, 2002; Tipper, Driver, & Weaver, 1991). A study by (Ogawa, Takeda, & Yagi, 2002) showed object-based inhibitory tagging in randomly-moving visual objects. Using a set of moving search items, they confirmed the earlier finding (Klein, 1988) that in difficult (non-popout) search, rejected nontarget items exhibit object-based inhibition, as assessed by a probe detection task. These cases all involve the inhibition of formerly attended items so are cases of IOR, as opposed to the sort of task-induced inhibition of irrelevant items that (Watson & Humphreys, 2000) referred to as “visual marking” and that, presumably, might be involved in the inhibition of nontargets in MOT. IOR differs in a number of ways from visual marking: it has a well-defined temporal pattern and reaches a maximum somewhere between 300 ms and 900 ms after initial attention-disengagement from the item. Moreover, in IOR studies, the inhibition is defined in relative terms by comparing detection performance on the inhibited item to performance elsewhere – typically on other items or in the space around the inhibited item. What has not been systematically studied are cases where both attention and inhibition may be involved. One exception is the work by (Ro & Rafal, 1999) that showed some independence of short-term (several hundred ms) enhancement and inhibition in situations where an uninformative cue attracts involuntary attention. There has been little evidence of object-based inhibition or visual marking occurring in paradigms such as MOT, where inhibition may function to facilitate performance, perhaps in addition to attentional enhancement. Consequently if punctate object-based nontarget inhibition of the visual-marking sort is to explain the selective identity-switching findings reported in (Pylyshyn, 2004) we need to show that individual randomly-moving nontargets are inhibited during tracking.

The possibility that nontargets are inhibited relative to the entire display (especially relative to the background), over-and-above any inhibition relative to the targets, has ramifications for theories of tracking such as the FINST Visual Index Theory (Pylyshyn, 2001). The FINST theory (as well as theories of MOT based on split attention Scholl, 2001) postulate a limited capacity mechanism that keeps track of target objects. According to such accounts, however, non-target objects are not tracked and therefore there is no provision for keeping inhibition attached to them in a punctate manner without at the same time inhibiting the entire extra-target region. Thus it is of some theoretical interest whether in tasks such as MOT inhibition occurs on nontargets relative to both targets and empty space. The present experiments were designed to examine this question.

General Method

The experiments reported here were designed to examine whether nontargets in the MOT task are inhibited relative to targets and *also* relative to the background of the display. The measure of inhibition used was the dot-probe detection task, a task used with success by (Watson & Humphreys, 1997) as well as others (Donk & Theeuwes, 2001; Olivers, Watson, & Humphreys, 1999; Theeuwes, Kramer, & Atchley, 1998; Watson & Humphreys, 1998b) to measure inhibition effects on specific visual items. The measure assumes that performance in detecting a small faint dot in a particular location provides an indication of the availability of attentional resources at that location, and therefore that it serves as a measure of either attentional enhancement or inhibition. Because we are interested in distinguishing attentional enhancement from inhibition, we need to compare the measure at three or more distinct locations: for example, on targets, on nontargets and in the empty space between them. If the effect is one of inhibition, then probe detection should not only be worse on nontargets than on targets, but it should also be worse on nontargets than at other locations. Experiment 1 presents the basic study. Other experiments control for various possible confounds and also explore the spatial distribution of attention or inhibition. The present section describes the method used in all 3 experiments.

Materials and apparatus

The experiments were programmed using the VisionShell© graphics libraries (Comtois, 2003) and were presented on iMac computers. The circles in the tracking task consisted of white outline rings with dark interiors and were displayed on a dark background. The interior dark region was drawn as opaque so that when one of the circles passed by another, occlusion cues (T-junctions) showed one of the circles to be in front of the other. The circles were 47 pixels or 2.7 degrees of visual angle with outer rings 2 pixels (approximately 0.12°) thick.

The motion algorithm is the same as that used in other recent MOT experiments. Each item was assigned a random initial location and also a random horizontal and vertical velocity component 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 motion. In the present experiments, this probability was set at 0.10, 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. In the resulting motion, items could move a maximum of 0.12º 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 deg/s, with an average velocity across all items and trials of 2.37 deg/s.

The probe dot used was adjusted so that the error rate would be most sensitive to our primary dependent variable (probe location). The probe was a red square 6 x 6 pixels (approximately 0.34º x 0.34º) displayed for 136 ms. Probes occurred on half the trials and were equally often distributed among the locations being tested in each experiment (e.g., in Experiment 1 they occurred equally often on targets, nontargets or in the space between them). On trials containing probes, the probes occurred once at a randomly chosen time in the third or fourth second of the 5 second trial.

Procedure

After being instructed on the tracking and probe detection responses required, observers were told that since only trials in which they correctly tracked the targets could be used, they should place special emphasis on the tracking part of the task. Participants pressed a key to start each trial. There were 5 practice trials at the beginning of each experiment. Each trial began with 8 static circles in the screen. 4 of these flashed on and off a few times, then all 8 circles began to move. After 5 seconds, all circles stopped moving. Observers then had to select the four circles that had been indicated as targets, using a computer mouse. After making these 4 responses, a screen appeared with the question: “Did a red dot appear anywhere during this trial?” and observers made a forced choice response by selecting one of two labeled buttons on the screen. All responses were recorded automatically and stored on the computer disk. Only after the set of 5 responses were completed was the next trial allowed to proceed.

Experiment 1

Participants

Eighteen Rutgers undergraduates participated either as part of their course requirements or for remuneration. Two additional participants were omitted from the analysis because their overall tracking performance or probe detection performance was too low (tracking below 65% or probe detection below 50%).

Method

The method was as described above. In the Empty Space condition a probe location was chosen at random subject to the constraint that it was located at least two diameters (5.4 °) from any other circle or from the edge of the screen. In the Target and Nontarget conditions the probe was always located at the center of the circle.

Results

Probe-dot detection performance was analyzed using a within-subject ANOVA. The effect of location was significant (F=21.3, df=2,34, p<000). A post-hoc paired comparison of the performance at 3 locations revealed that probe detection at the nontarget location was significantly worse (p <.001) than at either the Target Location or at the Empty Space location. There was no statistically reliable difference between the Target location and the Empty Space location (p>.32) (using the Bonferroni correction for multiple comparisons). These results are shown in Figure 2.

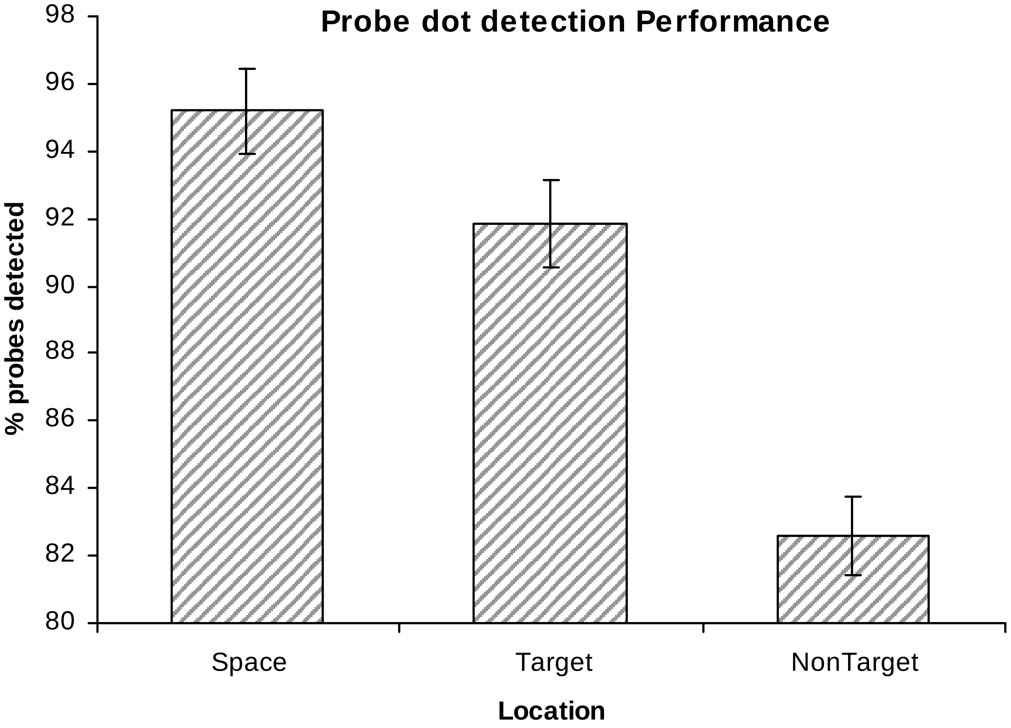

Figure 2. Performance in detecting a probe dot at three types of locations during a multiple object tracking task

Discussion

The results provided confirmation of the hypothesis that in MOT the Nontarget items are inhibited relative to the target items and also relative to the empty space between items. Probe detection on targets and on empty space did not differ significantly. Thus the results provide no support for the hypothesis that tracked targets are enhanced relative to the rest of the display, as opposed to being enhanced only relative to the nontargets.

Although the inside of the circular objects were the same color and brightness as the background, it is still possible that a probe occurring far from a moving object might be more easily detected than one occurring on an object, independent of any effect of the tracking itself.

In other words a probe surrounded by empty dark space may be less subject to masking effects and therefore might be more visible than one that occurs at the center of a 2.7 ° diameter circle. This would not affect the difference between probe detection on targets and nontargets, since these are physically identical, but it could effect the detection of probes in the empty space condition. Thus it might be that the effect we found, in which detection in empty space was more like that on targets, was the result of the superiority of empty space detection, superimposed on the difference between targets and non-targets. In other words it might be that the empty space is actually inhibited as much as the nontargets, but that the greater visibility in the empty region raised probe detection performance. If that were the case we would not be entitled to conclude that inhibition was specific to nontargets, as opposed to being a general inhibition of everything in the scene, and thus it might be that what we were observing was the effect of the relative enhancement of targets.

To control for this possibility, Experiment 2 introduced a control condition in which we tested the detection of probes at the same 3 sites when observers were not tracking anything, but merely watching the 8 objects moving on the screen. In this control condition, there is no distinction between “targets” and “nontargets” since none of the objects was flashed and observers were not asked to track any objects. The control condition was described to the participants simply as the task of detecting probes in the presence of moving distractors. In order to discourage observers from spontaneously tracking some of objects, the control task was presented first before the tracking condition – and before any mention of object-tracking.

Experiment 2

Participants

Twenty-four volunteers from the undergraduate subject pool participated to fulfill course requirements.

Method

The method is the same as in Experiment 1, with the addition of a block of control trials that were identical to the experimental trials except that they involved no tracking. The control trials preceded the tracking trials and involved only a single two-alternative forced choice response at the end of each trial. There were 60 control (no tracking) trials and 120 experimental (tracking) trials, in half of which there was no probe. In the experimental (tracking) trials observers were asked to first pick out the targets by clicking on them using a computer mouse and then to make a forced choice response to the question whether a probe had appeared in that trial, as described in the general method section above.

Results

An analysis of the non-tracking control trials revealed that performance on the probe detection task was indeed better when the probe appeared in the empty space than on the circles ($t=4.5$; $df=23$; $p<.000$), thus raising the possibility that the effect reported in Experiment 1 may be due to a combination of target enhancement and superior probe detection for the probes in empty space. Thus we proceeded to examine the quantitative relation among the probe detection performance at different locations in order to ascertain whether it is compatible with this interpretation. To do this we analyzed the control and experimental conditions together using a within-subjects analysis of variance.

The analysis of variance revealed a significant difference between control and experimental conditions ($F=8.38$; $df=1,23$; $p<0.01$), and between the three different probe locations ($F=28.27$, df=2,46; p<.000), as well as a significant interaction between these two factors (F=6.10; df=2,46; p<.01). A post-hoc analysis revealed that the locations were significantly different from one another, but the difference between control and experimental condition was only significant when the probe occurred on nontargets (t=4.7; df=24, p<0.000). In other words, only probe detection on nontargets was affected by the presence of the tracking task, over-and-above the matching control condition. This result supports the conclusion that tracking causes the inhibition of probe detection on nontargets, as opposed to enhancing the detection on targets (or inhibiting everything but targets). These results are shown in Figure 3.

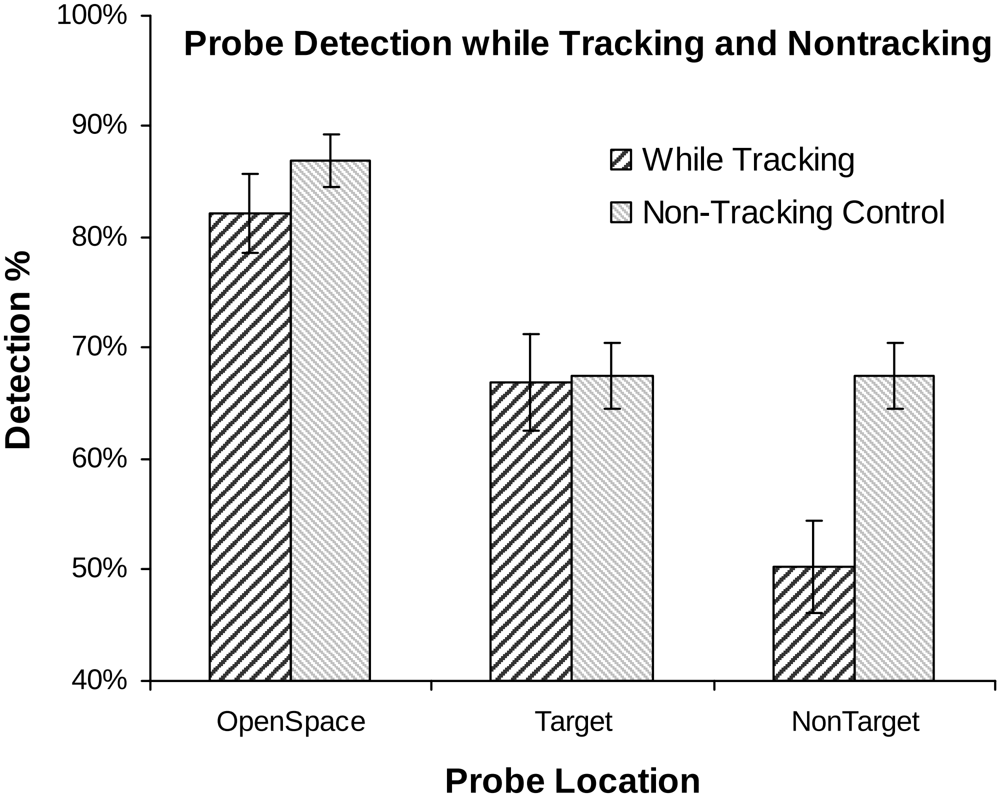

Figure 3. Performance in detecting a probe dot during tracking and also during the same probe detection task when there was no tracking. (Note: In the nontracking control task there is no distinction between targets and nontargets, so there are twice as many target/nontarget trials averaged together and reproduced twice in this graph, thus resulting in a lower confidence interval.)

The difference between the absolute values of probe detection performance in the control (nontracking) condition and the experimental (tracking) condition was confounded by the fact that the tasks were performed in separate blocks in a fixed order (nontracking first) in order to discourage tacit tracking. Moreover, since the experimental condition requires keeping two tasks in mind it may leave less attention for probe detection and result in lower overall detection performance. Because of this we adopted a second way of exhibiting the results which takes into account not only the baseline (nontracking) probe detection performance but also the statistical correlations between control and experimental conditions at each of the three locations. To do this we performed an analysis of covariance, using the method described in (Green, Salkind, & Aken, 2000, Lesson 26). The resulting graph of adjusted detection rates is shown in Figure 4, along with the unadjusted detection rate. It confirms the pattern found in the uncorrected detection means and shows, perhaps even more graphically, that only the nontarget performance was impaired relative to both target and empty space performance.

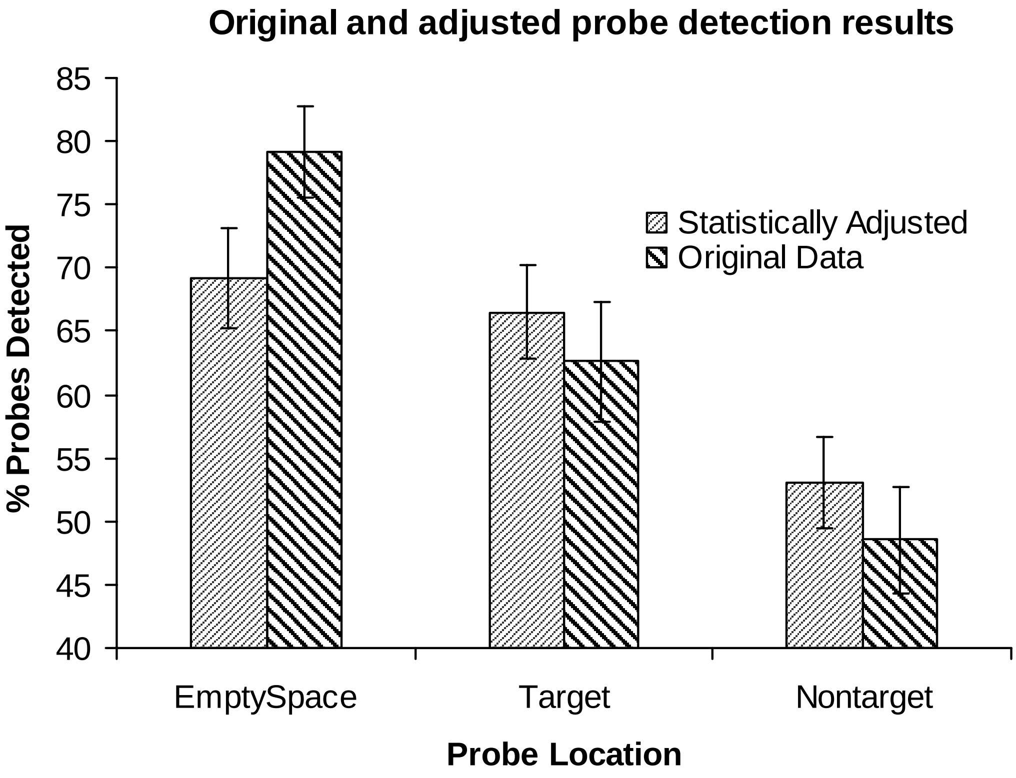

Figure 4. This graph shows the original probe detection results together with results obtained when the nontracking control task was used as a covariate for adjusting the probe detection score measured during tracking. A convariance analysis was used to predict what the performance would be if the three conditions were equated for their nontracking difficulty.

Discussion

The results of experiments 1 and 2 offer no support for the hypothesis that tracking enhances detection on targets. Rather they suggest that nontargets were inhibited in a punctate manner and that the inhibition appears to be object-based. These results do not, however, cast any light on how local the inhibition is and whether it drops off slowly with distance from the nontargets. The question of the locality of inhibition is important to theories of attention and inhibition since it is generally believed that attention drops off slowly as one goes away from the attentional focus (Cheal, Lyon, & Gottlob, 1994) and thus one might expect that inhibition does as well. The probe detection method has been used successfully to plot the gradient of attention in other tasks, including where moving objects are involved (Kerzel, 2003), so we continued to use that measure.

Experiment 3

In order to assess how localized the attention and inhibition was during tracking, Experiment 3 was designed to test additional locations near to targets and nontargets. In this study we tested five different locations with the probe-dot detection task. These included the three used in experiment 2 and also one that was one-radius (1.35°)away from the target and another that was one-radius away from the nontarget. In other words we presented a probe at the same distance from the circular contour as those that were on the targets or on the nontargets, but outside the circle. These are referred to as the Near Target and Near Nontarget conditions. A nontracking control condition was presented as in Experiment 2. In addition, in order to see whether there was any generalized effect of the tracking task, over and above what might be described as an effect of poorer visibility, crowding, or masking in the case of the probes closer to (or inside) the moving objects, we also included a control condition similar to the one used in Experiment 2, but in which the circles did not move (the “static control” condition). Both control conditions provide a baseline measure of probe detection unaffected by the distinction between targets and nontargets (since in neither case was the difference between targets and nontargets visually marked). The static control condition, however, was also free of any motion, and therefore provided a more direct test of the visibility/masking hypothesis. It also controlled for any possible implicit secondary task, such as involuntary tracking of moving objects.

Participants

The data for the experiment were collected on sixteen naïve volunteers who responded to a recruiting poster and participated for a small remuneration. Data from two additional participants were not used on the grounds that their probe detection scores in the moving control condition was at chance. In addition we recruited 4 volunteers who had considerable experience with MOT. These were added to the pool to make a total of 20 participants, although the experienced volunteers were also examined separately and their results reported.

Method

The method is the same as in Experiment 2 except that two additional probe locations were used and half of the control trials were ones in which the objects did not move. For the control condition, participants were told simply that the task was to see how well they could detect small red dots that occurred among static or moving circles. The control trials preceded the tracking trials and involved only one two-alternative forced choice response per trial. The experiment was preceded by 5 practice trials. Then there were 50 static trials and 50 moving trials randomly order in the control block followed by 100 trials in the experimental block. As before, half the experimental trials had no probes while the other half had probes distributed equally among the 5 locations. Three of the locations were the same as those used in Experiment 2; they were located in empty space, on target, and on nontarget. In addition there was one located one radius (approximately 1.35 º) from a Target (Near-Target), as well as one radius from a Nontarget (Near-Nontarget).

Results

Examination of the static control condition revealed that the difference in probe detection accuracy was not due to visibility or crowding or lateral masking. Despite having been collected at the very start of the experimental session, scores in the static control condition were at ceiling, ranging from 95.9% (on targets) to 98.9% (on Near Nontargets) and the difference among them did not approach significance (F=0.79, df=4,116, p>0.5). Therefore only the moving control condition was analyzed further.

A within-subjects analysis of variance showed that probe detection in the tracking condition was significantly lower than in the (moving) control condition (F=12.2, df=1,19; p<.02), the detection rate was significantly different among the 5 locations (F=15.6; df=4,76; p<.000), and the interaction of these two factors was also significant (F=2.6; df=1,76; p<.05). Figure 5 shows the probe detection scores for the control condition and for the tracking condition at each of the 5 locations. Planned comparison t-tests revealed that, as in Experiment 2, the only difference between the control and experimental condition that was statistically reliable (using the Bonferroni correction for multiple tests) was on the nontarget (t=4.5, df=19, p<.000). (The comparison of the means on the next largest pair, the empty space condition, had t=2.4, df=19, which gave a corrected p > .05)

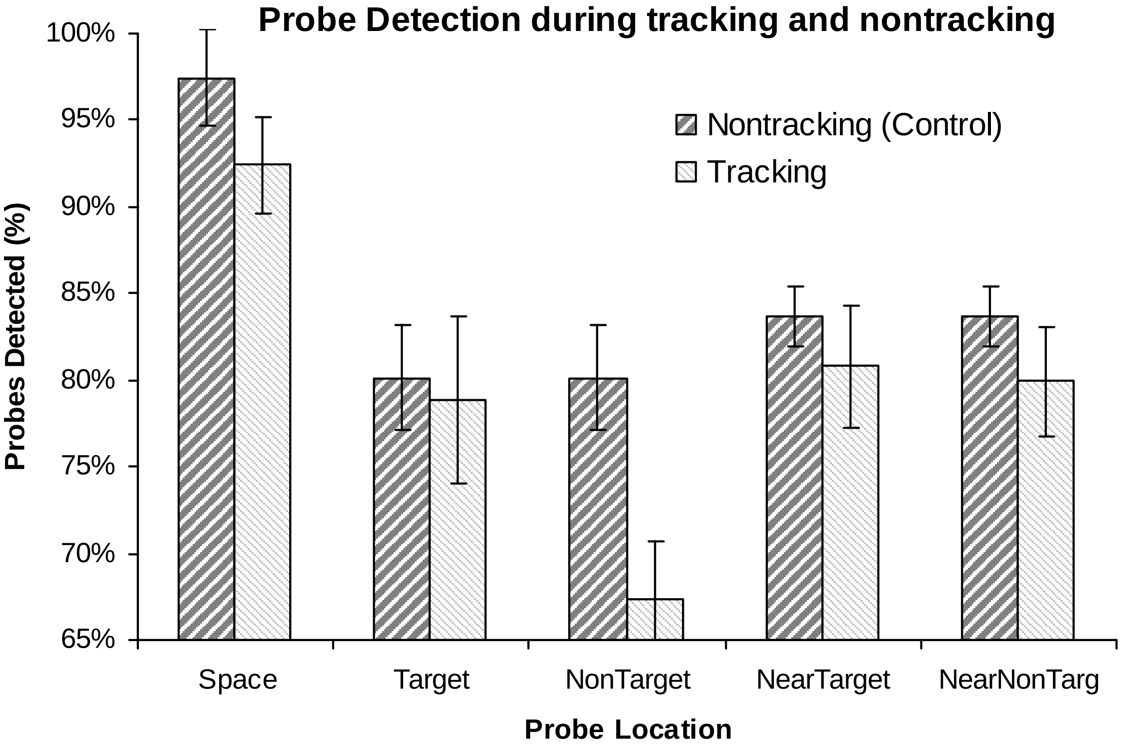

Figure 5. Probe Detection performance in the moving (nontracking) control and in the tracking condition, as a function of the location of the probes. Only the difference at the nontarget was significant.

As mentioned earlier, four of the participants had had considerable experience with the MOT task, having participated in previous experiments. These were also highly motivated and were willing to provide 600 trials in 3 one-hour sessions. Consequently it seemed prudent to see how they performed in relation to the general subject pool. The results for these expert subjects are shown in Figure 6, using the same scale as used to show the results for the other subjects in the previous figure. Even with only 4 subjects (over three blocks of trials), the results are statistically significant: there was a significant control vs tracking difference (F=17.6, df=1,3; p<.05), a significant probe location effect (F=8.4; df=4,12; p<.002), and a control-tracking by location interaction (F=3.5; df=4,12; p<.0.05). It is apparent from the figure that these subjects (a) performed better at detecting probes, especially on the targets, and (b) showed the same inhibition of nontargets as observed with the naïve participants.

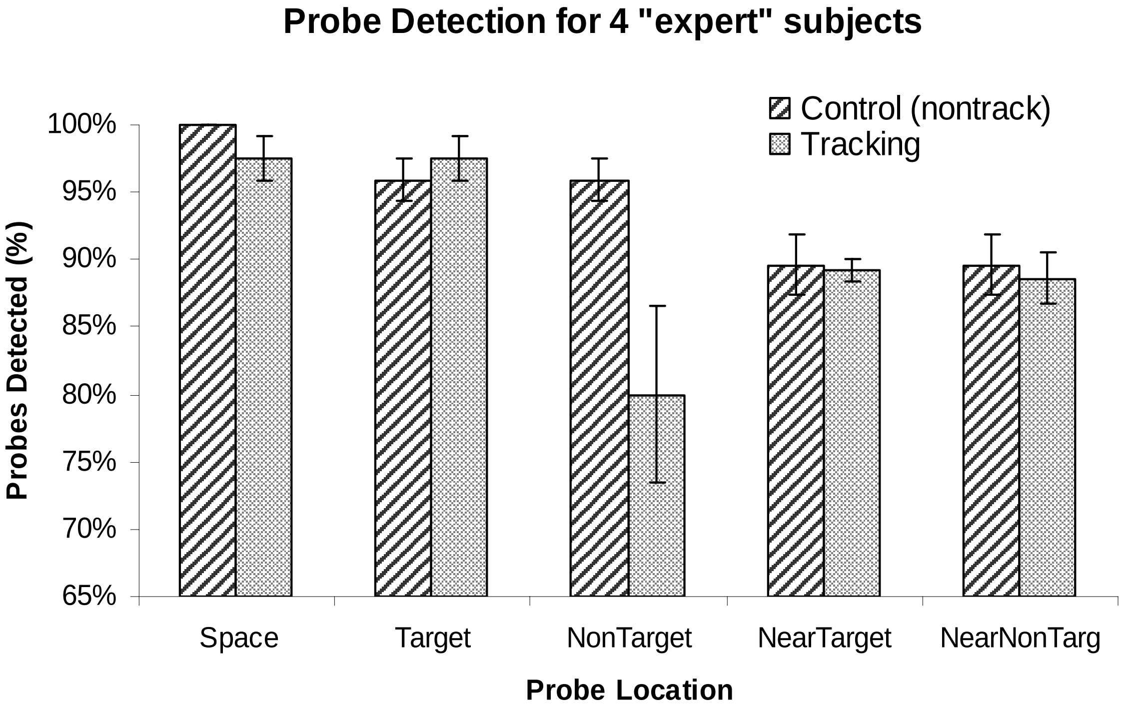

Figure 6. Graph of probe detection performance by four volunteers who had a great deal of experience with MOT and were willing to provide several hours of data. Although they performed better that the other participants, they show the same decrement on the nontargets.

The difference between the pattern of probe detection performance in the control condition and in the tracking condition is an indication of the degree of inhibition observed at each location. This difference is plotted in Figure 7 for the 20 participants and shows that inhibition is highly local, not even reaching one radius distance from the nontargets, even though these probes were the same distance from the outside circular contour as they were in the “on nontarget” probes. Other than the difference at the nontarget location, no other difference is significant although the effect in the empty space (more than 2 diameters from any circle) approached significance. As noted earlier, the absolute values depicted in this chart cannot be univocally interpreted since the control block always preceded the experimental block. Since the unexpected suppression effect at the empty space location may be due to some combination of an order effect and a general dual-task effect, we might take the value at empty space as a neutral baseline. In we show the origin at that value (as in the dotted line in Figure 7) we see that there is some basis for conjecturing that there may actually be some attentional enhancement at the target which even spread slightly to the nearby location. However the evidence for this in the present study is highly tentative.

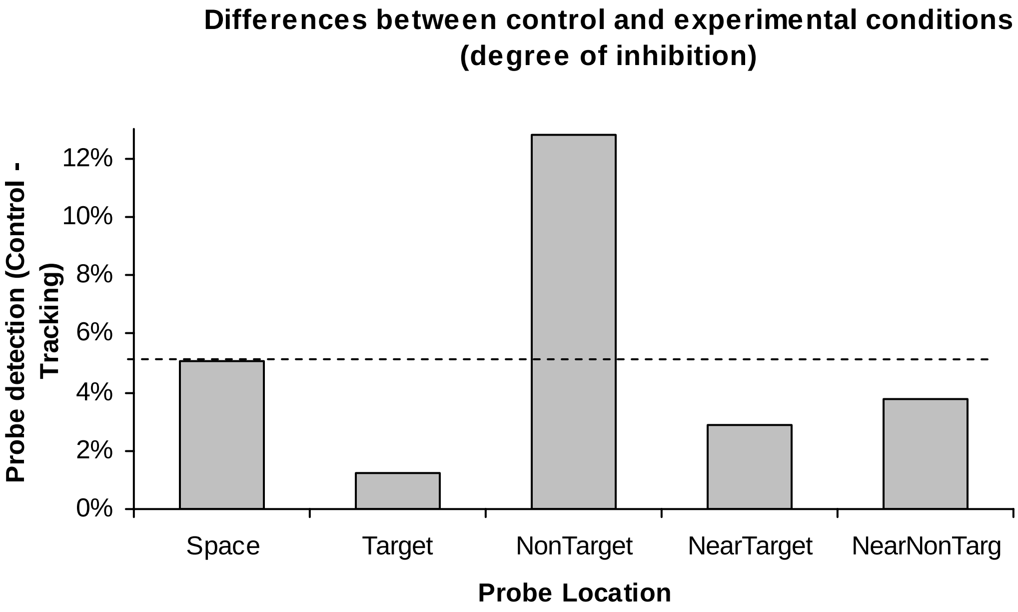

Figure 7. This figure shows the degree of inhibition at each probe location. The dotted line represents a possible baseline for measuring the degree of inhibition based on the assumption that the inhibition in empty space is due solely to the effect of a secondary task or of the order in which the control and experimental conditions were experienced. Based on that assumption one might take this figure as suggesting some degree of attentional enhancement at the targets (i.e., the 5% below this baseline at the target location could be seen as an enhancement), as well as a strong inhibition at the nontargets.

As in Experiment 2, a revealing presentation of these results might use statistical regression to adjust the probe detection rate based on the correlations between the control and tracking performance at the five locations. The covariance analysis revealed a significant effect of probe location after adjusting for the control data (F=2.58, df=4,94, p<.05) and also showed that the only pairs of locations that are significant (using the Bonferroni correction) are those between the nontarget position and all others. The result of this analysis is shown in Figure 8 and reveals clearly that after the statistical adjustment all locations are equal in the probe detection performance except for the significant depression at the nontarget location, again confirming that only the nontargets appear to be inhibited.

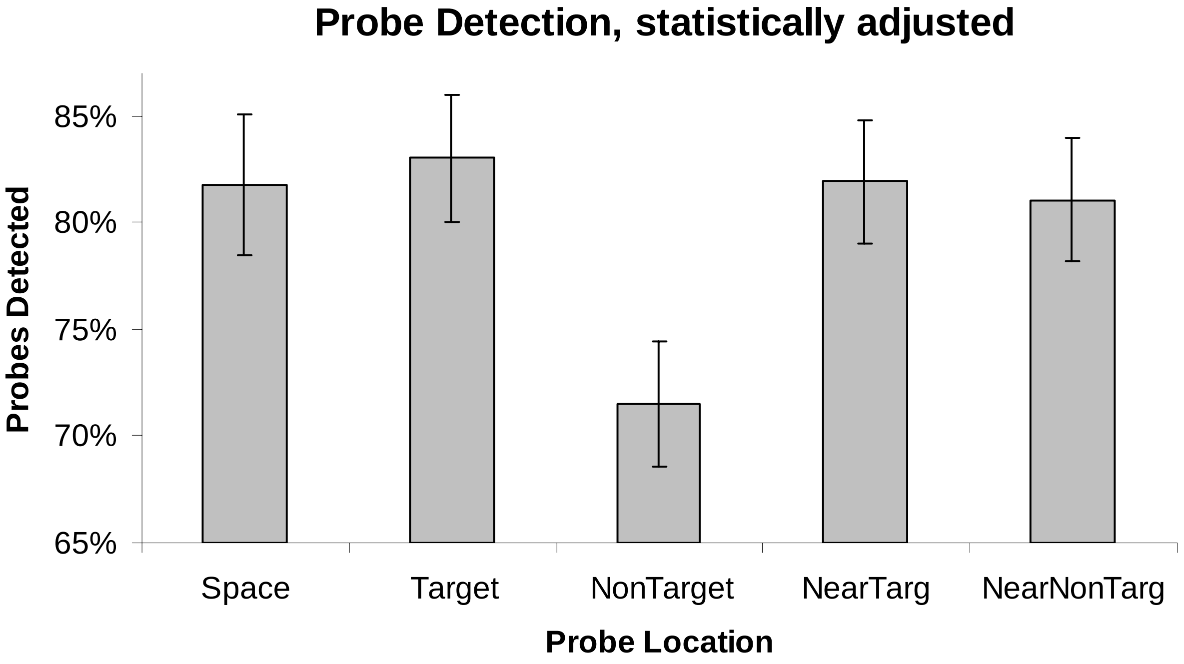

Figure 8. This figure shows the probe detection performance adjusted using a covariance analysis. The adjustment takes account of the correlations between the control and tracking conditions at each location.

Finally we performed an additional precautionary analysis of the records of trajectories actually used in this study. Although circles were located a random and moved in a random manner (subject only to speed and acceleration constraints), probe locations were subject to additional constraints. Probes on targets and nontargets were located at the center of the circles. Near Target and Near NonTarget probes were located at random subject to the constraint that they be one radius (1.35 °) from the relevant circle and more than one radius from any other circle and from the edge of the display. Empty space probes met the most stringent criterion as they had to be at least 2 diameters (5.4 °) from any circle. It is thus possible that in order to meet all these constraints, the probes in some conditions (e.g., the empty space condition) might have ended up more or less concentric than in other conditions. Since eccentricity could be a major factor in their visibility, this possibility needed to be excluded. Fortunately we had a record of the trajectories of the objects used in these studies, so we were able to take a sample of probes in each of the 5 conditions to compare their eccentricities. On a sample of 264 probes at each of the 5 locations we found no significant differences in their eccentricities (F=.732, df=4,1052; p>0.57). The empty space probes were not even nominally at the extremes of this distribution but somewhere between the targets/nontargets and the neartarget/nearnontarget eccentricities whose means lay in the range from 178 and 186 pixels, so that the means were within 0.5 º of each other.

Discussion

Results of experiment 3 were consistent with the hypothesis that nontarget items are inhibited in MOT and further showed that this effect appears to be confined to the immediate region of the moving nontargets. This raises some questions about the mechanism that may be responsible for this effect, which is discussed in the next section.

General Discussion

This study began with the hypothesis that in MOT, nontargets are segregated from targets by virtue of an inhibitory process that specifically affects the individual nontarget objects. The evidence presented here suggests that nontargets are inhibited over and beyond the enhancement of targets or the general inhibition of everything that is not being tracked, and that the inhibition is highly local to nontargets. This finding is consistent with our earlier hypothesis (Pylyshyn, 2004) that the reason that in MOT targets are more often confused with (i.e., identities are switched with) other targets than with nontargets, is that nontargets are suppressed. But the finding raises a further theoretical question: How can moving objects alone be inhibited without the inhibition affecting the space through which they travel? There are at least two possibilities. One is that only individual nontargets are inhibited and the inhibition travels with the nontargets as they move (i.e., that it is object-based, in the sense in which this term has been used in the attention literature). This possibility is consistent with the evidence on object-based inhibition of return (IOR) cited earlier. But the only way that inhibition could move with a moving object is if that object is being tracked in some way; if it is recognized as the same object over time. But Visual Index (FINST) Theory only provides a mechanism for tracking about 5 objects. If nontargets as well as targets are being tracked in MOT then up to 8 items would have to be tracked. This problem was noted by (Ogawa et al., 2002), who also found that up to 8 moving items could be inhibited in a search paradigm, leading them to suggest that "inhibitory tagging" involved some mechanism other than FINSTs.

A second possibility is that the inhibition does not actually move, but rather is directed in a more general manner that nonetheless excludes empty space. So, for example, inhibition might be directed at all unattended moving objects, or all unattended objects sharing some property, such as color or shape. There is evidence for the inhibition of groups of items sharing a common property such as color or shape (Kunar, Humphreys, & Smith, 2003), configuration (Kunar, Humphreys, Smith, & Hulleman, 2003), order of presentation (Watson & Humphreys, 1997), or time of onset (Watson, Humphreys, & Olivers, 2003) and that this selective inhibition may depend on the goals of the task (Watson & Humphreys, 2000). The current evidence does not allow us to choose between these two main possibilities, and indeed both may hold.

The FINST visual index theory only provides a mechanism for tracking up to 5 items, and therefore such a mechanism could not explain how nontargets in these experiments could be tracked, as required by the hypothesis that punctate inhibition follows the moving nontargets. Is there any other way that objects could be tracked other than through a mechanism such as FINST indexes or pointers? Perhaps we need to refine out concept of tracking.

There are independent reasons for thinking that some form of "tracking" must be possible for more items than the limit of 5 generally found in MOT. For example, in order to carry out a search on a large number of moving items (as in the experiment of Ogawa, Takedo & Yagi, as well as many other studies – e.g., Alvarez, Horowitz, & Wolfe, 2000; Cohen & Pylyshyn, 2002), vision must maintain the integrity of the candidate objects as they move, otherwise no two time slices would be perceived as containing the same set of objects, and thus only a repetitive exhaustive scanning of all locations in the display could lead to a successful match in such moving-search experiments. In addition, solving the ubiquitous "correspondence problem" appears to require the preattentive identification of large numbers of visual clusters.[1] The correspondence problem is a problem that is solved whenever two initially distinct visual tokens are put into correspondence and thereby treated by the visual system as arising from one and the same distal object. This problem is routinely solved in apparent motion and stereo, and moreover it appears to be solved over some prior segregation of visual tokens. For example, (Ullman, 1979) showed that apparent motion is computed over distinct tokens, as opposed to a smooth intensity map. Since apparent motion can involve large numbers of token elements (as in the "kinetic depth effect" – Wallach & O'Connell, 1953), the correspondence problem must be solved over many tokens which, in turn, means that many such tokens must be distinguished in early vision and assigned the same persisting identity – far more than the capacity of the FINST mechanism. The same is true of stereo vision, where tokens in each eye must be placed in correspondence in order to compute the disparity of the corresponding distal element. These phenomena all call for distinguishing a large number of token elements at the same time and keeping track of their persisting identity as they move. Since stereo can be computed over a moving field of dots (as in dynamic random-dot stereograms, Julesz, 1971), the stereo correspondence problem has to be solved even when the tokens are in motion which, in turn, means that the temporal correspondence must be solved first. Thus we have independent reasons to believe that segregation of moving elements takes place and is not subject to the same sorts of numerical limits as postulated by FINST theory, or as found in MOT.

This suggests that MOT, and other findings for which visual indexing has been invoked, involves at least two stages. Before visual objects can be indexed, a scene must first be parsed into tokens and the tokens merged so they refer to individual candidate objects or proto-objects. This can be carried out by a process operating in parallel across the scene. Such a process not only segments the scene into distinct spatial regions, but also solves the correspondence problem for the moving tokens, thus treating them as arising from persisting objects. Parallel processes of this sort are well-known in the study of early vision, and various models for their implementation have been proposed (see, Dawson & Pylyshyn, 1988; Koch & Ullman, 1985; Ullman, 1976). Only after a scene has been parsed into such persisting visual objects can pointers be attached to a subset of these objects. This idea is in fact explicit in the original FINST theory, where it is recognized that indexes are only assigned to a subset of the possible objects in a scene. What the present findings (as well as those of Ogawa et al., 2002, and the studies of object-based IOR cited above) suggest is that inhibition is applied to these persisting visual objects before they are indexed, and therefore at a stage prior to when they can be accessed. Such access is required for purposes such as responding correctly in MOT (by picking out the targets using a computer mouse), making judgments about them (as in computing "visual routines", Ullman, 1984), enumerating or subitizing them, and so on (for more on this notion of access see Pylyshyn, 1989, Chapter 5). Thus a clear prediction of this theory is that nontargets cannot be rapidly enumerated or subitized, nor can patterns such as collinearity be recognized over them, even though object-based inhibition has been applied to them. Thus the

---

[1] To keep certain distinctions clear and to remain as close as possible to general terminological usage in psychology, I refer to proximal figure-ground groupings as "tokens," and to the enduring distal elements to which these tokens correspond as "visual objects." In the past I have sometimes used the term "individual" (e.g., Pylyshyn, 2001) to refer to elements that meet the two conditions that: (a) they are distinguished from the region around them, as in Gestalt "figure-ground segregation" and (b) they have been merged by a correspondence operation so that they correspond to hypothetical enduring distal objects. In the more general literature, particularly in philosophy, an "individual" is a token that is not only distinct from the region around it but also from each of the other tokens: the signature property individuals is that they can, in principle be counted. To preserve the distinction between this more general sense of individual and the more limited sense that I have in mind here, I will follow the terminology that is common in psychology and refer to the referents of enduring post-correspondence tokens as "visual objects" or "proto-objects".

evidence for object-based inhibitory tagging provides a prima-facie argument for a certain sort of quasi-individuation as a separate stage of the indexing and accessing process.

References

Alvarez, G. A., Horowitz, J. M., & Wolfe, J. M. (2000). Multielement tracking and visual search use independent resources [Abstract]. *Investigative Ophthalmology & Visual Science, 41*(4).

Atchley, P., Jones, S. E., & Hoffman, L. (2003). Visual marking: A convergence of goal- and stimulus-driven processes during visual search. *Perception & Psychophysics, 65*(5), 667-677.

Baylis, G. C., Tipper, S. P., & Houghton, G. (1997). Externally cued and internally generated selection: Differences in distractor analysis and inhibition. *Journal of Experimental Psychology: Human Perception & Performance, 23*(6), 1617-1630.

Braithwaite, J. J., & Humphreys, G. W. (2003). Inhibition and anticipation in visual search: Evidence from effects of color foreknowledge on preview search. *Perception & Psychophysics, 65*(2), 213-237.

Cheal, M., Lyon, D. R., & Gottlob, L. R. (1994). A framework for understanding the allocation of attention in location-precued discrimination. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 3,* 699-739.

Christ, S. E., McCrae, C. S., & Abrams, R. A. (2002). Inhibition of return in static and dynamic displays. *Psychonomic Bulletin & Review, 9*(1), 80-85.

Cohen, E. H., & Pylyshyn, Z. W. (2002). Searching through subsets of moving items [Abstract]. *Journal of Vision, 2*(7), 541a.

Dagenbach, D., & Kubat-Silman, A. K. (2003). The principle of inhibition. In R. H. Kluwe & G. Lueer & e. al. (Eds.), *Principles of learning and memory.* (pp. 207-221). Cambridge, MA, US: Birkhaeuser.

Dawson, M., & Pylyshyn, Z. W. (1988). Natural constraints in apparent motion. In Z. W. Pylyshyn (Ed.), *Computational Processes in Human Vision: An interdisciplinary perspective* (pp. 99-120). Stamford, CT: Ablex Publishing.

Donk, M., & Theeuwes, J. (2001). Visual marking beside the mark: Prioritizing selection by abrupt onsets. *Perception & Psychophysics, 63*(5), 891-900.

Green, S. B., Salkind, N. J., & Aken, T. M. (2000). *Using SPSS for Windows (second edition).* London, UK: Printice-Hall Inc.

Houghton, G., & Tipper, S. P. (1996). Inhibitory mechanisms of neural and cognitive control: Applications to selective attention and sequential action. *Brain & Cognition, 30*(1), 20-43.

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

Julesz, B. (1971). *Foundations of Cyclopean Perception*. Chicago: Univ. of Chicago Press.

Kerzel, D. (2003). Attention maintains mental extrapolation of target position: Irrelevant distractors eliminate forward displacement after implied motion. *Cognition, 88*(1), 109-131.

Klein, R. M. (1988). Inhibitory tagging system facilitates visual search. *Nature, 334*(6181), 430-431.

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

Kunar, M. A., Humphreys, G. W., & Smith, K. J. (2003). Visual Change With Moving Displays: More Evidence for Color Feature Map Inhibition During Preview Search. *Journal of Experimental Psychology: Human Perception & Performance, 29*(4), 779-792.

Kunar, M. A., Humphreys, G. W., Smith, K. J., & Hulleman, J. (2003). What is "marked" in visual marking? Evidence for effects of configuration in preview search. *Perception & Psychophysics, 65*(6), 982-996.

Milner, P. M. (1957). The cell assembly: Mark II. *Psychological Review, 64,* 242-252.

Ogawa, H., Takeda, Y., & Yagi, A. (2002). Inhibitory tagging on randomly moving objects. *Psychological Science, 13*(2), 125-129.

Olivers, C. N. J., Watson, D. G., & Humphreys, G. W. (1999). Visual marking of locations and feature maps: Evidence from within-dimension defined conjunctions. *Quarterly Journal of Experimental Psychology A, 52A*(3), 679-715.

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. (2001). 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): I. Tracking without keeping track of object identities. *Visual Cognition*.

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

Ro, T., & Rafal, R. D. (1999). Components of reflexive visual orienting to moving objects. *Perception & Psychophysics, 61*(5), 826-836.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80*(1/2), 1-46.

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

Sears, C. (1991). *Information processing at multiple locations in the visual field.* Unpublished MA, University of Western Ontario, London, Canada.

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

Theeuwes, J., Kramer, A. F., & Atchley, P. (1998). Visual marking of old objects. *Psychonomic Bulletin and Review, 5*(1), 130-134.

Tipper, S. P., Driver, J., & Weaver, B. (1991). Object-centred inhibition of return of visual attention. *Quarterly Journal of Experimental Psychology A, 2,* 289-298.

Ullman, S. (1976). Relaxation and constrained optimization by local processes. *Computer Graphics and Image Processing, 10,* 115-125.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

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

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45*, 205-217.

Watson, D. G. (2001). Visual marking in moving displays: Feature-based inhibition is not necessary. *Perception & Psychophysics, 63*(1), 74-84.

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.

Watson, D. G., & Humphreys, G. W. (1998a). Visual marking of moving objects: A role for top-down feature-based inhibition in selection. *Journal of Experimental Psychology: Human Perception & Performance, 24*(3), 946-962.

Watson, D. G., & Humphreys, G. W. (1998b). Visual marking of moving objects: A role for top-down feature-based inhibition in selection. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 946-962.

Watson, D. G., & Humphreys, G. W. (2000). Visual marking: Evidence for inhibition using a probe-dot detection paradigm. *Perception & Psychophysics, 62*(3), 471-481.

Watson, D. G., Humphreys, G. W., & Olivers, C. N. L. (2003). Visual marking: using time in visual selection. *Trends in Cognitive Sciences, 7*(4), 180-186.