Tracking without keeping track: Some puzzling findings concerning multiple object tracking

Zenon W. Pylyshyn
Rutgers Center for Cognitive Science

Abstract

The task of visually tracking a subset (of about 4) 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 requires continually solving the correspondence problem for each target over time, and thus that an observer must keep track of each target as an enduring individual. This suggests that the identity of successfully tracked objects should be recalled easily. Yet in the present studies we show that observers are poor at recalling the identity of targets (as specified by a unique identifier associated with each target, such as a number or starting location). This discrepancy can be explained if the tendency to mistake a target for another target is greater that the tendency to mistake a target for a nontarget. We present evidence that this may indeed be the case: the identity of object pairs tends to be switched when they come close together and this tendency appears to be greater for target-target pairs than for target-nontarget pairs.

Background: Multiple Object Tracking

A typical MOT experiment is illustrated in Figure 1. A set of 8 to 10 simple objects is displayed on a screen. A number of these objects (about 4) are briefly made visually distinct, typically by flashing them a few times. Then, with all the objects identical in appearance, they move randomly and independently on a screen. 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 (about 10 seconds in our studies) the motion stops and observers are required to indicate which objects were the designated targets.

Figure 1: A typical MOT experiment, in which the subject indicates which items are targets at the end of the trial by moving a cursor to each target (shadow circles indicate items being flashed).
This paradigm has provided a number of interesting and often counter-intuitive findings concerning object-based visual attention (Blaser, Pylyshyn & Holcombe, 2000; Cavanagh, 1999; Culham, Brandt, Cavanagh, Kanwisher, Dale, & Tootell, 1998; He, Cavanagh & Intriligator, 1997; Intriligator & Cavanagh, 2001; 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; Viswanathan & Mingolla, submitted; Yantis, 1992). Perhaps the most surprising finding is the simple fact that observers are able easily to track 4 or even 5 such objects with an accuracy of over 80% using only the individual targets’ enduring identity. Observers can do this even when the movement parameters (speed, distances) precludes their doing so by scanning their attention sequentially to each item and updating its location in memory (as shown by Pylyshyn & Storm, 1988). Despite the simplicity of the paradigm there are a number of puzzles raised by this task that could usefully be clarified, both by a more formal analysis of the task and by further empirical explorations.

The theoretical framework that we adopted to explain the capacity to track visually-identical objects in MOT, is called Visual Indexing or FINST Theory. Because we believe that objects are indexed as individuals, rather than as occupants of certain locations or in virtue of their possessing certain properties, we view visual indexes as a form of direct preconceptual or “demonstrative” reference (Pylyshyn, 2000, 2001a, 2001b). The purpose of the present paper is not to argue in favor of this view, but rather to examine some of the logical requisites of MOT and their empirical entailments, and to report on some experiments designed to cast light on the nature of the tracking process.

The logic of Multiple Object Tracking

At the start of an MOT trial, \( n \) objects out of \( m \) are made visually distinct and then all \( m \) objects move in an unpredictable manner. After some time the motion stops and the observer reports which of the now-identical objects had earlier been singled out as “targets.” The critical aspect of this 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 provinence, which is traced back to the start of the trial. Technically, targethood is defined by historical continuity: a particular object is a target if it was a member of the set of targets in the immediately preceding instant. But how can one tell whether an object had been a member of the set of targets in the preceding instant?

In general, the only way to determine that a particular individual object belonged to the target set in the previous instant is by determining which particular item in the target set it had been, which leads to the well-known correspondence problem in vision [1] (Dawson & Pylyshyn, 1988; Ullman, 1980). An equivalent way to put the requirement of tracking objects’ individuality, which we call the Discrete Reference Principle (DRP), is as the claim that in effect a unique mental identifier or visual index \( V_i \) is assigned to each individual target during tracking. But if a distinct index is assigned to each successfully tracked target, it should be possible to associate a distinct overt label \( L_i \) with each target and subsequently to recall this experimenter-assigned label (which we will call the target’s “ID”). All that an observer needs to do in order to respond with the correct ID for each target is to learn a list of pairs \( V_i-L_i \) and to give the associated \( L_i \) for each tracked \( V_i \) at the end of the trial. Thus if a target is successfully tracked, an observer should be able to report its ID simply as a consequence of having tracked it, so long as the observer can recall a list of four paired-associates. This leads to the prediction that successfully tracked objects should be identifiable by their ID label. Experiment 1 was designed to test this prediction.
Experiment 1

Materials. In this MOT study observers were required not only to pick out the set of targets (referred to as the “tracking” performance measure), but also to identify each target by its ID (i.e., its distinct individual label). The stimuli were white rings with blue interiors that subtended 2.5 degrees of visual angle. Viswanathan & Mingolla (1998; submitted) have shown that such objects, which provided T-junction occlusion cues when the objects self-occluded, can easily be tracked.

Design. The duration of each trial was randomly chosen to be 2 s, 5 s or 10 s. During the initial two-second target-identification phase of each trial, target objects were flashed on and off several times and were also given an 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 flashing circles. In the Location condition, each target appeared in one of the four corner of the screen. At the end of the trial subjects had to move a cursor to the target objects and press a key corresponding to the object’s ID. The name condition and the location condition were run on different groups of subjects.

In addition to the main task (call it the Combined task) in which observers had to provide both ID and Tracking responses, subjects also took part in two baseline tasks: a static ID-only memory-control task that was the same as the Combined task but in which the targets did not move, and a Track-only task that was the same as the main Combined task, but did not require recalling the ID of the targets. These two baseline tasks were presented first, thus providing subjects with extra practice in both tasks and also providing a conservative baseline measure for each task alone. There were 10 subjects in the location condition and 9 subjects in the name condition (one was lost due to equipment failure).

Results

(a) Comparison of ID and Tracking performance. A mixed within- and between-subjects analysis of variance revealed that the overall effect of task-type (tracking vs. ID) in the Combined task 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 the solid lines in Figure 2, ID performance deteriorated much more rapidly than tracking performance as trial duration increased, reaching a value of less than 30% after 10 seconds of tracking.
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 target tracking.

(b) Baseline performance. In the baseline ID-only task, the ability to report the numerical names of the targets was nearly perfect (over 93%, at all durations) was significantly higher than tracking performance and decayed little with increasing trial duration, compared with tracking performance (not surprising since all subjects had to do was stare at the 8 circles and report the labels that had been in 4 of them at the start of the trial). This is confirmed by a within-subjects analysis of variance, 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) and of task x duration interaction (F=19.3, df=2,16; p < 0.001). Thus it does not appear to be the case that tracking by itself is easier than reporting the names of objects.

(c) ID performance on correctly tracked trials. Since tracking performance places an upper bound on ID performance (it is not possible to ID more objects than one has tracked), the poorer ID performance for longer trial durations could be an artifact of the dependence of the two measures. To control for this possibility, we analyzed ID performance on only those trials in which all targets had been successfully tracked. The resulting ID performance, shown in the dashed lines in Figure 2, exhibits the same significant drop with increasing trial length as was observed with the overall ID score. The dashed line results in Figure 2 are based on different groups of subjects so the data were analyzed using a mixed between and within subjects analysis of variance. The effect of trial duration (a within-subjects effect) was significant (F=54.3;
df=2,15; p < .001) but the difference between the two ID tasks (a between subjects effect) was not significant (F=0.02, df=1,16, p>.90), nor was the interaction of trial duration and task (F=4.0, df=1,16, p>.05). Although, as expected, ID performance on the perfectly tracked trials was slightly higher than the mean ID performance, it was still below 50% at the 5 s and 10 s trial durations. Consequently it appears that the rapid decline of ID performance was not a result of the ceiling imposed by tracking performance.

Experiment 2 (a second baseline study)

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 interferes with memory for the paired associates $V_iL_i$, which is a component of the ID task. To examine this possibility, we measured both ID and tracking performance under conditions in which the two tasks could interfere with each other, but in which the ID task was independent of the specific items being tracked. Experiment 2 was designed to measure recall performance on the ID task when the same intervening time was occupied by an independent tracking task, called the “embedded tracking task,” which involved tracking of different items. The study was simply a static recall experiment with an interspersed tracking task identical to the “names” condition of experiment 1. Eleven observers took part in this experiment for pay.

Results. Both tracking and ID performance remained high (above 80%), and the rate of decay of ID performance was even less pronounced than that of tracking performance. There was no statistical difference between the ID performance and tracking performance (F=0.06, df=1,10, p > .80) and overall performance decreased with trial duration (F=17.0, df=2,20, p < .001). However, a post-hoc analysis showed that this was due entirely to a drop in tracking performance: ID performance did not change significantly with trial duration (F=2.06, df=2,20, p>.15), whereas the tracking performance decreased significantly (F=21.7, df=2,20, p<.001), which is the opposite of what we observed in Experiment 1. Thus there is no support for the hypothesis that the rapid decay of ID performance was due to interference from the tracking task itself.

Discussion of results so far and motivation for experiment 3

These experiments leave us with a puzzle. We saw earlier that correct tracking assumes the discrete reference principle and this principle, together with the ability to recall the correspondence between given names and internal references, implies accurate ID recall for tracked targets. Yet we found that performance in maintaining the identity of individual tracked objects is poorer and decays more rapidly than the corresponding performance in tracking the objects, despite the fact that recall of the relevant paired associates remains nearly perfect under the conditions of this experiment.

However, tracking is not perfect and different types of errors have different effects on ID and on tracking performance. Confusing one individual object with another object does not show up as an error in tracking if the objects involved in this exchange are both targets. By contrast, switching the identity of a target with that of a nontarget does show up as a tracking error. So it is relevant to determine 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 (T-T) confusions more frequently than target-nontarget (T-N) confusions, since this tendency could account for the different rates of decay of the two tasks that we have observed.
Experiment 3

The purpose of this experiment was to establish whether there was greater interaction among switched target-target pairs than switched target-nontarget pairs. We defined two distinct interaction scores between pairs of objects. The duration score was defined as the number of frames in each trial that the two objects spent within a predefined distance of one another (3.1 degrees of visual angle or slightly over one diameter). The second interaction score took cognizance of the suggestion made by (He et al., 1997) that the minimum distance in any trial might be the more relevant measure of the degree interaction.

In order to explore the possibility that more switches occurred when two targets passed close to one another compared with when a target and nontarget passed close, we had to confine our analysis to pairs of objects that we could unambiguously determine had been switched. To do this we had to select trials differently for the target-target (T-T) and the target-nontarget (T-N) pairs. For the T-T pairs, we selected only trials in which all targets had been correctly tracked but in which two objects had switched IDs. For the T-N pairs we selected trials in which exactly one tracking error occurred, because we could then use the trajectory record to unambiguously determine which particular target failed to be tracked and which particular nontarget replaced it in the response (subjects always had to choose 4 objects in their response). Because of the very different selection criteria that had to be used in the T-T and the T-N cases, these two cases were not compared with one another in the same analysis. Instead, we carried out two separate analyses, one for T-T pairs and one for T-N pairs. In each case we compared the interaction score for a switched pair with the interaction score of an unswitched pair selected at random from the same trial. While this does not permit us to test the strong hypothesis that when T-T pairs came close together they have a greater tendency to switch than when T-N pairs came close together, it did allow us to test whether there was an asymmetry between the two cases insofar as the relation between interaction scores and switching probability differed in T-T cases compared with T-N cases. A difference between these tendencies would at least be compatible with the hypothesis that T-T pairs behave differently from T-N pairs.

Materials and method. Experiment 3 was essentially a replication of the name condition of the combined ID & Track tasks of experiment 1, while recording the entire trajectory of each object in the experiment so that the degree of interaction of pairs of objects could be computed. Because the number of name switches and tracking errors tended to be very low for short duration trials, only the 10-second trials were analyzed. In addition, in order to have stable data, we only analyzed results from subjects who met two conditions imposed for reasons discussed in the previous paragraph: (a) they had at least two trials with exactly one tracking error and (b) they had at least two trials with perfect tracking and with an ID switch. Of the twenty-seven subjects who served in this experiment only sixteen met these conditions and provided data for the final analysis.

Results

(1) Effect of target-target interaction on ID errors. The mean duration score between targets whose ID was switched was 43.1 frames (s =17.1) whereas the score between a randomly chosen pair of correctly identified targets was 62.8 (s = 20.3). A repeated measures t-test showed that this score was significantly higher for switched targets than unswitched targets (t=2.75, df=15, p < .01). Similarly, the minimum distance score for targets whose ID was switched was 37.4 pixels (s =16.3) whereas the score for a randomly chosen pair of correctly identified targets was 19.9 (s =
A repeated measures t-test showed that this difference was also statistically significant (t=3.40, df=15, p < .005).

(2) Effect of target-nontarget interaction on tracking errors. The mean duration score for misclassified T-N pairs was 49.7 (s = 22.5) while for randomly chosen correctly classified T-N pairs the score was 62.9 (s = 20.9). A repeated-measures t-test showed that this difference did not approach significance (t=1.68, df=15, p > .11). A similar result was found using the minimum distance score. The mean minimum distance score for misclassified T-N pairs was 33.1 (s = 23.0) while for randomly chosen correctly classified T-N pairs it was 17.6 (s = 15.0). A repeated-measures t-test showed that this difference again did not approach significance (t=1.94, df=15, p > .07).

Thus there is a significant tendency for switched target pairs to be associated with “close encounters” (i.e., higher interaction scores) compared with their unswitched controls, whereas there is no such tendency for target-nontarget pairs to be associated with higher interaction scores. Such an asymmetry may explain why, despite the discrete-reference principle that is required by the tracking task, we find that IDs are more poorly retained than is the target-nontarget distinction that leads to correct tracking performance. These results also provide further support for the view (proposed by Intriligator & Cavanagh, 2001) that errors in tracking are due to the failure of attentional resolution.

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 assigned names or their starting locations). Although this appears to be inconsistent with the need to keep track of each target as a distinct individual while tracking, it might be explained by the further hypothesis that target-target pairs may be more readily confused when they pass close to one another than are target-nontarget pairs. Experiment 3 provides some initial indirect support for that hypothesis by showing an asymmetry between the interaction (or proximity) scores for switched T-T cases relative to their unswitched controls compared to switched T-N cases relative to their unswitched controls.

Although an asymmetry between T-T confusions and T-N confusions associated with close encounters may account for the divergence between tracking and ID performance, we still do not know what mechanism is responsible for this asymmetry. One possibility that we are currently investigating involves the hypothesis of nontarget inhibition. Using a search task, (Watson & Humphreys, 1997) provided some evidence that when a subset of items is attentionally selected, the unselected items may actually be inhibited. If this were true in the MOT task, we would expect targets to be more often confused with other targets than with nontargets because the latter are inhibited.

References


NOTES

1 Under special conditions it might be possible to determine that a particular individual object had been a member of the target set without first solving the correspondence problem, but this will not work in the general case when objects are identical and move in unpredictable ways. Among the possible special circumstances would be cases where the set of objects was visibly distinguishable by its color or by its location as a group. In such cases, an object might be correctly classed as a target by virtue of its recognizable group membership, rather than by solving the correspondence problem for the individual object.