Distinguishing between parallel and serial accounts of multiple object tracking

Piers D. L. Howe   Brigham and Women's Hospital, Boston, MA, USA, & Harvard Medical School, Boston, MA, USA

Michael A. Cohen   Harvard University, Boston, MA, USA

Yair Pinto   Brigham and Women's Hospital, Boston, MA, USA, & Harvard Medical School, Boston, MA, USA

Todd S. Horowitz   Brigham and Women's Hospital, Boston, MA, USA, & Harvard Medical School, Boston, MA, USA

Humans can track multiple moving objects. Is this accomplished by attending to all the objects at the same time or do we attend to each object in turn? We addressed this question using a novel application of the classic simultaneous-sequential paradigm. We considered a display in which objects moved for only part of the time. In one condition, the objects moved sequentially, whereas in the other condition they all moved and paused simultaneously. A parallel model would predict that the targets are tracked independently, so the tracking of one target should not be influenced by the movement of another target. Thus, one would expect equal performance in the two conditions. Conversely, a simple serial account of object tracking would predict that an observer's accuracy should be greater in the sequential condition because in that condition, at any one time, fewer targets are moving and thus need to be attended. In fact, in our experiments we observed performance in the simultaneous condition to be equal to or greater than the performance in the sequential condition. This occurred regardless of the number of targets or how the targets were positioned in the visual field. These results are more directly in line with a parallel account of multiple object tracking.

Keywords: attention, motion-2D, computational modeling

Introduction

Attention plays a central role in visual cognition and has been a primary focus of research in cognitive psychology (Pashler, 1999). Recently, several studies have provided evidence that attention can be simultaneously directed to multiple locations (Awh & Pashler, 2000; McMains & Somers, 2004; Scharlau, 2004). While these are exciting and innovative reports, one might be forgiven for finding these discoveries unsurprising, since over two decades of research using the multiple object tracking paradigm (MOT, Pylyshyn & Storm, 1988) has demonstrated that attention can be simultaneously directed to multiple moving objects. Or has it? While the basic finding, that humans can track several independently moving objects, has been replicated many times (for reviews, see Cavanagh & Alvarez, 2005; Scholl, 2001), the assumption that this proves *simultaneous* attention to multiple objects has never been directly tested.

In a traditional MOT experiment, a trial might start with the presentation of ten identical black disks, a subset of which would then briefly turn red to indicate that they are the targets to be tracked. The targets would then revert to their original color and all the disks would move around the display in a random fashion. At the end of the trial, the observer would be asked to indicate which of the disks were the targets (Figure 1).

Depending on the exact experimental conditions, most observers are able to keep track of the locations of approximately four targets (Cavanagh & Alvarez, 2005; Pylyshyn & Storm, 1988). It could be that observers track each of the four targets simultaneously (Alvarez & Franconeri, 2007; Franconeri, Jonathan, & Scimeca, 2009; Kazanovich & Borisyuk, 2006; Pylyshyn, 2001; Scholl, in press; Yantis, 1992). Alternatively, tracking might be achieved by a serial process (d'Avossa, Shulman, Snyder, & Corbetta, 2006; Oksama & Hyona, 2008) or at least contain a significant serial component (Landry & Sheridan, 2001; Oksama & Hyona, 2004; Zelinsky & Neider, 2008).

In a basic serial model, an observer attends to each target in turn so that at any one time only a single target is attended. Every time a target is attended, the observer
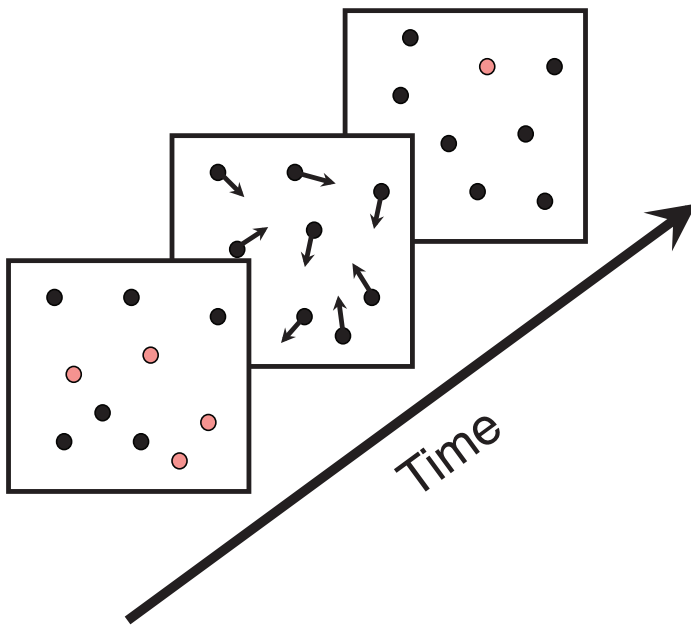
Figure 1. A typical multiple object tracking (MOT) trial. At the start of the trial four disks turn red to indicate that they are targets to be tracked. The disks then all become the same color and move about the display for several seconds. At end of the trial the disks stop moving and the observer indicates which were the targets.

notes its current location. When it is time to reattend a particular target, the observer assumes that whichever object is closest to that target's previously remembered position is the target. Mistakes occur when the target has moved so far since it was last attended that it is no longer the closest object to its previously remembered position. The quicker attention can be shifted between targets, the quicker any given target can be reattended, and the less likely mistakes are to occur.

In the past, it has proven difficult to evaluate the plausibility of serial accounts of MOT (Alvarez & Franconeri, 2007; Cavanagh & Alvarez, 2005). For example, Pylyshyn and Storm (1988) and Yantis (1992) both constructed serial models of MOT to explain their data and concluded that the scanning rate needed for successful tracking was implausibly high. Yantis (1992) claimed that his serial model would have required a spotlight of attention to scan between targets at a rate of at least 150°/s. However, if we reinterpret this scanning rate as a switching time, 150°/s becomes equivalent to switching attention between targets every 33 ms. This is at the lower end of estimates of attentional switching times, but not implausibly low (Egeth & Yantis, 1997; Horowitz, Wolfe, Alvarez, Cohen, & Kuzmova, 2009; Saarinen & Julesz, 1991; Wolfe, Alvarez, & Horowitz, 2000). In fact, Hogendoorn, Carlson, and Verstraten (2007) have provided evidence for such fast switching times in the context of attentional tracking. Given fast switching, a serial model could also explain the Pylyshyn and Storm (1988) data, especially if there were two

independent serial mechanisms, as described below. On average, each mechanism would need to track only half the targets, making its task that much easier.

The only previous empirical evidence against a serial account of MOT is by Alvarez and Cavanagh (2005), who demonstrated independent capacity limits for tracking in the left and right visual hemifields, such that observers could track twice as many objects when they were distributed across two hemifields as opposed to when they were confined to a single hemifield. This finding effectively rules out the possibility of a serial model with a single focus of attention, but allows for one with two foci of attention, each responsible for its own hemifield (cf. Luck, Hillyard, Mangun, & Gazzaniga, 1989). Thus, within each hemifield, targets might still be tracked by a serial mechanism. If true, this would necessitate a radical reevaluation of existing theories of MOT (Oksama & Hyona, 2008; Pylyshyn & Storm, 1988; Scholl, in press; Yantis, 1992).

In this paper, we address the serial versus parallel question by adapting the classic simultaneous-sequential paradigm (Eriksen & Spencer, 1969; Shiffrin & Gardner, 1972). In Experiment 1, we considered an MOT display that always contained two dot quartets, each quartet centered in its own quadrant (Figure 2). At the start of each trial, one member of each quartet would briefly turn red to indicate that it was a target to be tracked. Each quartet would then rotate about its own center point, 90 degrees at a time. There would be a pause between successive rotations so that the dots were moving only half the time. At the end of the trial, one of the quartets would disappear and the observer would use the mouse to indicate which of the four remaining dots was a target.

In the *simultaneous* condition, the two quartets rotated and paused synchronously. In the *sequential* condition, the two quartets rotated and paused asynchronously so that, at any one time, only one quartet was moving. Crossed with this manipulation, the two targets could either be in the same hemifield (either left or right) or one would be in the left hemifield while the other would be in the right hemifield. Crucially, the stimulus configuration was visually identical at the start and end of each quarter-turn rotation. Thus, if a target was not attended while it was moving, most likely it would be lost.

To illustrate the utility of this paradigm, let us consider two simple models of object tracking: a parallel model that assumes that objects are always tracked independently and in parallel regardless of where they are located, and a serial model that assumes that, when there are two objects each in a different hemifield, the objects are tracked independently but, when the two objects are located within the same hemifield, they are attended sequentially in that only one object is attended at any one time. In the case of the serial model, we further assume that attention is focused preferentially onto the moving target. That is, in the sequential condition, we assume that attention exploits the fact that only half of the
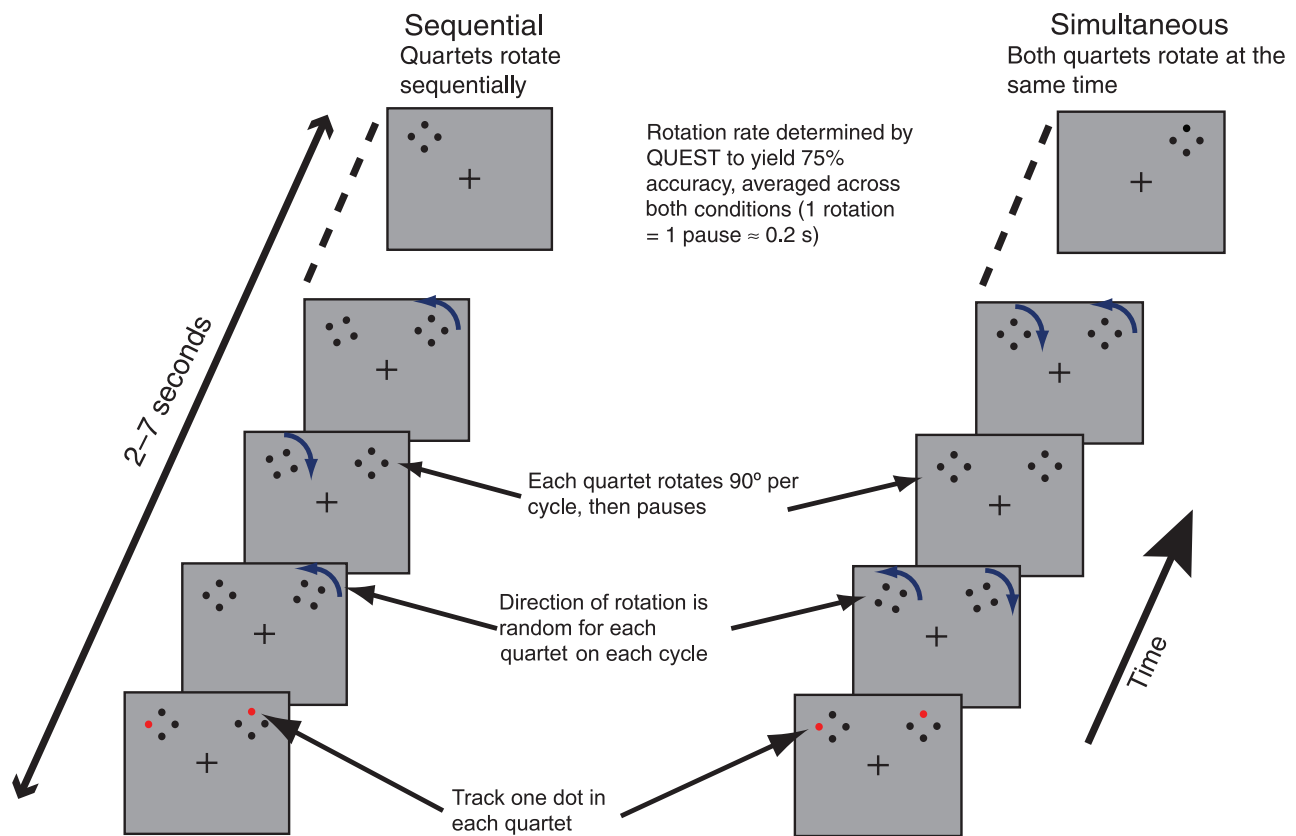
Figure 2. A cartoon of the method for Experiment 1. The displays always contained two quartets, each centered in its own quadrant. Here the quartets occupy the upper two quadrants. In another trial they might occupy the lower two quadrants, the left two quadrants or the right two quadrants. At the start of each trial, one member of each quartet would briefly turn red to indicate that it was a target to be tracked. Each quartet would then rotate about its own center point, 90° at a time. There would be a pause between successive rotations so that the dots were moving only half the time. At the end of the trial, one of the quartets would disappear and the observer would use the mouse to indicate which of the four remaining dots was a target. In the sequential condition, the quartets rotated asynchronously. In the simultaneous condition, they rotated synchronously so that at any one time only one quartet was rotating.

targets are moving and that only the moving targets are changing location, so need to be attended. We make this assumption for three reasons. First, there is evidence that motion onsets preferentially direct attention to moving targets (Abrams & Christ, 2003). Second, there is both fMRI (Howe, Horowitz, Morocz, Wolfe, & Livingstone, 2009) and EEG (Drew, Horowitz, Wolfe, & Vogel, submitted for publication) evidence that moving targets are attended more than stationary ones during MOT. Third, a previous MOT study has shown that observers preferentially attend to those targets that are in danger of being lost (Iordanescu, Grabowecky, & Suzuki, 2009), which in our study would be the moving targets.

We start by considering the case where the two targets are in different hemifields. Here, both models predict that the two targets should be tracked independently. Consequently, both models predict that the motion of one target should not affect how easy it is to track the other target, so simultaneous and sequential tracking performance should be equal.

Now we consider the case where the two targets are in the same hemifield. The parallel model continues to predict that the two targets are tracked independently and that tracking accuracy should continue to be equal in the two conditions. However, the serial model now assumes that the two targets are tracked in series. Thus, only one target is tracked at any one time. In the sequential condition, only one target moves at any one time, so only one target needs to be tracked. However, in the simultaneous condition, both targets move at the same time, so both targets would need to attended within each moving phase, making tracking more difficult. Consequently, the serial model predicts worse performance in the simultaneous condition.

To preview our results, in Experiment 1 we found tracking accuracy to be slightly greater in the simultaneous condition, regardless of whether the targets were located within the same or different hemifields. Nor was there any interaction between spatial arrangement and the difference in accuracy between the simultaneous and

sequential conditions. Both these findings argue against the serial model, and are more directly in line with the predictions of the parallel model. However, while the parallel model would correctly predict the lack of interaction, it would not predict the observed simultaneous advantage.

In Experiments 2, 3, 4, 5 and 7–8, we varied the number of targets and the locations of the targets. In all these experiments, performance in the simultaneous condition was at least as good as the performance in sequential condition. Experiment 6 was a control experiment designed to provide a test of the validity of the paradigm. Essentially, we biased the observers to track two targets in series and consequently observed tracking accuracy to be greatest in the sequential condition. Had a sequential advantage not been observed in this experiment, we would have been forced to question the validity of our experimental procedures.

Experiment 1

We adapted our displays from Alvarez and Cavanagh (2005). There were two groups of disks, each located in the center of its own quadrant (Figure 2). Each group comprised a quartet of disks. The two quartets were arranged so that they were both within the same hemifield (both quartets to the left or right of fixation) or distributed across both hemifields (both quartets above or below fixation). There could be only one quartet in a quadrant and the two quartets could not be located diagonally opposite each other. The observer's task was to track one disk in each quartet. In the simultaneous condition, both quartets rotated (and paused) simultaneously. In the sequential condition, one quartet rotated and then paused while the other quartet rotated.

As explained above, both the parallel and serial hypotheses assume that when the two targets are in different hemifields they should be tracked in parallel, so performance should be equal in the two conditions. The parallel hypothesis also makes this prediction when both targets are within the same hemifield. However, the serial hypothesis assumes that two targets in the same hemifield are tracked in series, and thus predicts a sequential advantage.

Method
Participants

There were twelve observers (age range 18–50). These observers had either normal or corrected-to-normal visual acuity, and none were colorblind. All observers provided informed consent, approved by the Brigham and Women's Hospital Institutional Review Board.

Apparatus and stimuli

Stimuli were presented on a 21-inch Mitsubishi Diamond Pro monitor at a resolution of 1280 × 960 and a refresh rate of 75 Hz, using a Macintosh PowerPc G4 and the Psychophysics toolbox (version 3) for MATLAB (Brainard, 1997; Pelli, 1997). The stimulus is cartooned in Figure 2. At the center of the display was a fixation cross. There were two quartets of black disks. Each quartet was centered in its own quadrant. The luminance of the background was 67 cd/m^2 and the luminance of the disks was <0.5 cd/m^2. Each disk subtended 0.4 degrees of visual angle. The diameter of each quartet was 3° and the center of each quartet was 6° from the fixation cross.

Procedure

Observers were instructed to fixate the central cross. At the start of the trial, one disk from each quartet would turn red for 1.5 seconds to indicate that it was a target to be tracked. Each quartet would then repeatedly rotate around its center point 90° at a time, with a pause between rotations so that each quartet rotated for only half of the time. The directions of successive rotations were random. In the sequential condition, the two quartets rotated asynchronously so that at any one time only one quartet of disks was rotating. The simultaneous condition was identical to the sequential condition except that both quartets rotated at the same time and paused at the same time. Each trial lasted a random duration between 2 and 7 seconds. At the end of the trial, one quartet disappeared and the observer was asked to indicate with a mouse which of the remaining four disks was a target. If the observer made a mistake, the observer would hear a tone and the word "Incorrect" would be displayed for 1.5 seconds. This feedback was given throughout the experiment.

The experiment began with ten practice trials. We then ran a block of 40 trials using the QUEST routine (King-Smith, Grigsby, Vingrys, Benes, & Supowit, 1994; Watson & Pelli, 1983) to find, *for each observer,* the rotation speed that would result in 75% accuracy averaged over both conditions and all four arrangements. This rotation speed was used in the third part of the experiment, in which the observer accuracy was measured for the different arrangements and conditions. 320 trials were run, equally divided between arrangements and conditions in a randomly interleaved fashion. The rotation speed determined the duration of one cycle (i.e. the time take for a quarter-turn rotation and a pause equal in duration to a quarter-turn rotation). Thus, the cycle duration was determined by the observer's performance. Averaging across observers, the cycle duration was 0.39 seconds (*SE* = 0.02).

Results and discussion

The results for this and the other experiments are shown in Figure 3. Consistent with Alvarez and Cavanagh
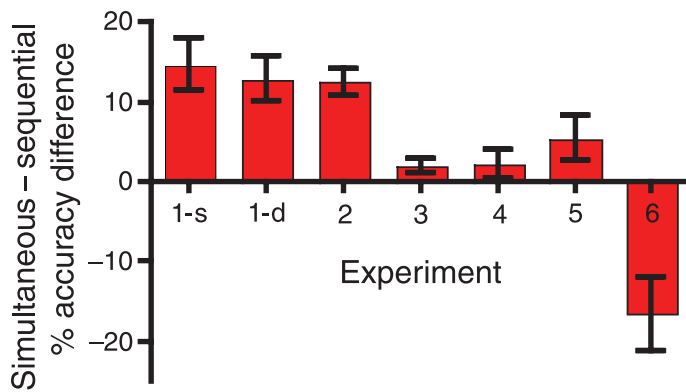
Figure 3. Results for Experiments 1, 2, 3, 4, 5, and 6. Note that 1-s (i.e. 1-same) refers to those trials of Experiment 1 where the two quartets were in the *same* visual hemifield (left or right), while 1-d refers to those trials of Experiment 1 where the two quartets were in *different* hemifields. The ordinate represents the accuracy difference between the simultaneous and sequential conditions. Error bars represent the standard error of the mean.

(2005), a two-way within-subjects ANOVA found accuracy to be greater in the different hemifields arrangements than in the same hemifields arrangements ($F(1,11) = 9.71$, $p = 0.01$, partial $\eta^2 = 0.469$). Inconsistent with a serial model, we found that accuracy was greater in the simultaneous condition than in the sequential condition ($F(1,11) = 24.07$, $p < 0.001$, partial $\eta^2 = 0.686$) and there was no interaction between spatial arrangement (i.e. same hemifield versus different hemifields) and temporal condition (i.e. simultaneous versus sequential; ($F(1,11) = 0.79$, $p = 0.393$, partial $\eta^2 = 0.067$).

The last finding is perhaps the most difficult for the serial model to explain. Even if it could explain why within-hemifield accuracy is not greater in the sequential condition, it still must predict that tracking in the different hemifields case to be fundamentally different than tracking in the same hemifield case, since the former must be parallel process (Alvarez & Cavanagh, 2005) while it assumes that the latter is a serial process.

Note that this is unlikely to be simply a case of our experiment lacking the statistical power to find the interaction. Not only was the effect size for the interaction quite small (partial $\eta^2 = 0.067$) in comparison to the main effects for spatial arrangement (partial $\eta^2 = 0.469$) and for temporal condition (partial $\eta^2 = 0.686$), but more importantly, the trend is in the opposite direction. The advantage for simultaneous motion is actually slightly greater when the stimuli are in the same hemifield.

Experiment 2

Experiment 1 considered only situations where there were just two targets. However, this is a low tracking load.

It is commonly reported that observers can track approximately four targets (Pylyshyn & Storm, 1988), or more if they move slowly enough (Alvarez & Franconeri, 2007). In Experiment 2, we asked whether increasing tracking load to four targets would force observers to adopt a serial strategy.

Method

The methods and stimuli were similar to the previous experiment except that in each display there were four quartets, each located in its own quadrant. The age range of observers was the same as in Experiment 1 and seven of the observers had participated in the previous experiment. As before, the center of each quartet was located $6°$ from the fixation cross and the quartets had a diameter of $3°$. In the simultaneous condition all quartets rotated and paused at the same time, whereas in the sequential condition diagonally opposite quartets rotated synchronously with each other but sequential with respect to the other two quartets. This meant that at any one time only two quartets were rotating. Having diagonally opposite quartets rotate together during the sequential condition meant that the two quartets within a single hemifield were rotating asynchronously in this condition. For each observer, the QUEST routine was used to find the rotation speed that resulted in 75% accuracy, averaged across condition. On average, one cycle lasted on 0.29 seconds ($SE = 0.01$). Using this speed, 300 trials were run, equally distributed between the two conditions in a randomly interleaved fashion.

Results and discussion

A 1-way within-subjects ANOVA found observers were more accurate in the simultaneous condition ($F(1,11) = 54.09$, $p < 0.001$, partial $\eta^2 = 0.831$). It was easier to track the targets when all the targets moved synchronously.

Experiment 3

The previous two experiments showed that, even within a hemifield, two targets could be tracked in parallel. However, these two targets were always located in a different quadrant. It could be that, instead of the two independent tracking mechanisms, each responsible for a different hemifield, there are actually four tracking mechanisms, each responsible for a different quadrant. Thus, the serial process would become apparent only when there is more than one target in a quadrant. Consistent with this hypothesis, tracking accuracy has been found to increase when two targets occupy different

quadrants as opposed to when they are placed within the same quadrant, all other factors held constant (Carlson, Alvarez, & Cavanagh, 2007). In Experiment 3, we tested this hypothesis by placing two targets in a single quadrant.

Method

In each trial, there were only two quartets and these were always placed within the same quadrant. The quadrant used was randomized between trials. As before, there were two conditions. In one condition, the two quartets rotated synchronously. In the other condition they would rotate asynchronously so that at any one time only one was rotating. The distance of the center of each quartet from the fixation cross was the same as before (6°). The quartet centers were separated by 4.8°. All other aspects of the stimulus and the procedure were the same as in the previous experiment. On average, one cycle (i.e. one quarter-turn rotation and one pause) lasted 0.33 seconds ($SE = 0.02$).

Results

Accuracy was again greater in the simultaneous condition, though this result was only marginally significant ($F(1,11) = 4.5$, $p = 0.057$, partial $\eta^2 = 0.290$). Even when there are two targets within a single quadrant, we can still find no evidence of a serial process.

Experiment 4

A concern with Experiment 3 is that observers may not have been maintaining accurate fixation on the central cross. If instead, the observers had fixated between the two quartets then the two quartets would, in effect, have been located in different quadrants. To control for this possibility, we reran Experiment 3 using a 60 Hz Arrington Research QuickClamp monocular eye tracker. As this reduced the viewing distance, the dimensions of the stimuli changed. The center of each group was now located 7.6° from the fixation cross and 6.1° from the center of the other group. The disks subtended 0.5°. If at any point the observer fixated more than 2° from the fixation point, the trial was aborted and redone. On average, a cycle lasted 0.31 seconds ($SE = 0.01$). All other aspects of this experiment were identical to the previous experiment.

There was no statistical difference between the results of Experiment 3 and Experiment 4 ($F(1,11) = 0.018$, $p = 0.894$, partial $\eta^2 = 0.001$). Since these experiments used the same procedure and stimuli, we combined the two data

sets. Across the two experiments, performance was significantly greater in the simultaneous condition than in the sequential condition ($F(1,23) = 4.622$, $p = 0.042$, partial $\eta^2 = 0.167$).

Experiment 5

According to Alvarez and Cavanagh (2005), objects are tracked independently in the left and right visual hemifields. Each visual hemifield has its own separate mechanism for tracking objects. It follows that, whenever an object crosses the vertical midline, the responsibility for its tracking would need to be transferred from one tracking mechanism to another. For example, when an object crosses the visual midline from left to right, it is initially tracked by the mechanism responsible for tracking objects in the left visual hemifield. After it crosses the vertical midline, it would then be tracked by the mechanism responsible for tracking objects in the right visual hemifield.

Although our above results demonstrate that, within a visual hemifield, objects are tracked in parallel, it could be that transferring tracking responsibility from one visual hemifield to the other occurs via a serial process. For example, as summarized above, Pylyshyn and Storm (1988) have suggested that objects are tracked by mental pointers known as FINSTs. Although these FINSTs are assumed to track the objects in parallel, with each FINST tracking its object independently of the other FINSTs, the initiation of the FINSTs might be serial (Pylyshyn, 1989). In particular, it is assumed that in some circumstances the FINSTs are attached to their respective targets one at a time using top-down attention (Pylyshyn, 1989). As the object crosses the vertical midline from left to right, the FINST responsible for tracking it in the left hemifield must be deleted and a new FINST created to track it in the right hemifield. If the creation of FINSTs is a serial process, then the transfer of tracking responsibility from one hemifield to the other must also be a serial process. If so, then one could not track in parallel two objects that repeatedly cross the vertical midline. In Experiment 5, we tested this hypothesis. The stimulus dimensions and experimental procedure were identical to Experiment 3 except that the two quartets were centered on the vertical midline and each trial lasted 10 seconds. On average, a cycle lasted 0.32 seconds ($SE = 0.01$). One quartet was above the fixation cross, the other below it. This ensured that the targets were continuously crossing the vertical midline thereby maximizing the chances of detecting a serial process, should it exist. 300 trials were run equally distributed among the two conditions. We found that accuracy was again greatest in the simultaneous condition, although the result was only marginally significant

$F(1,11) = 3.71$, $p = 0.08$, partial $\eta^2 = 0.252$. We were therefore unable to find any evidence that tracking is serial, even for this stimulus arrangement.

Experiment 6

At this point, one might wonder whether we could ever observe a sequential advantage in our paradigm. Experiment 6 was designed to address this concern. We modified the experimental arrangement to encourage observers to track objects in a serial fashion. As in Experiment 5, the stimulus comprised two quartets and a fixation cross. The dimensions of the stimuli were the same as in Experiment 5 except that the quartets were now aligned on the horizontal midline, one on each side of the fixation cross. The duration of a single trial was extended to a minimum of 15 seconds.

Before each quartet would rotate for a quarter turn and then pause, now each quartet would rotate continuously for 10 quarter turn rotations, each in a random direction, before pausing. In the sequential case, first one quartet would rotate. Then there would be a pause of 1.5 seconds after which the second quartet would rotate, followed by another pause for 1.5 seconds. This cycle was repeated until a minimum of 15 seconds had elapsed. This procedure encourages serial processing because the targets started and stopped moving in a predictable fashion, and the time from when one target stopped moving until the other target started moving (1.5 seconds) was more than sufficient for observers to transfer their attention from one target to another. Conversely, in the simultaneous condition, each quartet would rotate and pause for the same amount of time as it did in the sequential condition but the quartets would rotate and pause synchronously, thereby making it harder to track the targets using a serial mechanism.

The QUEST routine was used to find, for each subject, the rotation speed that resulted in 75% accuracy in the sequential condition. Across subjects, the average time for a quarter turn rotation that resulted in 75% accuracy was 0.22 seconds. Using this rotation speed, 100 trials were run, equally distributed between the two conditions in a random fashion. Unlike the previous experiments, performance was much greater in the sequential condition than in the simultaneous condition ($F(1,11) = 12.5$, $p = 0.005$, partial $\eta^2 = 0.532$). In agreement with the observers' subjective impressions, this result indicates that observers were tracking the targets in a serial fashion. The fact that the simultaneous-sequential paradigm unambiguously reports that tracking is serial in a circumstance where we would expect tracking to be serial increases our confidence in the paradigm's validity.

Experiment 7

As noted above, Alvarez and Franconeri (2007) suggested that up to eight targets can be tracked, providing they move slowly enough. However, Pylyshyn's FINST theory (Pylyshyn, 1989, 2001) suggests that observers should be able to track only up to 3–5 targets in parallel, because of a structural limit on the number of available visual indexes ("FINSTs"). One way to explain the Alvarez and Franconeri data from the FINST point of view would be to suggest that when the number of targets exceeds the number of indexes, observers adopt a serial strategy of switching indexes between subsets of the target set. Thus, while tracking four targets is accomplished by a parallel process, tracking eight targets may be accomplished by a serial process. Experiment 7 tested this hypothesis.

Whereas the previous experiments used rotating quartets, Experiment 7 used rotating doublets so as to keep the total number of disks down to a manageable number (Figure 4). The luminance of background was now 81 cd/m^2 and that of the disks was <0.5 cd/m^2. Each disk subtended 1° and the center-to-center of the disks in a doublet was 6°. The furthest doublets were centered 16° from the fixation cross, at which point the individual disks were still readily discernable.

As before, each doublet would rotate 90° at a time and then pause. Only one member of a doublet could be a target. There were always eight doublets, but not all doublets always contained a target. Depending on the condition, there could either be two, four, six or eight targets. In the simultaneous condition, all doublets would rotate and pause synchronously. In the sequential condition, the doublets were randomly dived into two groups, with each group containing the same number of targets. The two groups would rotate in alternation. At the end of the trial, the observer was required to click on the location of all the targets. A single error would result in the entire trial being categorized as a "miss" trial. As before, we used the QUEST routine to find the rotation speed that would result in 75% tracking accuracy. Unlike the previous experiments, we did this separately for each condition. Thus, for each observer, we measured eight different rotation speeds (each expressed in terms of radians per second).

Because of the geometry of the display, if a target is not tracked during each 90° rotation then it would likely to be lost. Thus, if the targets were tracked in series, in the simultaneous condition each target would need to be attended during each rotation phase. Conversely, in the sequential condition, only half the targets would need to be attended during each rotation phase. It follows that for the same tracking accuracy, the rotation phase would need to be twice as long in the simultaneous condition than in the sequential condition. Because the duration of the rotation phase is determined by the rotation speed, it follows that for equal tracking accuracy, the rotation
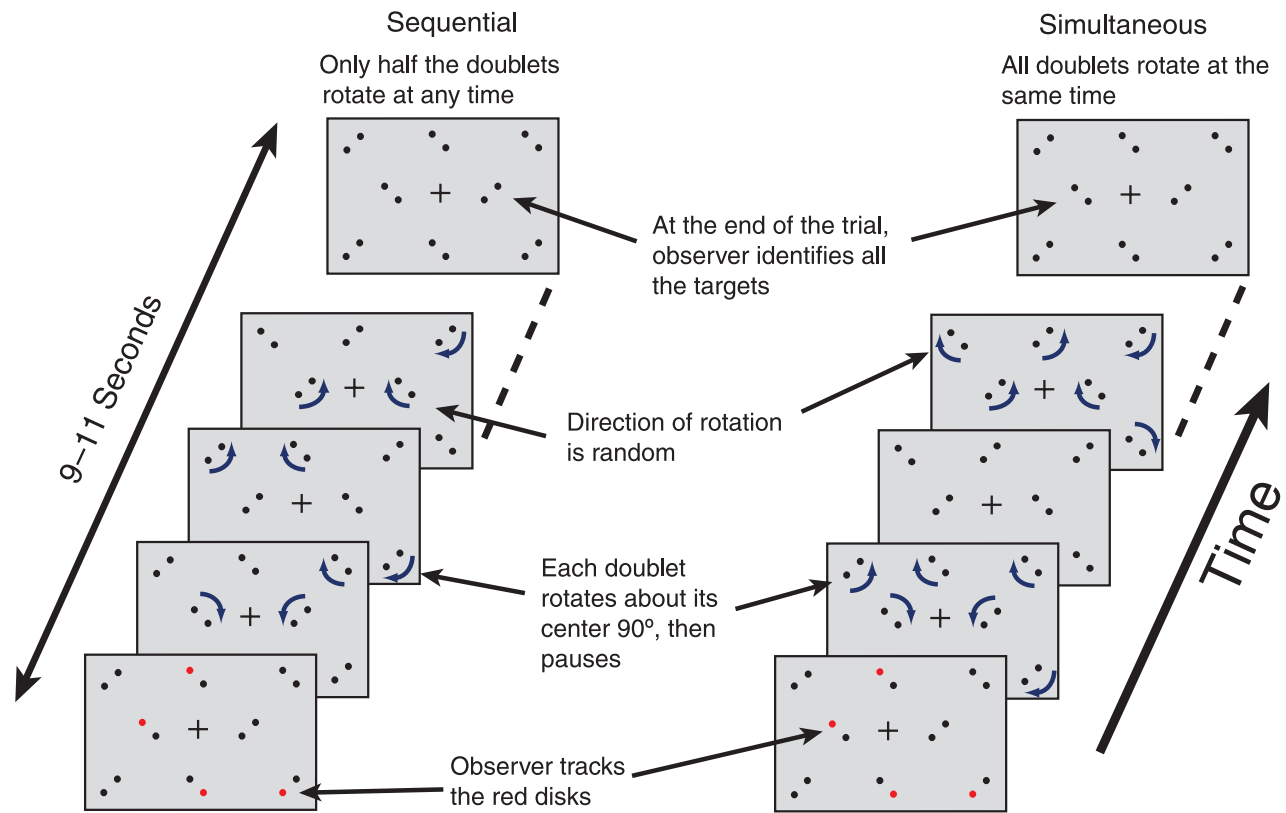
Figure 4. A cartoon of the method for Experiment 7. Each disk is paired with another disk to form a doublet. Each doublet rotates about its center, 90° at a time, with a pause between rotations. In the sequential condition only half the doublets are rotating at any given time whereas in the simultaneous condition all the doublets rotate and pause at the same time.

speed in the simultaneous condition would need to be half that in the sequential condition. Conversely, if the targets are tracked independently and in parallel we would expect that using the same rotation speed in the two conditions would result in equal tracking accuracy.

Because the measurement error was found to increase approximately linearly with rotation speed, we log-transformed our data so that the measurement uncertainty would be approximately equal for all data points. The results are plotted in Figure 5. There was a significant effect of target number ($F(3,33) = 146$, $p < 0.001$, partial $\eta^2 = 0.980$) and stimulus condition ($F(1,11) = 9.30$, $p = 0.01$, partial $\eta^2 = 0.458$), with accuracy being greater in the simultaneous condition. There was no interaction between these two factors ($F(3,33) = 0.662$, $p = 0.596$, partial $\eta^2 = 0.181$). Thus, there is no evidence of serial processing nor is any evidence of a change of strategy from parallel to serial with increasing target number.

tracking rotating stimuli. Rotating stimuli have been used before in MOT studies (e.g., Franconeri, Jonathan, & Scimeca, 2010), and Alvarez and Cavanagh (2005) demonstrated that their findings generalized to independently translating stimuli. Nevertheless, rotary motion might be qualitatively different from the translating motion typically used in MOT studies. Rotary target motion varies in only one dimension,
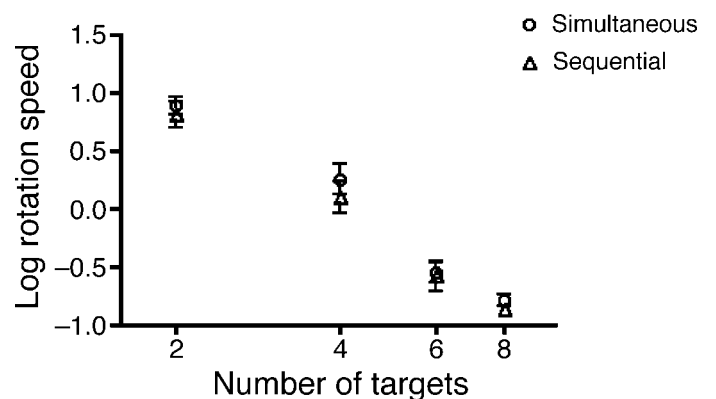

Figure 5. Results of Experiment 7. The rotation speed (measured in radians per second) that resulted in 75% correct accuracy is plotted as a function of the number of targets (both on a base 10 logarithm axis) for the simultaneous and sequential conditions. Error bars represent the standard error of the mean.

Experiment 8

At this point, we have six experiments suggesting that tracking of up to eight rotating stimuli is accomplished in parallel. However, this might be a peculiar property of

rather than two. Furthermore, targets and distractors within a quartet or doublet might be considered to be parts of the same object, whereas in the standard MOT paradigm targets and distractors are clearly independent objects. These factors might conspire to allow parallel tracking in these stimuli, even if "normal" tracking is carried out by a serial mechanism.

With these concerns in mind, in Experiment 8 we adapted the simultaneous-sequential paradigm to more typical MOT displays (Figure 6). There were a total of 16 green disks, each subtending 1°. Targets were briefly designated in red, at the start of the trial, then turned green. Each disk would move in a straight line until it bounced off another disk or the sides of the 40° by 30° display area. The luminance of the background was <0.5 cd/m^2 and the luminance of the disks was 58 cd/m^2.

Either two, four, six, or eight disks were designated as the targets at the start of the trial. As before, each disk would move and pause in alternation. The duration of both the movement and pause phases were set to 0.5 seconds. This duration was chosen to be similar to that used in the four target condition of Experiment 7.

In the simultaneous condition, all the disks would move and pause synchronously. In the sequential condition, the disks were randomly dived into two groups, each group containing the same number of targets. The two groups moved in alternation. At the end, of the trial the observer was required to click on the location of all the targets. A single error would result in the entire trial being categorized as a "miss" trial. We used the QUEST routine to find the disk speed (expressed in degrees per second) that would result in 75% tracking accuracy. As with Experiment 7, we did this separately for each condition. Thus, for each observer, we measured eight different disk speeds.

The results are shown in Figure 7. There was a significant effect of number of targets ($F(3,33) = 125$, $p < 0.001$, partial $\eta^2 = 0.976$), but not of stimulus condition ($F(1,11) = 0.615$, $p = 0.445$, partial $\eta^2 = 0.053$). There was no interaction between these two factors ($F(3,33) = 0.079$, $p = 0.970$, partial $\eta^2 = 0.026$). Again, there was no evidence for serial processing nor is any evidence of a change of strategy from parallel to serial with increasing target number.

General discussion

In this paper, we addressed the question of whether tracking is achieved by a serial or by a parallel process using a novel adaptation of the classic simultaneous-sequential paradigm (Eriksen & Spencer, 1969; Shiffrin &
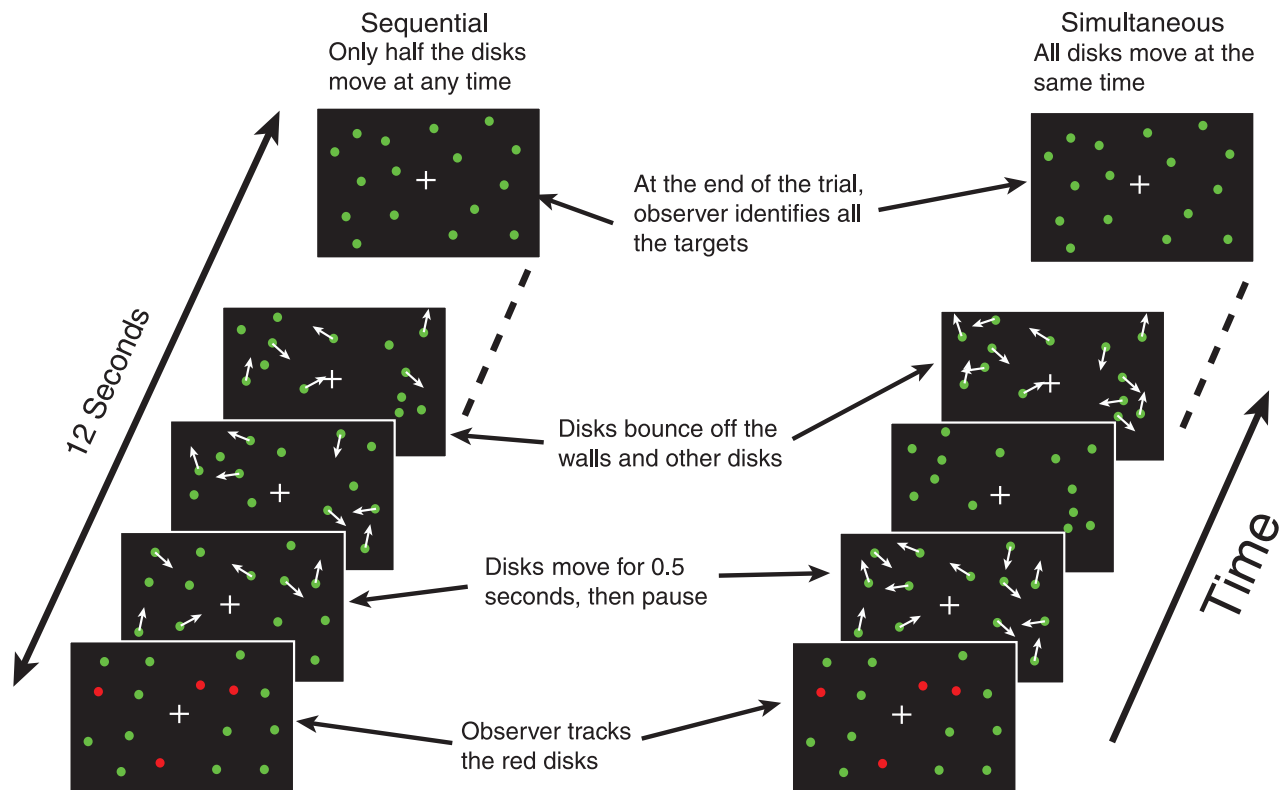

Figure 6. A cartoon of the method for Experiment 8. At the start of the trial a subset of the disks briefly turn red to indicate that these are the targets to be tracked. Each disk moved in a straight line (except when it bounced off the walls or another disk). Disks alternated between moving for 0.5 seconds and pausing for 0.5 seconds. In the sequential condition, only half the disks moved at any one time. In the simultaneous condition, all the disks moved and paused simultaneously. At the end of the trial, the observer was asked to click on all the targets.
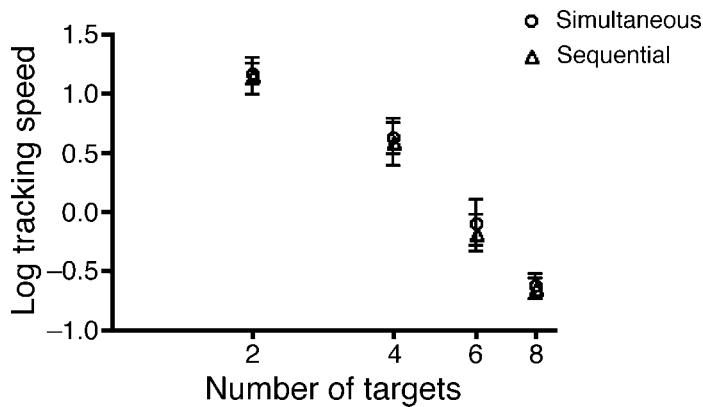
Figure 7. Results of Experiment 8. The tracking speed (measured in degrees per second) resulting in 75% correct accuracy is plotted as a function of the number of targets (both on a base 10 logarithm axis) for the simultaneous and sequential conditions. Error bars represent the standard error of the mean.

Gardner, 1972). In the simultaneous condition, all objects moved synchronously, then paused. In the sequential condition, half the objects (and half the targets) would move, then pause while the other half moved. According to the logic of the paradigm, if two targets are tracked in parallel, tracking accuracy should be equal in both the simultaneous and sequential conditions. Conversely, if the targets are tracked in series, we would expect performance to be greater in the sequential condition (Shiffrin & Gardner, 1972). In eight experiments, we found a sequential advantage only when we deliberately encouraged observers to adopt a serial strategy. Otherwise, across a variety of spatial arrangements and tracking loads, we found either equal performance or a simultaneous advantage.

We considered two simple models. The parallel model predicts that tracking of targets occurs in parallel regardless of where the targets are located. The serial model predicts that two targets will be tracked in parallel if one is the left hemifield while the other is the right hemifield (c.f. Alvarez & Cavanagh, 2005), but the two targets will be tracked in series if they are both placed in the same hemifield. Thus, the difference between simultaneous and sequential performance should strongly depend on the spatial arrangement of the stimuli. Experiment 1 tested these two competing predictions and found that tracking accuracy was greatest when the targets moved synchronously, regardless of whether the targets were in the same hemifield or in different hemifields.

Experiment 2 showed that even when there were four targets, each located in a different visual quadrant, tracking accuracy was still greater when the targets moved synchronously. Experiment 3 showed that this result held when there were just two targets, both located in the same visual quadrant. Experiment 4 confirmed these results with fixation enforced by eye tracking. Experiment 5

obtained a similar finding when targets were placed on the vertical midline so that they were repeatedly switching visual hemifields. No matter how we arranged the targets in space, no evidence for serial processes was found.

A possible concern was that the simultaneous-sequential paradigm might be flawed in that it might be impossible for accuracy to ever be greater in the sequential condition than in the simultaneous condition. Experiment 6 addressed this concern by using a stimulus where observers were obliged to track the targets in series. Here we found a substantial advantage for the sequential condition, as predicted. This bolstered our confidence in the validity of the paradigm.

In Experiment 7 we tested the hypothesis that tracking is parallel for up to four targets, but that observers must switch to a serial strategy when target load exceeds some threshold. We failed to find any evidence for a switch from parallel to serial tracking as the number of targets increased from two to eight. Experiment 8 repeated Experiment 7 using a more conventional MOT display in which the disks were free to move anywhere on the computer monitor. Again, we found no evidence for a switch in tracking strategy from parallel to serial with increasing target number.

In summary, our data argue strongly against the serial model proposed in the Introduction. However, this does not prove that it would be impossible to construct a more complex serial model that could explain our results, although our data are more directly in line with a parallel account of tracking.

Parallel models of MOT

Most models of MOT that have assumed that tracking is parallel. Perhaps the most influential model is that of Pylyshyn (1989, 2001), in which pre-attentive mental pointers, referred to as FINSTs (an acronym that stands for Fingers of INSTantiation) are attached to targets. The number of FINSTs limits the number of targets that can be tracked, which is about 3–5 depending on the observer. FINSTs do not encode any information about the identity of the targets but merely "point" at the targets' current locations. Crucially, each FINST tracks its target independently of the other FINSTs. Consequently, the FINST account predicts that the motion of one target should not affect the observer's ability to track a different target, regardless of the spatial arrangements of the targets. Thus, in Experiment 1, the model correctly predicts that performance will not be greater in the sequential condition and also correctly predicts that the difference in performance in the simultaneous and sequential conditions should not depend on the spatial arrangement of the targets. However, the model cannot explain the observed simultaneous advantage, instead predicting that performance should be equal in the simultaneous and sequential conditions.

Another concern with the FINST model is that it predicts that, because observers have only 3–5 FINSTs, observers should be able to track only this many objects in parallel. Tracking more than this many objects is predicted to require a serial process. However, in both Experiments 7 and 8 we could find no evidence of a switch from parallel to serial tracking as the number of tracked targets was increased from 2 to 8.

The model of Alvarez and Franconeri (2007) avoids this last concern. Like Pylyshyn and Storm, Alvarez and Franconeri predict that objects are tracked using mental pointers, which they refer to as FLEXs, an acronym that stands for FLexibly allocated indEXs. Unlike the FINST model, there is no theoretical limit on the number of FLEXs an observer may have. This would explain why no switch in tracking strategy was observed in Experiments 7 and 8. However, because FLEXs are predicted to act independently of each other, like FINST theory, FLEX theory must predict tracking performance to be equal in the simultaneous and sequential conditions, so cannot explain the simultaneous advantage generally observed in our experiments.

There have been several parallel models of MOT, variously positing that objects are tracked by being grouped into a virtual polygon (Yantis, 1992), by utilizing neural oscillators (Kazanovich & Borisyuk, 2006), or by a process where tracking performance is limited mainly by spatial interference (Franconeri et al., 2009). For our purposes, the differences between these models are less important than their similarities. Like the FLEX model, all three of the above models assume that a) there is no theoretical set limit on the number of targets that can be attended and b) the targets are all attended simultaneously. Consequently, they can explain the same aspects of our data as the FLEX model can but, like the FLEX model, have difficulty explaining the simultaneous advantage that we generally observed and that has also been regularly observed in the detection literature (Hung, Wilder, Curry, & Julesz, 1995; Shiffrin & Gardner, 1972). For a model to provide a complete description of MOT, it would need to provide an explanation for the simultaneous advantage. One possibility is that the simultaneous advantage is related to our use of rotary stimuli, since it was not observed (or at least substantially reduced) in Experiment 8, which utilized independent linear motions.

Conclusion

In summary, we provide the first direct empirical evidence against serial processing in multiple object tracking. Our version of the simultaneous-successive paradigm produced a sequential advantage (predicted by a basic serial model) only in Experiment 6, which was specifically designed to induce serial processing. In the other seven experiments, simultaneous tracking was equal or superior to sequential tracking, regardless of the spatial arrangement or the number of targets. These data should strongly constrain theoretical approaches to MOT.

References

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychology Science, 14,* 427–432. [PubMed]

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychology Science, 16,* 637–643. [PubMed]

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*(13):14, 1–10, http://www.journalofvision.org/content/7/13/14, doi:10.1167/7.13.14. [PubMed] [Article]

Awh, E., & Pashler, H. (2000). Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 834–846. [PubMed]

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. [PubMed]

Carlson, T. A., Alvarez, G. A., & Cavanagh, P. (2007). Quadrantic deficit reveals anatomical constraints on selection. *Proceedings National Academy of Science of the United States of America, 104,* 13496–13500. [PubMed] [Article]

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Science, 9,* 349–354. [PubMed]

d'Avossa, G., Shulman, G. L., Snyder, A. Z., & Corbetta, M. (2006). Attentional selection of moving objects by a serial process. *Vision Research, 46,* 3403–3412. [PubMed]

Drew, T., Horowitz, T. S., Wolfe, J. M., & Vogel, E. K. (submitted for publication). Isolating the neural signature of spatial updating during attentional tracking.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Reviews in Psychology, 48,* 269–297. [PubMed]

Eriksen, C. W., & Spencer, T. (1969). Rate of information processing in visual perception: Some results and methodological considerations. *Journal of Experimental Psychology, 79,* 1–16. [PubMed]

Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science, 21,* 920–925.

Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2009). Tracking multiple objects is limited only by interobject crowding, and not object speed. *Psychonomic Society Annual Meeting.*

Hogendoorn, H., Carlson, T. A., & Verstraten, F. A. J. (2007). The time course of attentive tracking. *Journal of Vision, 7*(14):2, 1–10, http://www.journalofvision.org/content/7/14/2, doi:10.1167/7.14.2. [PubMed] [Article]

Horowitz, T. S., Wolfe, J. M., Alvarez, G. A., Cohen, M. A., & Kuzmova, Y. I. (2009). The speed of free will. *Quarterly Journal of Experimental Psychology, 62,* 2262–2288. [PubMed]

Howe, P. D., Horowitz, T. S., Morocz, I. A., Wolfe, J. M., & Livingstone, M. S. (2009). Using fMRI to distinguish components of the multiple object tracking task. *Journal of Vision, 9*(4):10, 1–11, http://www.journalofvision.org/content/9/4/10, doi:10.1167/9.4.10. [PubMed] [Article]

Hung, G. K., Wilder, J., Curry, R., & Julesz, B. (1995). Simultaneous better than sequential for brief presentations. *Journal of Optical Society of America A, Optics, Image Science, and Vision, 12,* 441–449. [PubMed]

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision, 9*(4):1, 1–12, http://www.journalofvision.org/content/9/4/1, doi:10.1167/9.4.1. [PubMed] [Article]

Kazanovich, Y., & Borisyuk, R. (2006). An oscillatory neural model of multiple object tracking. *Neural Computation, 18,* 1413–1440. [PubMed]

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Research, 34,* 885–912. [PubMed]

Landry, S., & Sheridan, T. (2001). A methodology for studying cognitive groupings in a target-tracking task. *IEEE Transactions on Intelligent Transportation Systems, 2,* 92–100.

Luck, S. J., Hillyard, S. A., Mangun, G. R., & Gazzaniga, M. S. (1989). Independent hemispheric attentional systems mediate visual search in split-brain patients. *Nature, 342,* 543–545. [PubMed]

McMains, S. A., & Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron, 42,* 677–686. [PubMed]

Oksama, L., & Hyona, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Visual Cognition, 11,* 631–671.

Oksama, L., & Hyona, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology, 56,* 237–283. [PubMed]

Pashler, H. (1999). *The psychology of attention.* Cambridge, MA: MIT Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. [PubMed]

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

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80,* 127–158. [PubMed]

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179–197. [PubMed]

Saarinen, J., & Julesz, B. (1991). The speed of attentional shifts in the visual field. *Proceedings of the National Academy of Sciences of the United States of America, 88,* 1812–1814. [PubMed] [Article]

Scharlau, I. (2004). Evidence for split foci of attention in a priming paradigm. *Perception & Psychophysics, 66,* 988–1002. [PubMed] [Article]

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80,* 1–46. [PubMed]

Scholl, B. J. (in press). What have we learned about attention from multiple object tracking (and vice versa). In D. Dedrick & L. Trick (Eds.), *Computation, cognition, and Pylyshyn.* Cambridge, MA: MIT Press.

Shiffrin, R. M., & Gardner, F. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology, 93,* 72–83. [PubMed]

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33,* 113–120. [PubMed]

Wolfe, J. M., Alvarez, G. A., & Horowitz, T. S. (2000). Attention is fast but volition is slow. *Nature, 406,* 691. [PubMed]

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

Zelinsky, G. J., & Neider, M. B. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition, 16,* 553–566.