

Visual Attention Is Required for Multiple Object Tracking

Annie Tran and James E. Hoffman
University of Delaware

In the multiple object tracking task, participants attempt to keep track of a moving set of target objects embedded in an identical set of moving distractors. Depending on several display parameters, observers are usually only able to accurately track 3 to 4 objects. Various proposals attribute this limit to a fixed number of discrete indexes (Pylyshyn, 1989), limits in visual attention (Cavanagh & Alvarez, 2005), or “architectural limits” in visual cortical areas (Franconeri, 2013). The present set of experiments examined the specific role of visual attention in tracking using a dual-task methodology in which participants tracked objects while identifying letter probes appearing on the tracked objects and distractors. As predicted by the visual attention model, probe identification was faster and/or more accurate when probes appeared on tracked objects. This was the case even when probes were more than twice as likely to appear on distractors suggesting that some minimum amount of attention is required to maintain accurate tracking performance. When the need to protect tracking accuracy was relaxed, participants were able to allocate more attention to distractors when probes were likely to appear there but only at the expense of large reductions in tracking accuracy. A final experiment showed that people attend to tracked objects even when letters appearing on them are task-irrelevant, suggesting that allocation of attention to tracked objects is an obligatory process. These results support the claim that visual attention is required for tracking objects.

Keywords: attention, multiple object tracking, visual search

Many of our everyday activities, such as playing a team sport, driving on a busy street, or monitoring several children at a pool, require us to engage in multiple object tracking (MOT). Laboratory studies of MOT have generally been based on a task introduced by Pylyshyn and Storm (1988) in which observers attempt to keep track of a set of target objects embedded in identical moving distractors. Typically, observers are accurate at tracking up to about four objects before accuracy begins to drop precipitously (Cavanagh & Alvarez, 2005; Pylyshyn & Storm, 1988; Yantis, 1992), although the number of objects that can be tracked varies substantially depending on factors such as movement velocity, interobject spacing, and so forth (Alvarez & Franconeri, 2007).

A wide variety of models have been proposed to account for observed limits in our tracking ability. These models are similar to those proposed to account for comparable limits in maintaining visual objects for brief periods in visual short-term memory (VSTM). In fact, the underlying mechanism responsible for limits in these two domains may be related (Oksama & Hyona, 2008), so we might profit from integrating theoretical accounts developed in the domains of MOT and VSTM.

A first theoretical cut through the domain of tracking models rests on the distinction between explanations that assume some

sort of flexible allocation of a limited “processing resource” versus constraints attributable to built-in architectural limitations. In addition, within the domain of limited processing capacity, we can distinguish between claims that processing capacity is allocated in discrete chunks or as a continuous resource.

Discrete Resource Model of MOT

Pylyshyn (1989) proposed that people track objects using a pointer or indexing system. An index is assumed to be a primitive selection mechanism that can keep track of an object without necessarily providing any information about its identity or other properties. The number of indexes is assumed to be limited to three or four, thereby providing a straightforward account of the limited number of objects that can be tracked. Indexes appear to be similar to slots (Zhang & Luck, 2008, 2011), which have been proposed as an explanation for the limited number of items that can be stored in VSTM.

In Pylyshyn’s original theory, indexes were assumed to be “sticky” so that they remained attached to their corresponding objects without requiring the intervention of limited-capacity processes (Pylyshyn & Storm, 1988). However, in later versions, Pylyshyn (1994) suggested that visual attention may sometimes be allocated to tracked objects in order to “refresh” the indexes or to keep them from inadvertently “slipping off” onto nearby distractors. He suggested other reasons as well for a close relationship between the indexing and visual attention systems, namely that visual properties of an object, such as shape, color, identity, and so forth, only become available when people allocate visual attention to an object; the allocation of attention, in turn, depends on the prior assignment of an index to that object. In this model, visual attention has a unitary focus that can only be allocated to one

Annie Tran and James E. Hoffman, Department of Psychological and Brain Sciences, University of Delaware.

This research was supported by NSF Grant BCS-1059560 to J.E.H.

Correspondence concerning this article should be addressed to Annie Tran, Department of Psychological and Brain Sciences, Wolf Hall 108, 105 The Green, University of Delaware, Newark, DE 19716-2577. E-mail: atran@psych.udel.edu

object at a time even though several indexes can be simultaneously allocated to tracked objects.

Continuous Resource Model of MOT

Cavanagh and Alvarez (2005) challenged the claim that MOT is limited by a fixed set of indexes or pointers and suggested instead that attention, which was initially characterized as being allocated in the form of a single “spotlight,” could be divided into separate spotlights allocated to each of the to-be-tracked objects (the “multifocal attention model”). Increasing the number of targets would result in each of them receiving a smaller “slice of the pie” so that tracking accuracy would be expected to decline with increasing set size. Similar proposals that visual working memory involves competition for a limited, continuous resource were shown to account for set size effects on report accuracy (Bays, Catalao, & Husain, 2009; Keshvari, van den Berg, & Ma, 2013; Wilken & Ma, 2004). The continuous resource model nicely accounts for the finding that the number of objects that can be tracked varies over a wide range depending on “tracking difficulty” (close spacing or high velocity). For example, Alvarez and Franconeri (2007) found that people could track anywhere from one to 12 objects depending on their velocity. This result isn’t easily explained by the indexing/slot class of model.

Capacity-Free Models

Franconeri and colleagues (Franconeri, 2013; Franconeri et al., 2007, 2010, 2013) pointed out that explaining limits in performance in terms of a hypothetical, capacity-limited attention system may only give the illusion of providing an explanation: “Unfortunately, words like ‘capacity,’ ‘resources,’ and ‘load’ relabel the effect without explaining why it occurs” (p. 134). They suggested that limits in tracking accuracy reflect competition for representation in cortical maps. Attending to an object is thought to involve a central area of enhancement surrounded by a zone in which objects are inhibited or suppressed (Franconeri, 2013; Franconeri et al., 2013). If these suppressive zones are large relative to the total area available for representing objects, then attending to more than one object would increase the probability of these suppressive zones intruding on the central enhancement areas resulting in selection failures. Notice that this is somewhat similar to crowding limits in which neighboring objects in the periphery become difficult to individuate resulting in failures to correctly identify them. In this theory, deleterious effects of increasing velocity or increases in the number of objects tracked can be attributed to the increase in the number of close approaches between targets that accompanies increasing velocity. A similar argument can be made for the effects of increasing set size. The only limited resource in this model is the limited spatial resolution for individuating closely spaced objects associated with topographically organized visual cortex (Franconeri, 2013).

Vul, Frank, Alvarez, and Tenenbaum (2010) proposed a computational model to account for tracking performance without assuming capacity limits. They used a particle filter approach that predicts the future position of an object based on its current position and knowledge of its direction and velocity of movement. In this model, errors arise because of noise in these various estimates, which allowed the model to show effects of spacing and

velocity that were consistent with human performance. However, they also found that tracking declined with increases in the number of objects to be tracked and this could not be accounted for by their model. They concluded that set size effects implicate the need for a limited-capacity attentional system that plays a role in the tracking task. A later paper by Srivastava and Vul (2016) added an attentional component to the model in which a limited resource, assumed to correspond to visual attention, was divided among tracked objects and was dynamically allocated to targets when tracking difficulty increased. This model, which has similarities to the multifocal attention model, provides a good account of both set size and velocity effects on tracking accuracy. In agreement with their model, they found that people localized targets more accurately under crowded display conditions relative to sparse spacing (see Iordanescu, Grabowecy, & Suzuki, 2009 for a similar result) presumably because more attention was allocated to targets when they were approached by distractors.

The Current Experiments

There have been a number of models aimed at explaining our ability to keep track of moving objects. Some of these models have proposed that limits in tracking are attributable, at least in part, to the need to allocate a limited capacity resource to tracked objects. Further, this limited resource might correspond to visual attention, perhaps flexibly allocated in the form of separate “spotlights” bound to tracked objects (Cavanagh & Alvarez, 2005). In the first two experiments, we evaluate this class of models using letter probes appearing on tracked objects and distractors. Visual attention is known to improve the speed and/or accuracy of probe identification (Carrasco & McElree, 2001; Doran & Hoffman, 2010; Eriksen & Hoffman, 1972a, 1972b, 1973; Hoffman & Nelson, 1981; Posner, 1980; Szinte, Carrasco, Cavanagh, & Rolfs, 2015; Theeuwes & Van der Burg, 2007), so discrimination speed and/or accuracy for letter probes appearing on different objects can serve as an objective measure of how much attention those objects are receiving during MOT.

In addition, we evaluate whether advantages for probes appearing on tracked objects are simply due to a convenient choice of a default strategy or reflect the necessity of attentional allocation for tracking. We examined this issue by varying the probability of probes appearing on tracked objects versus distractors. If tracking doesn’t require visual attention, as claimed by index models, then participants should be free to allocate their attention to those objects that are likely to host a probe letter, independent of tracking requirements because tracking depends on indexes. In contrast, if tracking requires visual attention, as claimed by continuous resource models, allocation of attention should be strongly constrained by tracking, particularly in the set size three condition which would be expected to use most of the available resources. In this case, one should observe severely reduced identification of probe letters appearing on distractors.

Experiment 1

In Experiment 1, participants tracked one or three “figure eight” objects (see Figure 1) while also trying to identify probe letters appearing on a target or a distractor (see Sears & Pylyshyn, 2000 for a similar paradigm). The distractor was a different color than

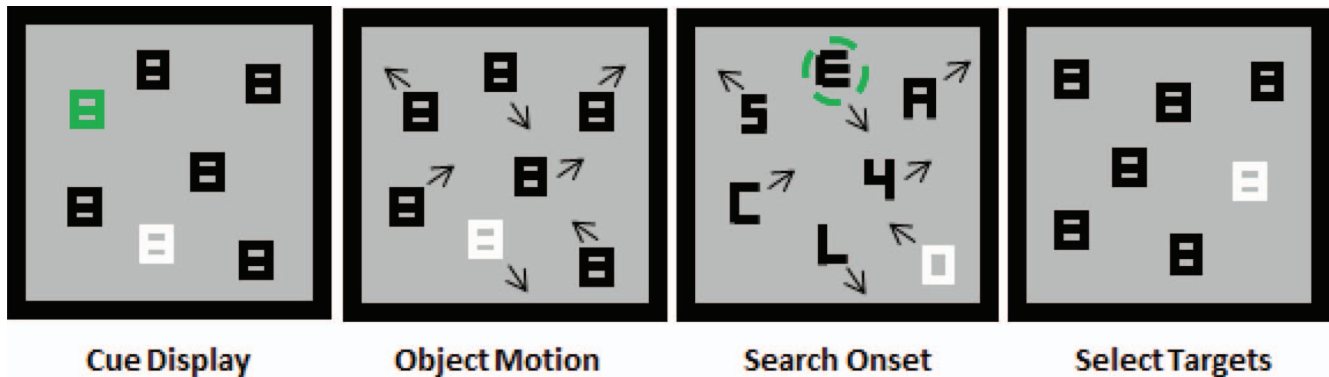


Figure 1. An example of a trial sequence in Experiment 1. The visual display consisted of seven objects, six of the same color and a singleton distractor having a different color (shown in white). MOT was paired with visual search in the dual task condition. The to-be-tracked objects were cued initially in green (colored gray in the gray-scale image) at the start of each trial. The probability of the search letters (i.e., E, H, F, P) appearing on a tracked object versus the singleton was varied. The circular indicator around the target letter is for clarification only. See the online article for the color version of this figure.

the other objects in the display (a singleton), allowing participants to attend to it without mistaking it for a to-be-tracked object (Tran & Hoffman, 2013). In different blocks of trials, the probe letter appeared on a tracked object with probability 1.0, 0.7, 0.3, or 0 and appeared on a distractor with the complementary probability. Participants were instructed to maintain their tracking performance at the same high level in all conditions and monetary rewards were used to reinforce this emphasis on the tracking task.

Method

Participants. There were 10 participants (2 male, 8 female) between ages 18 and 29 years old (mean age = 22.7 years old) in the track-1 group. Twelve participants were recruited for the track-3 group; however, two participants were dropped from analysis because of low tracking accuracy (below 70% correct). The remaining 10 participants (4 male, 6 female) in the track-3 group were between the ages of 18 and 29 years old (mean age = 23.3 years old).

Stimuli and procedure. Observers were tested in a dimly lit and acoustically isolated room. Displays were presented on a SAMSUNG 2233RZ 22" LCD Monitor (Wang & Nikolic, 2011) having $1,680 \times 1,050$ pixel resolution and a 120-Hz refresh rate. Viewing distance was fixed at 70 cm using a chin rest. Experiments were controlled by a Dell 3.60 GHz computer. The experiment was programmed in MATLAB (Mathworks, Natick, MA; <http://www.mathworks.com>) using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Eye movements were monitored with an Eyelink 1000 eye tracker using the Eyelink Toolbox extensions (SR Research, Ontario Canada; Cornelissen, Peters, & Palmer, 2002), which sampled eye position at 500-Hz. An eye movement was defined as three consecutive eye samples that were more than 1.4 degrees of visual angle (dva) from fixation.

Stimuli consisted of seven figure-eight objects which could have segments deleted to reveal the capital letters E, H, F, or P (see Sears & Pylyshyn, 2000). The size of each object was 1.1×1.3 dva in width and height, respectively. Six identical objects were presented in either black or white on a gray background with the

opposite color assigned to the singleton distractor. The assignment of the two possible colors to the singleton versus the other display objects was counterbalanced across participants. Tracking load (one vs. three) was varied between different groups of participants.

A schematic display of a typical trial sequence is shown in Figure 1. At the start of each trial, observers fixated a $.08 \times .08$ dva square at the center of the screen and pressed a mouse button to initiate the trial. The objects then appeared randomly positioned within a 12.4×9.8 dva rectangular area centered on fixation. The to-be-tracked objects were shown in green for one second at the start of each trial. Objects then moved along linear trajectories in random directions with a velocity of 6 dva/s and bounced off the edges of the rectangular region as well as each other when they were closer than 1.6 dva center-to-center. The objects also bounced off an invisible circular region with a 2.3 dva radius that surrounded the fixation point to prevent it from being occluded by the moving objects.

One of four different, randomly chosen search targets (E, H, F, or P) appeared on either the tracked object or the singleton between 1 and 3 seconds (uniform distribution) after the start of each trial. In the first block of trials, a target appeared on the *singleton* on 70% of the trials while in the second block, it appeared on the *tracked object* on 70% of the trials. The last two blocks were single-task search control conditions in which the target letter appeared only on a tracked object or on the singleton, respectively. The onset of the search target was accompanied by characters appearing on all the other objects, including the singleton. These characters were chosen randomly without replacement from a set of letters and numbers (i.e., A, C, J, L, O, S, U, 2, 4, 6, 7, 9). Letters were shown for a limited duration (determined by a staircase procedure described below) and then reverted to figure eights.

Following a four-second motion period, participants were required to select the tracked objects by clicking on them using the mouse. The unique color of the singleton was maintained during the response phase. Feedback was provided after each mouse click by changing the color of the chosen letter to red (incorrect) or green (correct). Participants then chose the search target by click-

ing on one of the four possible target letter choices. Once again, feedback was provided in the form of a color change to red or green. There were three runs, each containing 3 blocks of 48 trials each for a total of 432 trials per participant. Participants were told that they would receive a monetary reward on each trial of 1.5 cents each for correct performance on tracking and probe identification. *This reward was conditional on correct tracking on that trial.* This was designed to encourage participants to prioritize the tracking task while also rewarding them for correctly carrying out the search task.

Presentation durations for the search target were adjusted using a three-up, one-down staircase procedure in which duration was increased following an error and decreased following three consecutive correct responses. This procedure should converge on an accuracy level of approximately 80%. The staircase procedure continued until four reversals were reached, and an estimated asymptotic value was obtained by taking the average of these reversals. Separate durations were estimated for search targets appearing on the tracked object and the singleton. Note that there are two different ways one might adjust the exposure duration in the single task search condition. One way would be to have people search through the entire display for the target letter. The other is to indicate a subset of objects (one or three) that will always contain the target. We chose the latter procedure because maximum search performance, which is what the single-task search performance should reflect, will occur when targets appear in known locations and people are not required to make a tracking response (although they are probably tracking the cued locations as part of attending to them). This also allowed us to find different durations for the targets appearing on the tracked object(s) and the singleton in order to make performance comparable across physically different objects (singleton vs. nonsingletons) and differing numbers of relevant locations (as in set size 1 vs. set size 3).

During the staircase procedure in the search-only condition, participants were instructed that one or three objects (depending on the group) would be cued by changing to a green color at the start of each trial and that search targets would only occur on the cued objects. No tracking response was required however. Similarly, when the staircase was applied to targets appearing on the singleton object, participants were instructed that search targets would only appear on that object. The staircase procedures were conducted before the start of the experiment and between each run if search accuracy deviated more than 10% from the criterion accuracy of 80%.

Results and Discussion

In the set size 1 condition 10.4% of trials, and 15.11% of trials in the set size 3 condition, were rejected because of eye movements. Trials containing eye fixations deviating 1.87 dva or more from the fixation point were rejected. In the set size 1 condition, the average probe durations were 164.0 ms for the tracked object and 157.5 ms for the singleton. The corresponding values for the set size 3 condition were 319 ms for tracked objects and 108 ms for the singleton. Not surprisingly, a longer exposure duration was required to achieve the same probe accuracy when monitoring three locations compared to one. Average accuracy in the single-task search condition was 84.68% for probes on cued objects and 81.61% for probes on the singleton object for set size 1. For set

size 3 condition, the corresponding values were 85.19% and 83.72%. These values were normalized (Alvarez, Horowitz, Arsenio, Dimase, & Wolfe, 2005) as described below in the *Search Accuracy* section.

Tracking accuracy. Tracking accuracy was analyzed using a mixed ANOVA with a between-subjects factor of Set Size (one vs. three) and a within-subjects factor of Target Probability (Probability of target appearing on a tracked object: 30%T vs. 70%T). Although the presented means for each condition below are the untransformed accuracy scores, an arcsine transformation was performed on the data before conducting the ANOVA. There was a small but significant advantage in tracking accuracy when probability favored the search target appearing on a tracked object ($M = 96.8$) compared with the singleton ($M = 95.02$), $F(1, 18) = 9.73$, $p = .006$. There was no main effect of Set Size, $F(1, 18) = 3.74$, $p = .069$, or its interaction with probability, $F(1, 18) = 1.16$, $p = .29$. These results indicate that participants were largely successful in maintaining high tracking accuracy in all conditions.

Search accuracy. Visual search accuracy for targets appearing on tracked objects and distractors is shown in Figure 2 in the form of an attention operating characteristic or AOC (Alvarez et al., 2005; Hoffman & Nelson, 1981; Sperling & Melchner, 1978). Each data point in the interior of the figure represents joint search performance for targets appearing on tracked objects and singleton distractors for each probability condition. Search accuracy was normalized to a scale extending from 0 to 100 using the following formula from Alvarez et al. (2005): normalized accuracy = [(dual-task accuracy – chance)/(single task accuracy – chance)] × 100. Chance performance on the search task was 25%. This transformation expresses dual-task accuracy as a percentage of single-task

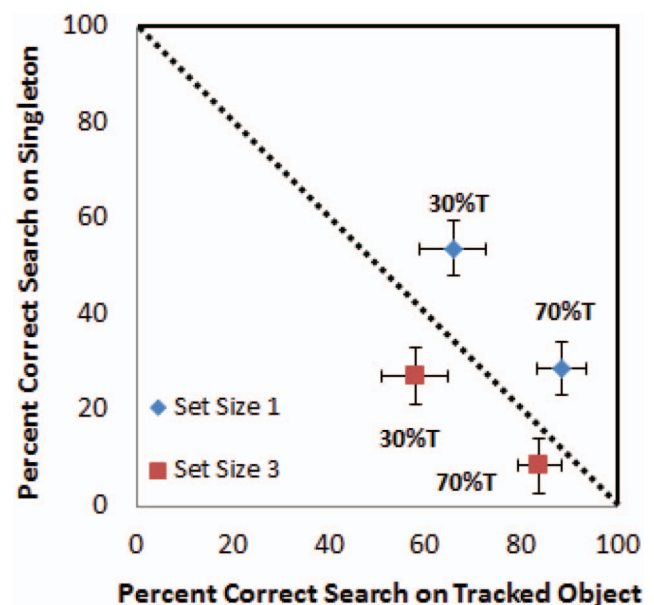


Figure 2. Attention operating characteristic (AOC) showing relationship between normalized search accuracy for letters appearing on the tracked object versus the distractor as a function of the number of objects being tracked (one vs. three) and the probability of targets appearing on the tracked object in Experiment 1. Error bars represent the standard error of the mean. See the online article for the color version of this figure.

performance. When performance in dual-task and single-task conditions is the same, the normalized score for the dual-task condition is 100%. Therefore, independence in performing the two tasks (i.e., no dual-task interference) is represented by a point in the upper right corner of the figure which corresponds to 100% normalized performance on both tasks. The diagonal line connecting the two single-task search conditions is the line of “mutual exclusivity” and represents dual-task performance in which the two tasks are mutually exclusive. In this case, accuracy gains in one task are accompanied by equivalent losses in the other task.

We evaluated whether the dual-task points were on the line of mutual exclusivity by computing the signed distance between each dual-task point and the diagonal line and using single sample *t* tests to determine whether this distance was greater than zero (Alvarez et al., 2005). In the track-3 condition, the distances between the two dual task points and the diagonal were not significantly greater than zero: 30%T, $t(9) = -1.52, p = .16$ and 70%T, $t(9) = -0.99, p = .35$. In contrast, in the track-1 condition, the dual task point for the 30%T condition was significantly above the diagonal, $t(9) = 2.28, p = .048$, whereas in the 70%T condition it was not, $t(9) = 2.03, p = .07$. Performance above the line would be expected if tracking a single object resulted in reserve capacity that could be allocated to the singleton in order to identify the target letter appearing there. Tracking three objects would presumably use more capacity than tracking a single object, leaving a smaller supply to be allocated to the distractor, resulting in reduced accuracy in identifying a target appearing there.

Non-normalized visual search accuracy was also analyzed using a mixed ANOVA with a between-subjects factor of Set Size (one vs. three) and within-subject factors of Target Probability (Probability of target appearing on tracked object: 30%T vs. 70%T) and Target Location (Tracked object vs. Distractor). There was a main effect of target location, $F(1, 18) = 116.8, p < .001$, reflecting better search performance for targets appearing on tracked objects ($M = 71.2%$) versus the singleton ($M = 43.6%$). There was no main effect of probability, $F(1, 18) = .201, p = .659$. The interaction of probability and target location was significant, $F(1, 18) = 39.47, p < .001$, demonstrating a trade-off in search accuracy between the two locations as a function of probability. Increasing the probability of a search target appearing on a tracked object improved accuracy when it occurred there, and this effect was accompanied by a decrease in accuracy for a target appearing on the distractor. Superimposed on this trade-off is the overall superiority for targets appearing on tracked objects. This can be seen most clearly in the condition where the search target was most likely to appear on the distractor (30%T condition). Search accuracy here was equal for the two locations (tracked object and distractor) despite the fact that the search target was more than twice as likely to occur on the distractor. Participants appeared to be unable to allocate a majority of their attention to the singleton when the target was highly likely to occur there, presumably because the requirement to maintain accurate tracking required a certain minimum amount of attention to be allocated to the tracked objects.

These results then are consistent with the idea that tracking depends on a limited-capacity attentional resource with the number of objects being tracked affecting the amount of resources that are available to be allocated to nontracked objects.

In Experiment 1, accuracy for discriminating targets on tracked objects was always equal to or better than accuracy for targets appearing on a distractor, regardless of probability. This is reflected in the fact that all four dual-task points in Figure 2 are located in the lower right quadrant, corresponding to good performance for targets appearing on a tracked objects coupled with relatively poor performance for targets appearing on the distractor. This was the case even when the target was more than twice as likely to appear on the distractor compared to a tracked object. These results suggest that tracking may require some minimum allocation of visual attention to the tracked object in order to maintain high tracking accuracy and this minimum allocation limits how much attention can be allocated to the distractor. Effects of target probability would then have to work within the allocation limits set by the tracking task. If tracking three objects compared with one requires more visual attention then there should be less attention left over to allocate to the distractor resulting in poorer performance in identifying targets appearing there which is what we observed.

Maintaining some minimum allocation of resources to tracking may simply be a rational strategy induced by our instructions and rewards which prioritized the tracking task over search. If so then we should see a more symmetric trade-off in performance across tracked objects and distractors if we remove the emphasis on tracking accuracy. In this case, observers should be free to shift a larger portion of their attentional resources away from the tracked objects to the distractor object when targets are likely to occur there; doing so, however, should incur a cost in terms of significant decreases in tracking accuracy because the amount of attention allocated to the tracked object will fall below the minimum required for accurate tracking.

Experiment 2

In Experiment 2, we changed our instructions in several ways to shift the emphasis from maintaining high tracking accuracy to performing well on the search task. A larger reward was provided for being accurate on the search task compared to tracking. In addition, search rewards were no longer contingent on correct tracking. Note that observers are still rewarded for correct tracking so the optimal strategy is to be accurate on both tasks if possible. However, if search and tracking are competing for the same limited resource, as we are claiming, then tracking performance should drop when search targets are likely to appear on a distractor and participants shift a large share of their attentional resources to the distractor.

In this experiment, the tracking load was set to two objects and the velocity of the moving objects was adjusted using a staircase procedure designed to keep tracking performance at approximately 80% in the track-only condition.

Method

Participants. Ten participants (3 male, 7 female) ranging from 18 to 24 years of age (mean age = 18.9 years) completed Experiment 2. All participants reported normal or corrected-to-normal visual acuity. Monetary payment was provided for participation.

Stimuli and procedure. Stimuli and procedures were identical to those used in Experiment 1 except for the following changes.

Observers tracked two objects among four identical distractors and one singleton distractor. A single-task tracking condition was included to provide a comparison to the dual-task condition. At the beginning of the experiment, object speed was adjusted using a three-up, one-down staircase procedure designed to keep tracking performance near 80%. This procedure was repeated between each experimental run if accuracy deviated more than 10% from the criterion accuracy of 80%. There were two runs in the experiment, each containing five blocks with 60 trials each in the first two dual-task blocks and 20 trials each in the remaining three single-task blocks. The dual-task blocks began with the condition in which there was a 30% probability of a target occurring on a tracked object (30%T), followed by the 70% probability (70%T) condition. The subsequent single-task blocks were presented in this order: (a) single-task tracking, (b) single-task search with targets appearing only on cued objects, and (c) single-task search with targets appearing only on the singleton. Participants were informed that they would receive a monetary reward on each trial of one cent for correct tracking and three cents for correct search identification. This reward system was designed to emphasize accuracy on the search task while still rewarding accurate tracking.

Results and Discussion

As a result of eye movements, 11.68% of trials were rejected. The average speed of the objects across participants, as determined by the staircase procedure, was 13.5 dva/s. The average search target durations for tracked objects and the singleton were 234.17 ms and 130.83 ms, respectively. Average non-normalized accuracy in the single-task tracking condition was 78.08%. The average non-normalized accuracy in the single-task search condition was 79.87% on cued objects and 83% on the singleton object.

Search accuracy. Visual search accuracy is shown in Figure 3 in the form of an attention operating characteristic (AOC), in which each data point represents the joint normalized search accuracy for targets appearing on the tracked objects and the distractor across the different probability conditions. The signed distance between each data point and the diagonal line was analyzed using t tests to examine whether the distance was significantly greater than zero (Alvarez et al., 2005). The distance was not significantly different from zero in either the 30%, $t(9) = -1.13, p = .29$ or 70%, $t(9) = 0.56, p = .59$, probability conditions. Similar to the case of tracking three objects in Experiment 1, the two dual-task data points for this experiment appear to lie along the negative diagonal representing performance on tasks that are mutually exclusive.

The new finding in the present experiment is that participants apparently shifted a large portion of their attention to the distractor when the target was likely to occur there, resulting in increased target accuracy for that object (see Figure 3). However, this increase in performance for distractor targets was accompanied by a large reduction in accuracy for targets appearing on tracked objects. Normalized accuracy for targets appearing on tracked objects was considerably lower (32%) here than it was in Experiment 1 (normalized accuracy of 58% and 66% correct for set size 3 and 1, respectively). In addition, unlike Experiment 1, participants in this experiment were able to achieve higher accuracy for probes appearing on the distractor relative to the tracked objects but only at a cost of sharply reduced accuracy on the tracking task.

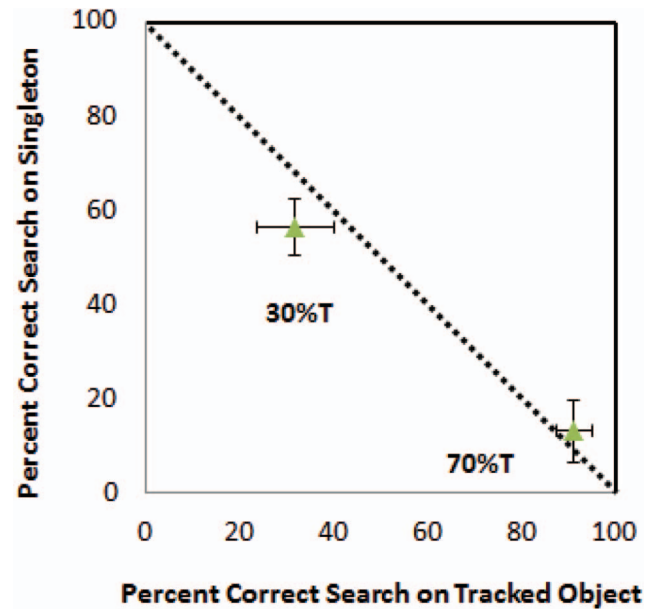


Figure 3. Attention operating characteristic (AOC) showing joint normalized search accuracy for letters appearing on a tracked object versus a distractor as a function of the relative probabilities of targets appearing in the two locations in Experiment 2. Error bars represent the standard error of the mean. See the online article for the color version of this figure.

In Experiment 1, participants were able to increase their attention to the distractor with only minor reductions in tracking accuracy (1.78%) whereas in this experiment, tracking accuracy decreased by approximately 18% compared with single-task performance. These different outcomes appear to be attributable to at least two differences in method between Experiments 1 and 2. First, the velocity of the moving objects in Experiment 2 was adjusted so that tracking performance was below ceiling and above chance. This allowed us to meaningfully assess whether *tracking accuracy* was affected by changes in attention allocation to the two tasks. Second, both the instructions and the reward structure were changed to encourage participants to be willing to sacrifice tracking accuracy to maximize rewards on the search task in the case where targets were most likely to appear on the distractor.

Here we report traditional analyses of the non-normalized search and tracking accuracy outside of the AOC framework. Visual search accuracy was analyzed using a 2 (Probability of target appearing on tracked object: 30%T vs. 70%T) \times 2 (Target location: Tracked object vs. Distractor) repeated-measures ANOVA. There was no main effect of probability, $F(1, 9) = 3.1, p = .11$. In keeping with our previous findings, we found a main effect of target location, $F(1, 9) = 17.37, p = .002$, reflecting more accurate identification of targets appearing on tracked objects ($M = 58.48$) compared with the singleton ($M = 45.13$). The crucial point of interest is the interaction of probability with target location, $F(1, 9) = 77.03, p < .001$, which showed a significant trade-off in accuracy between the search locations as a function of probability.

Tracking accuracy. Tracking accuracy is shown in Figure 4. We analyzed tracking accuracy using a one-way within-subjects ANOVA which revealed a significant effect of task condition

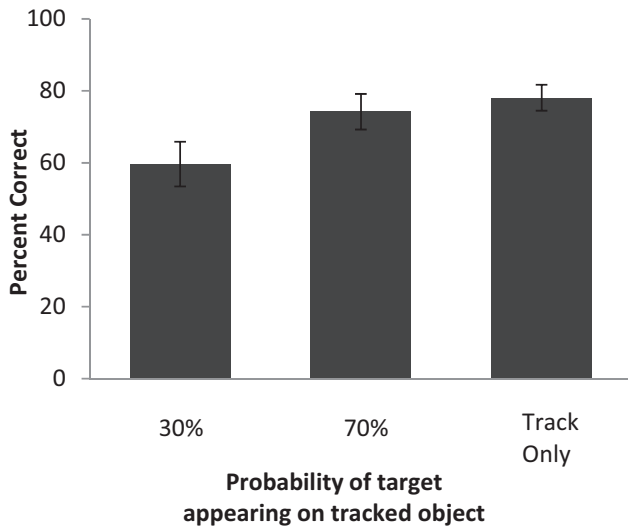


Figure 4. Tracking accuracy for the dual-task and single task condition from Experiment 2. Accuracy measures are presented as a function of the probability of targets appearing on the tracked object.

(30%T vs. 70%T vs. single-task tracking), $F(2, 18) = 24.12, p < .001$. Least Significant Difference (LSD) pairwise comparisons were performed to examine differences within condition pairs. The LSD test is restricted to cases with a significant main effect and three levels of the independent variable in which case there is no requirement to correct for multiple comparisons (Cardinal & Aitken, 2006; Howell, 2013). The LSD post hoc tests do not provide the value of the t -statistic; the provided p values are equivalent to that of t tests. Tracking accuracy was significantly lower in the 30%T condition ($M = 59.63$), in which targets were most likely to occur on the singleton, than the 70%T condition ($M = 74.22, p = .001$) and single-task tracking condition ($M = 78, p < .001$). Tracking accuracy was not significantly different between the 70% and single-task tracking conditions, $p = .11$.

These results show that tracking is dependent on visual attention and that once attention to a tracked object falls below some minimum level, people tend to lose track of it. Attention, as measured by the ability to discriminate briefly presented shapes, appears to be closely linked to object tracking. Tracking involves attending to the tracked objects and attending to objects involves tracking them.

Experiment 3

The previous two experiments indicate that people allocate visual attention to tracked objects and that when attention is withdrawn, tracking performance suffers. It appears, therefore, that attention plays a critical role in the task of MOT. However, in both of these experiments, letters that were relevant for the search task appeared on the display objects and, because visual attention is needed for target identification during search, one could argue that people do not need to attend to the objects for tracking but they do so in order to identify the search targets that sometimes appear there. This doesn't account for the ubiquitous superiority of detecting targets on the tracked objects relative to distractors or the other systematic patterns we observed that implicate visual atten-

tion as the principal mechanism for performing tracking. However, we think it is still worthwhile to examine whether people continue to allocate visual attention to tracked objects even when they have no reason to pay attention to them other than for tracking.

In this experiment, we removed the need to identify targets appearing on tracked objects (see Figure 5). Observers were required to track a set of moving objects while monitoring a stationary stream of letters for the occurrence of a target letter (an "E" or an "H"). Similar letter streams continuously appeared on the moving objects but these were irrelevant to the search task. A *task-irrelevant* target letter was appearing in one of the moving object streams when a *task-relevant* target appeared in the fixed location stream. The irrelevant target letters appeared equally often on a tracked object or a distractor and were either compatible or incompatible with the relevant target appearing in the fixed location. Similar to the flanker task (Eriksen, 1995; Eriksen & Eriksen, 1974; Eriksen & Hoffman, 1973), response competition should be observed in the form of increased errors and slowed reaction time (RT) to a target when it is accompanied by attended but irrelevant, response-incompatible distractors. If people attend to the moving objects as part of the tracking process then they should be slower on trials in which an incompatible versus a compatible letter appears on a tracked object. No compatibility effects should occur in response to target letters appearing on nontracked objects.

Method

Participants. Twenty students from the University of Delaware were recruited for this experiment; however, two students had incompatible minus compatible RT difference scores that were more than three standard deviations from the mean of all subjects

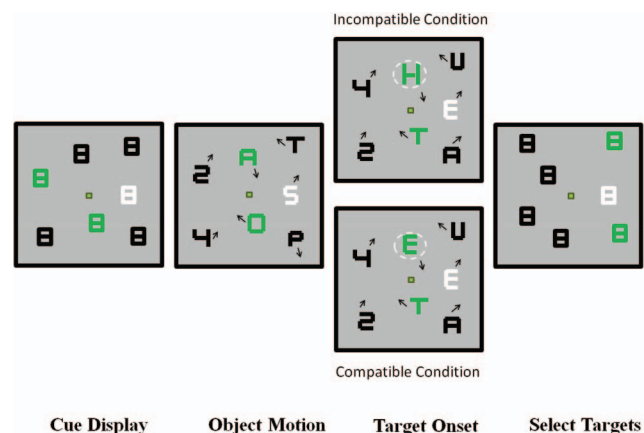


Figure 5. An example of a trial sequence from Experiment 3. In the dual task condition, observers made a speeded response to a target letter ("E" vs. "H") appearing in a stationary stream of letters (shown in white). They also had to track two moving objects (shown in green/light gray) among four moving distractors. Irrelevant streams of letters also appeared on the moving objects with the stationary stream. The onset of a target on the stationary object was accompanied by an irrelevant target appearing on either a tracked object or a distractor. These irrelevant targets were either compatible (e.g., "E" and "E") or incompatible (e.g., "E" and "H") with the relevant target. Colors of objects following the cue display are for illustration only. All objects were black. See the online article for the color version of this figure.

(see results). These participants were excluded from the experiment. The remaining 18 students (8 male, 10 female) were between 18 and 30 years old (mean age = 22.1 years old). All participants reported normal or corrected-to-normal visual acuity, and received monetary payment.

Stimuli and procedure. In this dual-task paradigm, participants were required to keep track of two of the six moving figure eight objects appearing in the display (see Figure 5). The objects were contained within a $12.4 \text{ dva} \times 9.8 \text{ dva}$ rectangular area centered on fixation. At the start of each trial, the to-be-tracked objects were presented in green for 500 ms. Objects then moved along linear trajectories in random directions with velocity adjusted based on a staircase algorithm described below. Objects bounced off the edges of the rectangular region and the stationary letter stream as well as each other when they were closer than 1.6 dva center-to-center. The objects also reversed directions when they reached an invisible circular region with a 2.3 dva radius that surrounded the fixation point to prevent it from being occluded by the moving objects.

The stationary letter stream was presented 3.3 dva to the left or right of the fixation point, with the location alternating across trials. A stream of random alphanumeric characters chosen from the set (A, C, J, L, O, S, U, 2, 4, 6, 7, 9) appeared in the stationary location and on each of the moving objects at a presentation duration of 400 ms per item. The stream of random characters appeared on each of the moving objects immediately at the start of each trial followed by the stream of characters in the stationary location at a delay of 200 ms. A target letter (E or H) appeared in the stationary letter stream at a random delay of 2 to 4 seconds (uniform distribution) relative to the beginning of the trial. Additionally, an irrelevant target letter (E or H) appeared in one of the streams corresponding to a tracked object or a distractor. This letter occurred randomly but equally often on tracked objects and distractors and was equally likely to be compatible or incompatible with the relevant target letter appearing at the fixed location.

Observers were instructed to make a speeded discrimination response to the target letter appearing in the fixed location and ignore targets appearing on the moving objects. Object movement ceased as soon as they made their response at which time they selected the target objects they were tracking by clicking on them with the mouse. They were provided with accuracy feedback after each choice. They were then provided with accuracy and RT feedback for the search task.

Object velocity was adjusted using a three up, one down staircase procedure designed to keep tracking performance close to 80% correct. This procedure occurred at the beginning of the experiment and between each experimental run if accuracy deviated more than 10% from the criterion of 80% correct. There were three runs of 3 blocks, each containing 48 trials for a total of 432 trials per participant. A monetary reward system was implemented to encourage participants to maintain high task performance. To emphasize the priority of the tracking task in the dual-task condition, participants had to be correct on tracking before they could earn a reward for the search task, and a larger reward was provided for correct tracking. Participants received 1 cent for every correct search response (earned only if the tracking response was correct) and double that amount for every correct tracking response.

The experiment consisted of separate blocks of trials devoted to three different tasks: single-task search, single-task tracking, and

dual-task (as described above). In the single-task search condition, observers were only required to identify the letter target appearing at the fixed location as an E or H. In the dual-task condition, the letter identification task was combined with the tracking task. For the single-task tracking condition, observers were only required to track two objects. Irrelevant letter streams were still presented in this condition to make displays comparable with those in the dual-task condition.

Results and Discussion

We restricted our analyses to trials without eye movements (15.43% of trials were rejected for eye movements). Outliers in the visual search task were defined as RTs that were more than three standard deviations from the mean for that participant. A total of 0.51% of the RTs were rejected as outliers. In the dual-task condition in which an irrelevant target letter appeared on a tracked object, two subjects had mean incompatible minus compatible RT difference scores on the search task that were more than three standard deviations from the mean (one was below the mean and the other was above). These subjects were excluded from further statistical analyses.

Tracking accuracy. Not surprisingly, tracking accuracy was significantly higher in the single-task tracking condition ($M = 91.13$) compared with the dual task condition ($M = 82.31$), $t(17) = -6.26$, $p < .001$. Tracking accuracy in the dual-task condition was analyzed with a 2 (Compatible vs. Incompatible irrelevant target) \times 2 (irrelevant target location: Tracked object vs. Distractor) repeated-measures ANOVA. There was no significant main effects of compatibility, $F(1, 17) = .31$, $p = .59$, or target location, $F(1, 17) = 3.2$, $p = .09$, or their interaction, $F(1, 17) = .25$, $p = .62$.

Search reaction time. The Search RT data are shown in Figure 6. To assess the compatibility effect on search RT as a function of target location we computed RT difference scores (incompatible minus compatible RT) for letters appearing on tracked objects versus distractors. RT difference scores were also computed for the single-task search condition in which task-irrelevant target letters could appear on any one of the moving

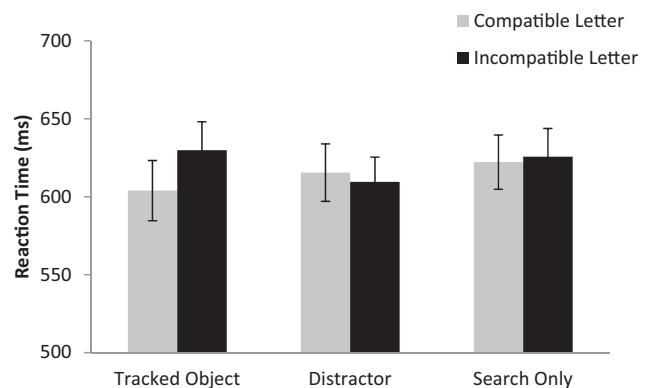


Figure 6. Search RT as a function of whether the irrelevant target occurred on a tracked object versus a distractor or in the search only condition for Experiment 3. The irrelevant target was compatible or incompatible with the relevant target occurring in the stationary letter stream. The error bars reflect the standard error of the mean.

objects. A one-way ANOVA revealed that the size of the difference scores varied across conditions (single-task search vs. dual-task search on tracked object vs. dual-task search on distractor), $F(2, 34) = 7.16, p = .003$. LSD contrasts showed that difference scores were significantly larger when targets occurred on tracked objects ($M = 25.9$ ms) relative to distractors ($M = -5.9$ ms; $p = .002$) and task-irrelevant moving objects in the single-task search condition ($M = 3.5$ ms; $p = .021$). RT in the single-task search condition was not significantly different from RT for distractors in dual-task search, $p = .29$.

The appearance of task-irrelevant incompatible letters only interfered with the search task when they occurred on tracked objects, but not when they occurred on distractors. Similarly, little interference was found when irrelevant targets occurred on unattended moving objects in the single-task search condition.

Search accuracy. Visual search accuracy is shown in Figure 7. Similar to the preceding analyses, we computed search accuracy difference scores (incompatible – compatible percent correct) for letters on tracked objects versus distractors, as well as for the single-task search condition in which participants did not track any objects. A one-way ANOVA showed that the difference scores were significantly different across the three conditions, $F(2, 34) = 6.26, p = .006$. LSD tests revealed a larger compatibility effect when letters appeared on tracked objects ($M = -5.62$) compared with distractors ($M = -1.25; p = .04$) and nontracked objects ($M = 1.16, p < .005$). Target identification on distractors in dual-task search was comparable to target identification on nontracked objects in the single-task search condition, $p = .18$.

The results of this experiment show that observers preferentially identified letters on tracked objects even though they were irrelevant and often harmful to the task of identifying letters in the stationary stream. Apparently, attention to tracked objects is not contingent on task-relevant targets appearing on them but is a necessary component of the tracking process.

Although the compatibility effects observed in this experiment are statistically significant and comparable with those observed in other flanker paradigms (Eriksen, 1995; Eriksen & Eriksen, 1974; Eriksen & Hoffman, 1973; Miller, 1988; Miller, 1991), one might have expected them to be much larger. In the typical flanker

paradigm, participants are trying to ignore the flanker letters whereas in this study, when the incompatible letter appears on a tracked object, it is presumably being attended for the purpose of tracking. The answer to this puzzle is that participants are probably allocating less attention to the tracked object than the stationary letter stream because, as suggested by the results of Experiment 1, accurate letter discrimination may require more attention than object tracking. This relationship could change, for example, if a distractor made a close approach to a target, in which case, one would presumably have to increase the attention allocated to the tracked object resulting in a greater flanker effect.

General Discussion

We first summarize the main results of our experiments and then relate them to current controversies in the area of MOT. The goal of our experiments was to determine whether MOT depends on a limited capacity resource that is preferentially allocated to tracked objects and whether that resource can be identified as visual attention. In the first two experiments, we measured attentional allocation during MOT by examining the accuracy of identifying probe letters appearing on tracked objects and distractors. In the first experiment where participants had to maintain high tracking accuracy in all conditions, we found large and robust performance advantages for probes appearing on tracked objects, regardless of whether probes were likely to occur on tracked objects or the distractor. We took this as evidence that maintaining high tracking accuracy requires participants to devote at least some minimum amount of attention to the tracked objects. In other words, tracking constrained the way in which attention was allocated to tracked objects and distractors. In the second experiment, the search task was emphasized and people were now able to achieve greater accuracy for probes on the distractor compared with the tracked objects but only at the cost of a sharp reduction in tracking accuracy (18%).

The third experiment examined whether people attend to tracked objects even when the probes appearing there are never task-relevant. We found that the identities of letters appearing on tracked objects affected the speed and accuracy of responding to the relevant letters. Importantly, letters appearing on distractors had no effect. These results show that when people track an object they pay attention to it even when the letters appearing on those objects are irrelevant and potentially harmful to their performance on the search task. Tracking appears to *require* visual attention.

We note that although a variety of studies have found similar advantages for probes appearing on tracked objects compared to distractors (Bahrami, 2003; Sears & Pylyshyn, 2000), not all studies report a significant effect. For example, Alvarez and colleagues (2005) used methods similar to those employed here to examine trade-offs between tracking accuracy and performance on a visual search task as the relative importance of the two tasks was systematically varied. Their results, however, appear to be different from ours. In three of their experiments they varied whether search targets appeared on tracked objects or distractors and reported accuracy advantages for targets of 18% (Experiment 2) and 17% (Experiment 3), as well as an RT advantage of 140 ms (Experiment 8), but none of these differences was statistically significant. We found generally similar size advantages for probes appearing on tracked objects in accuracy (accuracy advantage of

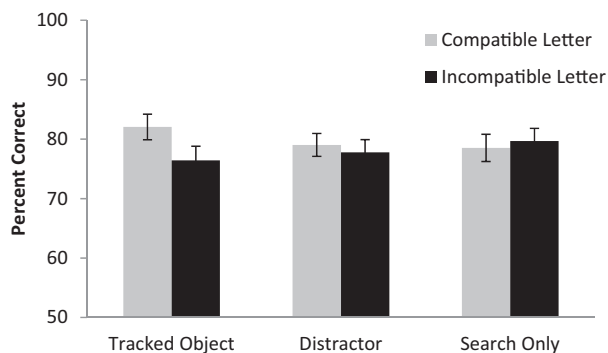


Figure 7. Visual search accuracy as a function of whether the irrelevant target occurred in the search-only condition or on a tracked object versus a distractor for Experiment 3. The irrelevant target was compatible or incompatible with the relevant target occurring in the stationary letter stream. The error bars reflect the standard error of the mean.

32% for set size 3 and 23% for set size 1 in Experiment 1), but these effects were highly significant. We can only speculate that search performance may have been more variable in the Alvarez et al. experiments because they used larger search arrays and set sizes. In any case, an examination of their effects and those reported in this paper show that participants identify shapes appearing on tracked objects more quickly and accurately than those on distractors.

Overall, these results make a strong case that MOT is dependent on a limited capacity resource that can be allocated to different objects in varying amounts. Allocation of attention to an object improves the ability to track it and also improves the ability to identify its shape. This pattern is inconsistent with the claim assumed by capacity-free models that tracking is limited only by a resource over which observers have no control. For example, Franconeri and colleagues (2010) suggested that tracking is primarily limited by the spacing between objects and this limit is set by the resolution of cortical receptive fields responding to the objects. According to this model, if the distance between a tracked object and a distractor is less than the spatial resolution of the receptive fields responding to objects in that area of the retina, then observers may fail to individuate the two nearby objects and may end up tracking the distractor (see also Bae & Flombaum, 2012; Intriligator & Cavanagh, 2001; O’Hearn, Hoffman, & Landau, 2010). They showed that this model is capable of accounting for many of the variables that are known to affect tracking accuracy, such as the number of objects to be tracked, spacing, and velocity (Franconeri, Jonathan, & Scimeca, 2010). For example, increasing the number of objects to be tracked decreases tracking accuracy because increases in the number of tracked objects increases the probability that at least one of them will pass close enough to a distractor to produce a tracking error.

These spacing effects surely play an important role in tracking but they cannot be the whole story. For example, it wouldn’t account for the present findings showing that even when display spacing is constant, participants are still able to vary how much attention they allocate to tracked objects and this allocation can have strong effects on tracking accuracy (our Experiment 2; see also Holcombe, Chen, & Howe, 2014 who showed that set size affects tracking accuracy even with large spacing that should minimize deleterious spatial interactions). Similarly, maintaining high tracking accuracy constrains how much attention can be allocated to other task-relevant objects in the display (Experiment 1).

Our findings are consistent with the multifocal attention theory proposed by Cavanagh and Alvarez (2005) as well as Srivastava and Vul’s (2016) computational model of tracking. In these models, increased visual attention can be allocated to tracked objects to improve the spatial resolution of a target’s representation, but this occurs at the cost of decreased resolution for other targets. These various findings suggest that visual attention is a limited-capacity resource and that increasing set size reduces the amount of attention available to each tracked object. This reduction in attention eventually results in increased tracking errors.

We have emphasized that visual attention is the limited resource used for object tracking. However, we need to also consider other potential limited-capacity mechanisms that might play a role. For example, Pylyshyn (1989) proposed that tracking is accomplished by assigning pointers or indexes to tracked objects. Both indexing

and visual attention can account for selection of objects for tracking but they differ in terms of the information they provide about the tracked objects. According to Pylyshyn (1989), an index is a preattentive mechanism that essentially “points to” a relevant object without necessarily providing any information about its features or identity. In addition, indexes differ from visual attention in that an index is a discrete resource and cannot be divided while visual attention is assumed to be a resource that can be continuously divided between different objects. Our results support the visual attention interpretation as opposed to indexing. Tracking an object appears to involve attending to it as shown by the improvement in the ability to identify shapes appearing on the tracked object compared to distractor objects, a result not directly predicted by the indexing account which holds that objects can be tracked without necessarily paying attention to them.

Furthermore, the indexing model cannot easily explain the results of our last experiment that showed that people allocate visual attention to a tracked object even when the letters appearing there are not task relevant and in fact can be harmful to performance of the relevant task. Attending to an object appears to be an obligatory part of tracking it, even when it isn’t advantageous to do so. However, later versions of the indexing model may be able to account for these results as well. Pylyshyn (1994) proposed a hybrid model in which indexes are the principal mechanism supporting object tracking while visual attention plays a “helper role.” He suggested that attention might be necessary to “refresh” the indexes, which would otherwise decay. In addition, visual attention might be involved in maintaining an index on the correct object when distractors are nearby as well as during “error recovery” when indexes are temporarily lost during tracking. Given the somewhat nebulous account of the circumstances in which attention might intervene to maintain tracking, this particular hybrid model appears to be largely indistinguishable from the multifocal attention model of Cavanagh and Alvarez (2005). In a recent review, Scholl (2009) concluded that there isn’t any independent evidence for an indexing system that is distinct from visual attention, and that parsimony favors postulating a single mechanism (visual attention) responsible for MOT. That conclusion is consistent with our results that show the involvement of a continuous resource that is divisible among different objects and is involved in object identification. As Scholl (2009) recently put it: “there may be nothing to MOT beyond attention” (p. 55). In any case, our results cannot refute the possibility that tracking is accomplished through a combination of indexing and visual attention. They do, however, show that visual attention plays a critical role in object tracking (see Iordanescu, Grabowecy, & Suzuki, 2009 and Horowitz & Cohen, 2010 for additional evidence that a continuous resource plays a role in object tracking).

References

- 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, 1–10. <http://dx.doi.org/10.1167/7.13.14>
- Alvarez, G. A., Horowitz, T. S., Arsenio, H. C., Dimase, J. S., & Wolfe, J. M. (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology: Human Perception and Performance*, 31, 643–667. <http://dx.doi.org/10.1037/0096-1523.31.4.643a>

- Bae, G. Y., & Flombaum, J. I. (2012). Close encounters of the distracting kind: Identifying the cause of visual tracking errors. *Attention, Perception, & Psychophysics*, *74*, 703–715. <http://dx.doi.org/10.3758/s13414-011-0260-1>
- Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition*, *10*, 949–963. <http://dx.doi.org/10.1080/13506280344000158>
- Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*, 1–11. <http://dx.doi.org/10.1167/9.10.7>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. <http://dx.doi.org/10.1163/156856897X00357>
- Cardinal, R. N., & Aitken, M. R. F. (2006). *ANOVA for the behavioral sciences researcher*. Mahwah, NJ: Erlbaum Inc.
- Carrasco, M., & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *98*, 5363–5367. <http://dx.doi.org/10.1073/pnas.081074098>
- Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*, 349–354. <http://dx.doi.org/10.1016/j.tics.2005.05.009>
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*, 613–617. <http://dx.doi.org/10.3758/BF03195489>
- Doran, M. M., & Hoffman, J. E. (2010). The role of visual attention in multiple object tracking: Evidence from ERPs. *Attention, Perception & Psychophysics*, *72*, 33–52. <http://dx.doi.org/10.3758/APP.72.1.33>
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143–149. <http://dx.doi.org/10.3758/BF03203267>
- Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition*, *2*, 101–118. <http://dx.doi.org/10.1080/13506289508401726>
- Eriksen, C. W., & Hoffman, J. E. (1972a). Some characteristics of selective attention in visual perception determined by vocal reaction time. *Perception & Psychophysics*, *11*, 169–171. <http://dx.doi.org/10.3758/BF03210367>
- Eriksen, C. W., & Hoffman, J. E. (1972b). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics*, *12*, 201–204. <http://dx.doi.org/10.3758/BF03212870>
- Eriksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selecting encoding from visual displays. *Perception & Psychophysics*, *14*, 155–160. <http://dx.doi.org/10.3758/BF03198630>
- Franconeri, S. L. (2013). The nature and status of visual resources. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 1–16). New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/oxfordhb/9780195376746.013.0010>
- Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences*, *17*, 134–141. <http://dx.doi.org/10.1016/j.tics.2013.01.010>
- Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can be selected at once? *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1003–1012. <http://dx.doi.org/10.1037/0096-1523.33.5.1003>
- 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. <http://dx.doi.org/10.1177/0956797610373935>
- Hoffman, J. E., & Nelson, B. (1981). Spatial selectivity in visual search. *Perception & Psychophysics*, *30*, 283–290. <http://dx.doi.org/10.3758/BF03214284>
- Holcombe, A. O., Chen, W.-Y., & Howe, P. D. L. (2014). Object tracking: Absence of long-range spatial interference supports resource theories. *Journal of Vision*, *14*, 1–21. <http://dx.doi.org/10.1167/14.6.1>
- Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception, & Psychophysics*, *72*, 1765–1775. <http://dx.doi.org/10.3758/APP.72.7.1765>
- Howell, D. C. (2013). *Fundamental statistics for the behavioral sciences* (8th ed.). Independence, KY: Cengage Learning.
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*, 171–216. <http://dx.doi.org/10.1006/cogp.2001.0755>
- 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*, 1–12. <http://dx.doi.org/10.1167/9.4.1>
- Keshvari, S., van den Berg, R., & Ma, W. J. (2013). No evidence for an item limit in change detection. *PLoS Computational Biology*, *9*, e1002927. <http://dx.doi.org/10.1371/journal.pcbi.1002927>
- Miller, J. (1988). Response-compatibility effects in focused-attention tasks: A same-hand advantage in response activation. *Perception & Psychophysics*, *43*, 83–89. <http://dx.doi.org/10.3758/BF03208977>
- Miller, J. (1991). The flanker compatibility effect as a function of visual angle, attentional focus, visual transients, and perceptual load: A search for boundary conditions. *Perception & Psychophysics*, *49*, 270–288. <http://dx.doi.org/10.3758/BF03214311>
- O’Hearn, K., Hoffman, J. E., & Landau, B. (2010). Developmental profiles for multiple object tracking and spatial memory: Typically developing preschoolers and people with Williams syndrome. *Developmental Science*, *13*, 430–440. <http://dx.doi.org/10.1111/j.1467-7687.2009.00893.x>
- Oksama, L., & Hyönä, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology*, *56*, 237–283. <http://dx.doi.org/10.1016/j.cogpsych.2007.03.001>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. <http://dx.doi.org/10.1163/156856897X00366>
- Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*, 3–25. <http://dx.doi.org/10.1080/00335588008248231>
- Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*, 65–97. [http://dx.doi.org/10.1016/0010-0277\(89\)90014-0](http://dx.doi.org/10.1016/0010-0277(89)90014-0)
- Pylyshyn, Z. (1994). Some primitive mechanisms of spatial attention. *Cognition*, *50*, 363–384. [http://dx.doi.org/10.1016/0010-0277\(94\)90036-1](http://dx.doi.org/10.1016/0010-0277(94)90036-1)
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197. <http://dx.doi.org/10.1163/156856888X00122>
- Scholl, B. J. (2009). What have learned about attention from multiple object tracking (and vice versa)? In D. Dedrick & L. Trick (Eds.), *Computation, cognition, and Pylyshyn* (pp. 49–78). Cambridge, MA: MIT Press.
- Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *54*, 1–14. <http://dx.doi.org/10.1037/h0087326>
- Sperling, G., & Melchner, M. J. (1978). The attention operating characteristic: Examples from visual search. *Science*, *202*, 315–318. <http://dx.doi.org/10.1126/science.694536>
- Srivastava, N., & Vul, E. (2016). Attention modulates spatial precision in multiple-object tracking. *Topics in Cognitive Science*, *8*, 335–348. <http://dx.doi.org/10.1111/tops.12189>

- Szinte, M., Carrasco, M., Cavanagh, P., & Rolfs, M. (2015). Attentional trade-offs maintain the tracking of moving objects across saccades. *Journal of Neurophysiology*, *113*, 2220–2231. <http://dx.doi.org/10.1152/jn.00966.2014>
- Theeuwes, J., & Van der Burg, E. (2007). The role of spatial and nonspatial information in visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1335–1351. <http://dx.doi.org/10.1037/0096-1523.33.6.1335>
- Tran, A., & Hoffman, J. E. (2013, May). Visual attention is required for multiple object tracking. Poster session presented at the meeting of Vision Sciences Society, Naples, FL.
- Vul, E., Frank, M. C., Alvarez, G., & Tenenbaum, J. B. (2010). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. Paper presented at the Neural Information Processing Systems 2010.
- Wang, P., & Nikolić, D. (2011). An LCD monitor with sufficiently precise timing for research in vision. *Frontiers in Human Neuroscience*, *5*, 85. <http://dx.doi.org/10.3389/fnhum.2011.00085>
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*, 1120–1135. <http://dx.doi.org/10.1167/4.12.11>
- Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, *24*, 295–340. [http://dx.doi.org/10.1016/0010-0285\(92\)90010-Y](http://dx.doi.org/10.1016/0010-0285(92)90010-Y)
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235. <http://dx.doi.org/10.1038/nature06860>
- Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science*, *22*, 1434–1444. <http://dx.doi.org/10.1177/0956797611417006>

Received July 16, 2015

Revision received May 5, 2016

Accepted May 5, 2016 ■