How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism

George A. Alvarez
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology,
Cambridge, MA, USA

Steven L. Franconeri
Department of Psychology Northwestern University,
Evanston, IL, USA

Much of our interaction with the visual world requires us to isolate some currently important objects from other less important objects. This task becomes more difficult when objects move, or when our field of view moves relative to the world, requiring us to track these objects over space and time. Previous experiments have shown that observers can track a maximum of about 4 moving objects. A natural explanation for this capacity limit is that the visual system is architecturally limited to handling a fixed number of objects at once, a so-called magical number 4 on visual attention. In contrast to this view, Experiment 1 shows that tracking capacity is not fixed. At slow speeds it is possible to track up to 8 objects, and yet there are fast speeds at which only a single object can be tracked. Experiment 2 suggests that that the limit on tracking is related to the spatial resolution of attention. These findings suggest that the number of objects that can be tracked is primarily set by a flexibly allocated resource, which has important implications for the mechanisms of object tracking and for the relationship between object tracking and other cognitive processes.

Keywords: multiple object tracking, MOT, multi-element tracking, attentive tracking, FINST, FLEX, attention, capacity

Citation: Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 0*(0):1, 1–10, http://journalofvision.org/0/0/1/, doi:10.1167/0.0.1.

Introduction

Tracking moving objects over space and time is a fundamental part of making sense of a dynamic visual world. Whether driving on a busy highway, playing team sports, or watching one's children at the playground, one often maintains attention on multiple moving objects simultaneously. To explore this ability in the laboratory, researchers have employed the multiple object tracking task (Pylyshyn & Storm, 1988). Typically, a set of identical items is presented and a subset of target items is cued, then all items move randomly about the screen for several seconds. During this time, all of the items appear identical and the eyes can only fixate directly on one target at a time. Thus, to track multiple targets concurrently, observers are required to "mentally track" the target items as they move about the display. At the end of the trial, all of the items stop and the observer must indicate which items were the original targets.

Studies employing this task have been used to investigate a wide range of topics in visual cognition, including determining what counts as an object for object-based attention (Scholl & Pylyshyn, 1999; Scholl, Pylyshyn, & Feldman, 2001), the dynamics of attention in depth (Viswanathan & Mingolla, 2002), the coordinate systems underlying attention (Liu et al., 2005), the limits on divided or multifocal attention (Alvarez, Horowitz, Aresenio, DiMase, & Wolfe, 2005; Cavanagh & Alvarez, 2005), age differences in attention (Trick, Audet, & Dales, 2003), and deficits in attention for different patient populations (Ho et al., 2006; O'Hearn, Landau, & Hoffman, 2005).

Given the broad range of work that employs the multiple object-tracking task, it is important to understand the nature of limits on tracking at a basic level. In the current paper, we investigate whether the limit on the number of objects that can be tracked is fixed (the *fixed-architecture model*), or whether the limit on tracking is set by a resource that can be flexibly allocated to objects depending on the demands of the task (the *flexible-resource model*).

The argument for the fixed-architecture model

Surprisingly, across multiple studies, researchers have consistently found that approximately 4 objects can be tracked (Intriligator & Cavanagh, 2001; Pylyshyn & Storm, 1988; Yantis, 1992). The similarity of these estimates, combined with the frequency with which 4-item limits arise in other attention tasks, suggests the possibility that there is a "magical number 4" in visual attention (Cowan, 2001; Pylyshyn, 1989). This 4-item limit implies an architectural constraint on multiple object tracking. That is, there appears to be a fixed number of mechanisms

doi: 10.1167/0.0.1 Received March 15, 2007; published 0 0, 2007 ISSN 1534-7362 © ARVO

used for tracking, and the number of these mechanisms sets the limit on the number of objects that can be tracked. These mechanisms could take the form of "FINSTs" (which "stick" to objects; Pylyshyn & Storm, 1988) or object files (which track objects via spatiotemporal information; Kahneman, Treisman, & Gibbs, 1992; Mitroff & Alvarez, in press).

The argument for the flexible-resource model

While the apparently high agreement in capacity estimates across studies suggests there exists a fixed number of tracking mechanisms, the data are by no means conclusive. There is a great deal of variability in the tracking capacity across individuals (Oksama & Hyona, 2004), expertise can increase the number of objects tracked (Allen, McGeorge, Pearson, & Milne, 2004), playing video games increases the number of objects that can be tracked (Green & Bavelier, 2006), and grouping targets into a virtual polygon improves tracking accuracy (Yantis, 1992). While it is conceivable that different individuals would be born with different numbers of tracking mechanisms, explaining individual differences and expertise effects, it is less clear how playing video games or using a grouping strategy would increase the number of tracking mechanisms a particular individual has. Thus, it is worth considering alternatives to the fixed-architecture view, such as an attentional resource theory (Allen et al., 2004; Yantis, 1992).

A resource theory would hold that there is a pool of resources required for tracking objects, and that the limit on tracking depends on the resource demands required to track each object. For example, if the tracking task were so difficult that tracking one target consumed all available tracking resources, then only a single item could be tracked. However, if each item only required 1/4th of the total available resources, then four objects could be tracked. Thus, the number of objects that could be tracked would be inversely related to the resource demands for each individual object.

The fixed-architecture model and the flexible-resource model present a fundamental division between potential tracking mechanisms. Thus, interpreting the results of studies employing the multiple object-tracking task will be influenced by which theory best explains limits on this task. Beyond object tracking, describing the visual system's mechanisms for maintaining attention on moving objects is critical to understanding broader phenomena, such as spatial vision and imagery (Pylyshyn, 1989; Pylyshyn, 1998), our stable percept of the visual world across eye and body movements (Pylyshyn, 1989), the development of object knowledge in infants (e.g., Carey & Xu, 2001; Leslie, Xu, Tremoulet, & Scholl, 1998), and the development and operation of our numerical concepts (Carey & Xu, 2001). Distinguishing between these alternate models of the limits on multiple object tracking would inform a variety of problems within vision and within cognitive psychology more generally. In the current study, we investigate whether the tracking limit is set by a fixed number of tracking mechanisms, or by a resource limitation.

Experiment 1: Evidence for a resource limit on tracking

We used object speed to manipulate the demands of the tracking task and to determine whether the number of objects that can be tracked is fixed, or whether there is a tradeoff between the difficulty of tracking targets and the number that can be tracked. We asked observers to track 1 to 8 objects, and we estimated the maximum speed at which they could perform the task (see Figure 1a). If there is a fixed number of independently functioning tracking mechanisms, and only the number of tracking mechanisms imposes a limit on tracking, then the maximum tracking speed should be the same from 1 to *N* targets,

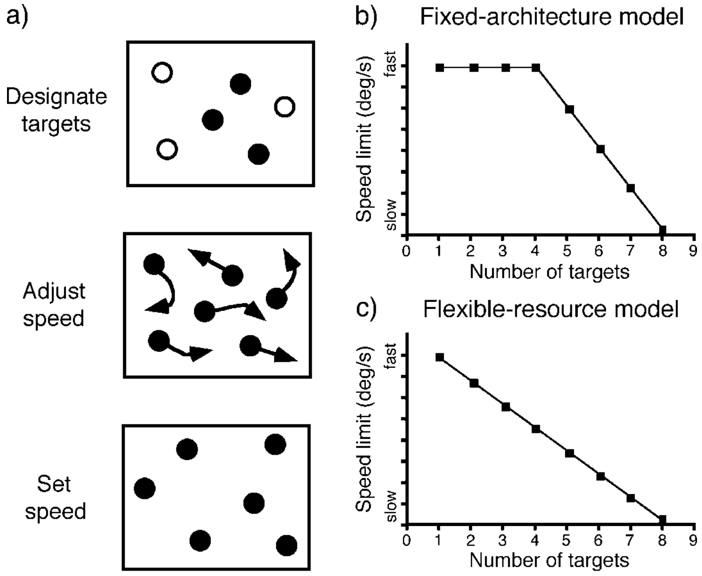

Figure 1. Task and predictions for Experiment 1. (a) A schematic depiction of the tracking task in Experiment 1. At the beginning of each trial, a subset of items were identified as targets. Then all items appeared identical and observers adjusted the speed to the maximum at which they could perfectly track the items for about 5 s. The trial ended when the observer selected a speed. The accuracy of these speed limit settings was verified in a separate session. (b) The *fixed-architecture model* predicts that the speed limit will be the same from 1 to *N,* where *N* is the number of tracking mechanisms available (shown as 4 here) and then will decline beyond that point. (c) The *flexible-resource model* predicts that with each increase in the number of targets the speed limit will decrease.

where N is the number of tracking mechanisms (see Figure 1b). In contrast, if tracking capacity is limited by some flexible resource, then as the number of targets tracked increases, the amount of this resource allocated to each individual object will decrease. Assuming the maximum speed at which an object can be tracked depends on the amount of resource devoted to that object, the speed limit should decrease as the number of targets increases (see Figure 1c). We depict a linear tradeoff in Figure 1c, but the function need only be monotonically decreasing.

Method

Participants

Fourteen observers reported normal or corrected-to-normal vision, gave informed consent and were paid or received course credit.

Stimuli

Sixteen green circles (diameter 1.25°) were presented on a black background (30° × 24°). A gray fixation point ("+") subtending 1° × 1° was presented at the center of the display. The circles moved at a constant speed (between 0°/s and 42°/s) and were "repelled" by each edge of the display and by other items with decreasing strength over distance, such that the items "avoided" each other. The circles changed direction to avoid other items and were never closer than 4° (center to center) to another circle.

Procedure

There were two sessions: a speed limit session where observers would estimate their top tracking speeds for each number of targets, and an accuracy check session, which would confirm whether their estimates were correct. In the speed limit session, each trial began with the presentation of 16 circles with a subset of green distractor circles and red target circles. Once observers noted the red subset they pressed the down arrow key to "hide" the targets (they turned green and appeared identical to the other circles on the screen). Then observers adjusted the speed of the circles by pressing the arrow keys (left arrow to slow down, right arrow to speed up). Observers were instructed to increase the speed until they found that they were moving too fast to track. At that point, observers were instructed to decrease the speed, and then press the up arrow to "show" the targets again (the original target set was turned red again). Observers were instructed to repeat this procedure a few times until they reached the maximum speed at which they could perfectly track all of the targets for about 5 s. Once the observers were confident they had found their speed limit, they pressed the space bar to enter their setting. They were then prompted to confirm their selection, and then the next trial began. If they could not track the number of targets required observers were instructed to set the speed to zero (stationary). This procedure was repeated 3 times each for 1 to 8 targets, for a total of 24 settings.

In the second session, observers performed a tracking task with the speed set to their personal speed limit for 1 to 8 targets. At the beginning of each trial, 1 to 8 targets were highlighted in red, and then all of them turned green. The items then moved for 6 s at the observer's speed limit setting for that number of targets. At the end of the trial, all of the circles stopped moving, and then randomly one of the circles turned red (half of the time it was a target and half of the time it was a distractor). The task was to indicate whether the red item was one of the targets, or one of the distractors by pressing left arrow key to indicate "target" and the right arrow key to indicate "distractor." Critically, this probe method equates response demands and chance performance (50%) across all numbers of targets. Observers completed a total of 80 trials in this accuracy check session.

Although eye movements were not monitored, observers were informed that our primary interest was in how well they could track objects by paying attention to them in their peripheral vision, rather than by moving their eyes around to follow them and were asked to keep their eyes focused on the central "+" throughout the experiment.

Results

Data for two observers were discarded because their error rates in the tracking task (averaged across numbers of targets) were about 3 standard deviations above the mean. Analysis of speed limit settings and tracking accuracy was performed for the remaining 12 observers.

Speed settings

Figure 2a illustrates the average speed limit setting as a function of the number of targets. There appears to be a continuous function relating the number of targets tracked to the speed limit. The speed limit decreased significantly with each increase in the number of targets (1 vs. 2, $t(11) = 5.7$, $p < .001$; 2 vs. 3, $t(11) = 5.1$, $p < .001$; 3 vs. 4, $t(11) = 5.2$, $p < .001$; 4 vs. 5, $t(11) = 8.0$, $p < .001$; 5 vs. 6, $t(11) = 6.7$, $p < .001$; 6 vs. 7, $t(11) = 5.9$, $p < .001$; 7 vs. 8, $t(11) = 3.0$, $p < .05$). Although we had no a priori expectation for what the shape of the speed limit × number of targets function would be, upon inspection it appeared logarithmic. We plotted the speed limit versus the log of the number of targets (see Figure 2b) and found a strong linear correlation ($r^2 = .996$) and

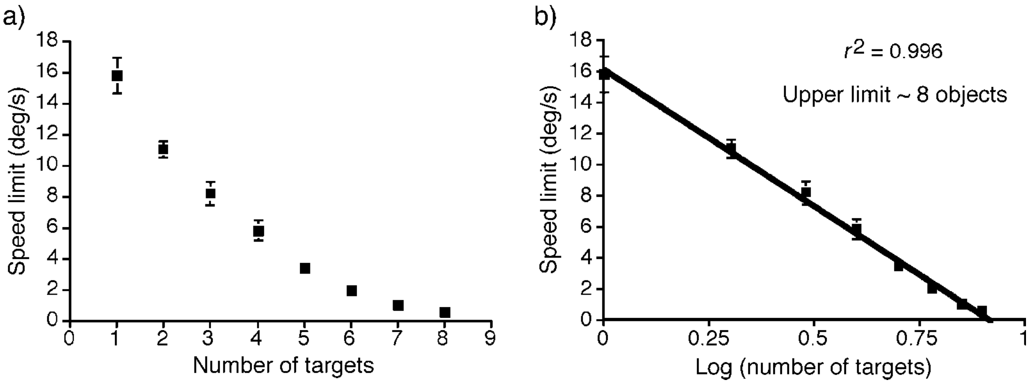

Figure 2. Results of Experiment 1. (a) Estimated speed limit in degrees per second as a function of the number of targets in Experiment 1. Error bars are presented where they are larger than the data symbols and represent one standard error of the mean. (b) Plotting the estimated speed limit as a function of the log of the number of targets shows a strong correlation and a maximum upper limit of about 8 on the number of objects that can be tracked.

extrapolating this function to speed zero suggests an upper limit on tracking capacity of about 8 objects.

Accuracy check

Observers accurately estimated their personal speed limits for tracking different numbers of targets. Tracking accuracy was high (~94% overall) and did not vary as a function of the number of targets when the speed was set to each individual observer's speed limit for each number of targets ($F(7, 77) < 1$, $p = .53$). None of the t-tests comparing accuracy for different numbers of targets were significant (the comparison for 2 vs. 3 targets, approached significance at $p = .053$, but none of the other 27 comparisons were significant, with uncorrected p values greater than .11 for each comparison).

Discussion

The results of this experiment show that with each increase in the number of targets tracked, there is a decrease in the maximum speed at which those targets can move and still be tracked accurately. For example, increasing the number of tracked targets from 1 to 2 decreased the speed limit by 30%. If the allocation of attention to an object were set to a certain fixed amount, then the speed limit would not change when the number of tracked targets increases (assuming the capacity limit was greater than one, see Figure 1b). The gradual decrease in speed limit with the number of targets tracked is inconsistent with a fixed-architecture model that assumes number of objects tracked is limited primarily by a fixed number of independent tracking mechanisms. However, the results are consistent with a flexible-resource model that assumes attention can be flexibly allocated to tracked objects. When 1 object is tracked, all resources are devoted to that one target and it can be tracked at a fast speed. When 2 objects are tracked, resources are divided among the targets, and the speed limit is reduced. In general, as the number of targets increases, the amount of resource devoted to each object decreases, reducing the maximum speed of tracking.

The accuracy of these speed limit settings was verified in a block of trials in which participants tracked 1–8 targets at their own personal speed limit settings. The average accuracy was 94% and did not vary as a function of the number of targets, suggesting that the speed measurements accurately reflect the maximum speed at which participants can track all of the targets.

The subjective experience of trying to track a large number of objects (e.g., 4) at a very fast speed (e.g., the speed limit for 1 item) is quite compelling: as soon as the targets begin to move, they "scatter" and are completely untrackable. In fact, it seems that if one tries to track all 4 targets, they will all be lost. For readers interested in observing this result first hand, we have posted a demonstration online at http://cvcl.mit.edu/george/demos.htm. While these online displays have fewer items than in Experiment 1, they nevertheless provide a clear demonstration of this effect.

We interpret these results as evidence for a resource limit on the number of objects that can be tracked. This conclusion rests on an important distinction between processes that are primarily data-limited and those that are primarily resource-limited (Norman & Bobrow, 1975). For example, if the task was to identify a letter among white noise, the task could become impossible simply because there is not enough signal in the noise, even with 100% of the available resources devoted to the task. In general, when the quality of the data is the primary limit on performance, devoting more resources to that task will not improve performance. It is important to note that the difficulty in tracking multiple objects at fast speeds in the current study cannot be attributed to data limitations. For any individual, there is a fast speed at which a single target can be tracked accurately without errors, but no more than one object can be tracked at that speed. The fact that one target can be tracked indicates that the quality of the image data is sufficient to support accurate tracking. The failure to track more than one object at such high speeds must therefore result from a lack of available attentional resources.

The current results also constrain any "hybrid" model that assumes there is both a fixed number of tracking mechanisms and a resource limit on tracking. According to such a hybrid account, there should be a decrease in the speed limit from 1 to N targets because less resource is available to each tracker as the number of targets increases. Beyond N, there should be a breakdown in performance because the number of targets exceeds the number of tracking mechanisms. At best, tracking is aided by an "offline" spatial memory that is much less effective than the "online" continuous operations of the tracking mechanisms. However, there is no evidence for such a discontinuity in the function relating the speed limit to the number of targets tracked. Thus, any such hybrid model would have to be modified to account for the continuous transition from the online tracking system to the offline spatial memory system. While it is difficult to rule out all classes of hybrid models, the important point for our purposes is that any hybrid account would have to include a resource-limited component that acts as the *primary* determinant of the number of objects that can be tracked.

Experiment 2: Attentional resolution limits

If the limit on the number of objects that can be tracked is set primarily by a flexibly allocated resource, then it is important to understand the role this resource plays in tracking. What are the advantages of allocating more tracking resources to an object? Previous researchers have proposed that attention refreshes tracking indexes to overcome decay or interference (Pylyshyn et al., 1994), facilitates tracking through anticipation or error recovery (McKeever & Pylyshyn, 1993), or maintains a higher order object representation (a "virtual polygon") to facilitate tracking (Yantis, 1992). In addition to these factors, we propose that the allocation of attention affects the spatial resolution with which information is represented (Yeshurun & Carrasco, 1998), and that spatial resolution imposes important constraints on multiple object tracking (Intriligator & Cavanagh, 2001).

Previous research has shown that the number of locations that spatial attention can select at once depends on the precision required to isolate target locations from distractor locations (Franconeri, Alvarez, & Enns, in press). When the spacing between items was small, requiring precise selection regions, only 2–3 locations could be selected. But when the spacing between items was large, allowing selection regions to be coarser, up to 6–7 locations could be selected. This suggests there is a tradeoff between the number of items selected, and the precision with which those items can be selected: The greater the number of items selected, the coarser the selection.

The tradeoff between the number of items selected and the spatial precision of the selection can explain why there is a limit to the number of objects that can be tracked at a particular speed. On this view, when a single item is tracked, its position can be selected very precisely because all resources are devoted to tracking that one item. As more objects are tracked, each item's position must be selected more coarsely. Eventually, increasing the number of objects tracked will result in such coarse selections that distractors will fall within the selected region and become confused with targets, leading to a decrease in performance. Thus, the maximum allowable window of selection around the target (which depends on how close the targets are allowed to come to distractors), will set the limit on the number of objects that can be tracked.

With an additional assumption, we can also explain the speed limit on tracking observed in Experiment 1 in terms of spatial resolution. Specifically, if we assume that faster moving objects require a coarser selection window than slow moving objects, then increasing the speed should decrease the number of objects that can be tracked. This hypothesis is based in part on the relationship between velocity sensitivity and receptive field sizes, which are positively correlated in the cat and monkey, such that receptive fields of cells tuned to faster speeds tend to be larger (Mikami, Newsome, & Wurtz, 1986; Orban, Kennedy, & Bullier, 1986; Orban, Kennedy, & Maes, 1981). Attentive tracking most likely relies on inputs from such motion sensitive mechanisms, and thus it is possible that tracking faster moving objects relies on a spatially coarser representation than tracking slower moving objects.

Thus, we propose a resolution-based account for the resource limit on tracking accuracy with two important claims: (1) the more items that are tracked, the coarser the selection; and (2) the faster the tracked items move, the coarser the selection. In the current experiment, we varied the required resolution of selection by varying the minimum spacing between items. Our hypothesis predicts that the number of objects that can be tracked will decrease as the spacing between targets and distractors decreases because a more precise selection window is required. We also predict that the cost for decreasing the spacing between targets and distractors will be greater for fast moving targets than for slow moving targets because selection regions are necessarily coarser for fast items than for slow items. Alternatively, it is possible that there is a fixed resolution limit, a lower bound on the resolution of attention (Intrilgator & Cavanagh, 2001), and that this will be the same for fast and slow targets.

Method

Participants

Twelve observers reported normal or corrected-to-normal vision, gave informed consent, and were paid for their participation.

Stimuli

Eight black circles (diameter = 0.67°) were presented on a gray background (23° × 23°). The number of targets was fixed at 4, the speed was either slow (7°/s) or fast (14°/s), and the minimum spacing between items varied (0.67°–4.67°, in 1° intervals). As in Experiment 1, the items repelled each other to avoid collisions and bounced off of the edges of the display to remain on the screen.

Procedure

At the beginning of each trial, 8 items were presented, and a subset of 4 items blinked off and on at 2 Hz for 2 s to designate them as targets for the tracking task. Then all of the items moved at a constant rate for 12 s and stopped. Participants used the mouse to highlight and click on the 4 target items. Participants completed 8 trials for each combination of speed (slow and fast) and the 5 minimum spacings between items (0.67°–4.67°), with the order of conditions randomized.

Results

Tracking accuracy was more sensitive to the spacing between items when the items moved at a fast speed than when they moved at a slow speed (see Figure 3a). A 2 × 5 ANOVA on tracking accuracy with speed and minimum spacing as factors showed a significant main effect of speed ($F(1, 11) = 62.8$, $MSE = 54.7$, $p < .001$, $\eta_p^2 = 0.85$), indicating that tracking was more accurate for slow moving targets than fast moving targets. There was also a significant main effect of spacing ($F(4, 44) = 23.1$, $MSE = 24.5$, $p < .001$, $\eta_p^2 = 0.68$), indicating the tracking accuracy was higher the more widely spaced the items were. Most importantly, there was a significant interaction between speed and minimum spacing ($F(4, 44) = 5.9$, $MSE = 24.2$, $p < .001$, $\eta_p^2 = 0.35$), indicating that the crowding effect of the distractors was greater for fast moving targets than for slow moving targets (the drop in accuracy for the smallest spacing compared to the largest spacing was 18.4% for fast targets, and 5.9% for slow targets).

The interaction does not appear to be due to the general difficulty of tracking the faster targets. Although there was a trend for better tracking accuracy at the slower speed for each spacing, at the largest spacing tracking was high for both speeds and not significantly different (slow, $M = 89.8\%$, $SEM = 3.4\%$; fast, $M = 93.5\%$, $SEM = 3.9\%$; $t(11) = 2.08$, $p = .062$, $r^2 = .28$). At all smaller spacings, the difference in tracking accuracy for slow and fast targets was significant (all p values <.05, all r^2 values greater than .44).

To estimate the number of objects tracked as a function of speed and the minimum spacing between items, we used the following equation:

$$P(\text{correct}) = [C + (n - C) * (n - C)/(m - C)]/n. \quad (1)$$

Where P(correct) is the average proportion of targets accurately clicked, C is the number of targets actually tracked, n is the number of targets, and m is the total number of items in the display. An example illustrates the logic of this equation. Say an observer is asked to track 4 out of 8 items, but is only able to actually track 3 of the targets. We can assume that the subject will click on the 3 tracked targets but will then guess for the remaining 1 target among 4 distractors (a 20% chance of

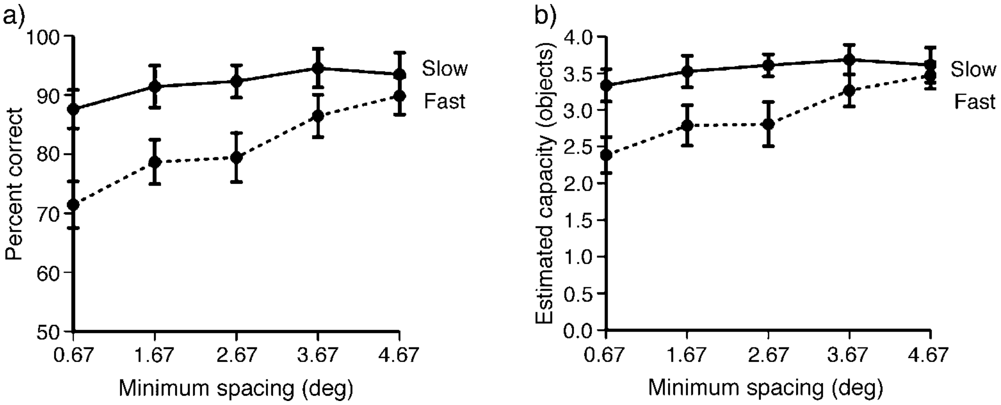

Figure 3. Results of Experiment 2. (a) Decreasing the minimum spacing between items decreased tracking accuracy more when the items move at a fast speed than when they move at a slow speed. (b) Results in terms of tracking capacity (the number of objects tracked) reveal that the number of objects that can be tracked decreases as the minimum spacing decreases.

correctly guessing). On average, this observer would click on 3 + (4 − 3) * (4 − 3) / (8 − 3) = 3.2 targets out of 4, yielding a proportion correct of .80 on average.

Figure 3b shows the results in terms of estimated number of objects tracked. As the minimum spacing decreases from 4.67° to 0.67°, the number of objects that can be tracked at a slow speed drops a small but significant amount from a mean of 3.6 ± 0.3 objects to 3.3 ± 0.2 objects ($t(11) = 4.12$, $p < .01$, $r^2 = .61$). At a fast speed, the drop was even greater, from a mean of 3.5 ± 0.2 objects to a mean of 2.4 ± 0.3 objects ($t(11) = 6.51$, $p < .001$, $r^2 = .79$). The difference in tracking capacity for fast and slow moving targets was not significant at the at the largest spacing of 4.67° ($t(11) = 1.31$, $p = .215$, $r^2 = .14$), but was significant at the smaller spacing of 0.67° ($t(11) = 8.20$, $p < .001$, $r^2 = .86$).

Note that our Equation 1 is mathematically equivalent to Equation 6 in Hulleman (2005). As Hulleman described, this method of estimating the number of items tracked from percent correct assumes that participants have no knowledge about the distractor identities. Recent work suggests that this assumption is valid. When the load of tracking targets is high, observers have little to no information about the location of individual distractors during multiple object tracking (Alvarez & Oliva, 2007). Moreover, although assuming knowledge of distractors would change our overall capacity estimates, it would not change the relative difference in performance we see in Figure 3b. For example, capacity estimates computed using Hulleman's maximum number of objects tracked (Equation 8), and minimum number of objects tracked (Equation 9), changed the absolute value of capacity estimates but showed the same relative pattern of performance as that shown in Figure 3b. Specifically, the number of items tracked decreased as the spacing between items decreased, and this effect was greater for the faster moving items.

Discussion

This experiment shows two important results. First, it is not possible to track as many targets when the spacing between items is small (requiring more precise selection) as when the spacing between items is large (allowing coarser selection). This finding is consistent with previous research on the tradeoff between the number of items selected at once and the spatial resolution of attention (Franconeri et al., in press). Second, the cost for decreasing spacing is greater for fast moving targets than for slow moving targets. This novel finding suggests that it is possible to track slow moving targets with a "tighter" focus of attention, enabling distractors to be ignored or suppressed even when they are close to the targets. In contrast, when targets move quickly, a "coarser" focus of attention appears to be necessary, causing nearby items to impair tracking accuracy to a greater extent.

The current results can also be interpreted in terms of positional uncertainty. On this view, tracking mechanisms estimate the position of targets with some uncertainty. As the number of targets increases, or the speed at which targets move increases, the positional uncertainty increases. On this positional uncertainty account, the coarseness in the spatial resolution of selection arises by the accumulation of local errors over time.

General discussion

The current study represents a challenge to the hypothesis that the number of objects that can be tracked is a fixed number, set by an architectural constraint. Experiment 1 showed a systematic decrease in the maximum speed of tracking as the number of targets tracked increased. This finding suggests that the limit on tracking is not determined by a fixed number of tracking mechanisms, but instead that it is primarily set by a shared resource. In Experiment 2, fewer items could be tracked when precise selection windows were required than when coarse selections were possible, and the effect of required precision was greater for faster moving objects. These results suggest that the number of tracked objects and the speed of the tracked objects affect the spatial resolution of attention: increasing the number of objects tracked or the speed of tracked objects increases the size of the selection window. Combined, these results suggest that the number of objects that can be tracked depends on a flexibly allocated resource, and that allocating more resources to tracking a particular object increases the precision with which that object is selected.

These findings are consistent with the more general claim that attentional processing is not limited to a fixed number of items (Davis, 2004; Davis, Welch, Holmes, & Shepherd, 2001; Tripathy & Barret, 2004; Tripathy, Narasimhan, & Barret, 2007). For example, the ability to discriminate changes in the trajectory of moving items drops off dramatically as the number of tracked trajectories increases beyond 1 (Tripathy & Barret, 2004). This suggests that the resolution required to detect a deviation is the primary limit on the number of trajectories that can be tracked, not the number of trajectories (Tripathy et al., 2007). While our conclusions are similar, there are several reasons to believe that the constraints on trajectory tracking are different than those on multiple object tracking. First, the trajectory tracking task places heavy demands on visual memory. To determine whether an item has changed direction, it is necessary to compare its current direction to its previous direction. Indeed, visual memory limitations may be the primary determinant of the limits on trajectory tracking (Narasimhan, Tripathy, & Barret, 2005). In contrast, the multiple-object tracking task does not require a direct comparison of the current features of an object to its previous features, and iconic memory is unlikely to play an essential role in this task. Second, observers with amblyopia are impaired in multiple object tracking (Ho et al., 2006), but not in the trajectory task (Levi & Tripathy, 2006). Thus, while both multiple object tracking and trajectory tracking appear to be resource-limited, the resource in multiple object tracking appears to be attention (consistent with the attentional resolution results of Experiment 2), whereas in trajectory tracking it appears to be visual sensory memory.

Another line of related research comes from the object-based attention literature. In the standard object-based attention paradigm, participants are required to make a speeded judgment about two features which either appear on the same object, or which appear an equal distance apart but on separate objects. The typical finding is that there is a cost for dividing attention across objects (Duncan, 1984). However, if the amount of "perceptual information" is equated in the 1-object and 2-object conditions, this cost is eliminated (Davis et al., 2001). Davis et al. (2001) concluded that attention is not limited to selecting a fixed number of objects but instead is limited by binding operations. Specifically, attention is limited in the number of within-object and between-object "links" it can maintain.

A within-object link represents the relationship between features of a single object (e.g., shape, texture, color), whereas a between-object link represents the relationship between features of separate objects (Davis, 2004). According to Davis (2004), the number and strength of these links imposes the limit on the number of objects that can be attended. This theory of attention would need to be expanded to account for the current results. For example, in Experiment 2, we found that target-distractor spacing and speed impact the number of objects that can be tracked, but the number and appearance of display items was constant across conditions. Thus, the number of within-object links and between-object links was constant, and therefore the number of links cannot explain the current results. However, the link model could potentially account for the current results if it were modified to specify that the between-object links become weaker as speed increases and as inter-item spacing decreases.

Our proposal differs from these previous proposals in its focus on (1) inter-item interference and (2) the decrease in spatial resolution as the number of targets increases and as target speed increases. To account for our results, we propose that the number of tracking mechanisms that can be deployed is flexible and limited by a shared resource. We introduce the term FLEX (a Flexibly allocated indEX) to refer to these flexibly allocated tracking mechanisms. It is possible to envision a variety of models that produce a drop-off in spatial precision as the number of selected items increases. A parallel account would hold that there is no limit on the number of FLEXs, but there is a cost for each additional FLEX deployed: as the number of FLEXs increases, the efficiency with which each individual FLEX can track decreases because they all draw on a common resource. For example, if the objects were tracked as the vertices of a single deforming object (Yantis, 1992), then increasing the number of vertices in this object may place greater demands on the shape memory system underlying tracking.

An alternative, serial account assumes that there is only one FLEX, and that this single FLEX is moved serially from object to object. The FLEX marks each target location with a placeholder and returns to that placeholder after sampling other targets. If there were a fixed sampling rate or if sampling at a faster rate reduced the accuracy with which placeholders could be positioned, then increasing the number of targets would decrease the precision of tracking. Pylyshyn and Storm (1988) initially proposed and ruled out a serial tracking mechanism based on a model that consisted of several conservative assumptions concerning the sampling mechanism (e.g., the rate at which attention could move from item to item). However, the modeling assumptions about the speed of attention shifts may not have been appropriate (e.g., it is unclear that attention can be described as having a set "speed" for switching between objects, see Egeth & Yantis, 1997), and the distinction between serial and parallel processing is notoriously difficult to make empirically (Townsend, 1990). Thus, we refrain from making any claims about the serial versus parallel nature of the tracking mechanism until direct empirical evidence favors one model over the other.

The results characterizing tracking as a resource-limited raise many important questions. What is this resource? How does it determine the number of items that can be tracked? Why is there a tradeoff between the number of items tracked and the spatial resolution with which each item is represented? Why are faster moving objects tracked with a coarser selection window? Are there multiple FLEXs, or is there just a single FLEX? Raising these questions is an important benefit of characterizing tracking as resource-limited. If we cannot explain the limits on attentive tracking by assuming that the number of tracking mechanisms alone explains the limit, then we must seek a more detailed understanding of the mechanisms underlying tracking. Discovering the important role of attentional resolution in Experiment 2 was an initial step in this direction.

The implications of characterizing tracking as primarily resource limited are not restricted to object tracking. Limits on tracking have influenced theories of other aspects of cognitive processing, such as the ability to rapidly enumerate small numbers of items (Trick & Pylyshyn, 1993), memory storage (Cowan, 2001), the object concept in infants (Carey & Xu, 2001), as well as number perception in infants (Feigenson, Carey, & Hauser, 2002), and non-human primates (Nieder & Miller, 2004). The current results indicate that a common capacity limit of 4 items is not enough to make or to dismiss the connection between these processes and the object tracking system in adults. If these systems are all tapping the same underlying mechanism, then they should show resource limitations similar to those shown for tracking, such as sensitivity to speed or a loss of precision with the number of items tracked. Given the great deal of data making connections between these systems it is still likely *that* they are related, but understanding the nature of the resource limits can take us further to show *how* they are related. In this way, characterizing tracking as a resource limited mechanism can lead to a richer understanding of attentive tracking and its relation to other cognitive processes.

Acknowledgments

This research was supported by NIH/NEI Fellowship #F32 EY016982 to G.A.A.

Commercial relationships: none.

Corresponding author: George A. Alvarez.
Email: alvarez@mit.edu.
Address: 77 Massachusetts Avenue, 46-4078c, Cambridge, MA 02138.

References

Allen, R., McGeorge, P., Pearson, D. G., & Milne, A. B. (2004). Attention and expertise in multiple target tracking. *Applied Cognitive Psychology, 18,* 337–347.

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 attentional resources? *Journal of Experimental Psychology: Human Perception and Performance, 31,* 643–667. [PubMed]

Alvarez, G. A., & Oliva, A. (2007). The representation of ensemble visual features outside the focus of attention [Abstract]. *Journal of Vision, 7*(9):129, 129a, http://journalofvision.org/7/9/129/, doi:10.1167/7.9.129.

Carey, S., & Xu, F. (2001). Infants' knowledge of objects: Beyond object files and object tracking. *Cognition, 80,* 179–213. [PubMed]

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

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–185. [PubMed]

Davis, G. (2004). Characteristics of attention and visual short-term memory: Implications for visual interface design. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences, 362,* 2741–2759. [PubMed] [Article]

Davis, G., Welch, V. L., Holmes, A., & Shepherd, A. (2001). Can attention select only a fixed number of objects at a time? *Perception, 30,* 1227–1248. [PubMed]

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113,* 501–517. [PubMed]

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

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science, 13,* 150–156. [PubMed]

Franconeri, S., Alvarez, G. A., & Enns, J. (in press). How many locations can you select at once? *Journal of Experimental Psychology: Human Perception and Performance.* AQ1

Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition, 101,* 217–245. [PubMed]

Ho, C. S., Paul, P. S., Asirvatham, A., Cavanagh, P., Cline, R., & Giaschi, D. E. (2006). Abnormal spatial selection and tracking in children with amblyopia. *Vision Research, 46,* 3274–3283. [PubMed]

Hulleman, J. (2005). The mathematics of multiple object tracking: From proportions correct to number of objects tracked. *Vision Research, 45,* 2298–2309. [PubMed]

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention, *Cognitive Psychology, 43,* 171–216. [PubMed]

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175–219. [PubMed]

Leslie, A., Xu, F., Tremoulet, P., & Scholl, B. (1998). Indexing and the object concept: Developing "what" and "where" systems. *Trends in Cognitive Sciences, 2,* 10–18.

Levi, D. M., & Tripathy, S. P. (2006). Is the ability to identify deviations in multiple trajectories compromised in amblyopia? *Journal of Vision, 6*(12):3, http://jounralofvision.org/6/12/3/, doi:10.1167/6.12.3. [PubMed] [Article]

Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., et al. (2005). Multiple-object tracking is based on scene, not retinal, coordinates.

Journal of Experimental Psychology: Human Perception and Performance, 31, 235–247. [PubMed]

McKeever, P., & Pylyshyn, Z. W. (1993). Nontarget numerosity and identity maintenance with FINSTs: A two component account of multiple object tracking (Technical Report: Cogmem 65). London, Ontario: Centre for Cognitive Science, University of Western Ontario.

Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex: II. Spatiotemporal range of directional interactions in MT and V1. *Journal of Neurophysiology, 55,* 1328–1339. [PubMed]

AQ2 Mitroff, S. R., & Alvarez, G. A. (in press). Space and time, not surface features, guide object persistence. *Psychonomic Bulletin & Review.*

Narasimhan, S., Tripathy, S. P., & Barrett, B. T. (2005). The decay of trajectory-traces in memory when tracking multiple trajectories. *Perception, 34,* 77.

Nieder, A., & Miller, E. K. (2004). Analog numerical representations in rhesus monkeys: Evidence for parallel processing *Journal of Cognitive Neuroscience, 16,* 889–901. [PubMed]

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource limited processes. *Cognitive Psychology, 7,* 44–64.

O'Hearn, K., Landau, B., & Hoffman, J. E. (2005). Multiple object tracking in people with Williams syndrome and in normally developing children. *Psychological Science, 16,* 905–912. [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.

Orban, G. A., Kennedy, H., & Bullier, J. (1986). Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: Influence of eccentricity. *Journal of Neurophysiology, 56,* 462–480. [PubMed]

Orban, G. A., Kennedy, H., & Maes, H. (1981). Response to movement of neurons in areas 17 and 18 of the cat: Velocity sensitivity. *Journal of Neurophysiology, 45,* 1043–1058. [PubMed]

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

AQ3 Pylyshyn, Z. (1998). The role of visual indexes in spatial vision and imagery. In R. Wright (Ed.), *Visual attention*. New York: Oxford University Press.

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

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

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

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

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like tweedledum and tweedledee, but they can (and should) be distinguished. *Psychological Science, 1,* 46–54.

Trick, L. M., Audet, D., & Dales, L. (2003). Age differences in enumerating things that move: Implications for the development of multiple-object tracking. *Memory & Cognition, 31,* 1229–1237. [PubMed]

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

Tripathy, S. P., & Barrett, B. T. (2004). Severe loss of positional information when detecting deviations in multiple trajectories. *Journal of Vision, 4*(12):4, 1020–1043, http://journalofvision.org/4/12/4/, doi:10.1167/4.12.4. [PubMed] [Article]

Tripathy, S. P., Narasimhan, S., & Barrett, B. T. (2007). On the effective number of tracked trajectories in normal human vision. *Journal of Vision, 7*(6):2, 1–18, http://journalofvision.org/7/6/2/, doi:10.1167/7.6.2. [PubMed] [Article]

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

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

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature, 396,* 72–75. [PubMed]

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1: Please update the publication status of Franconeri et al., in press.
AQ2: Please update the publication status of Mitroff & Alvarez, in press.
AQ3: Please provide page range.

END OF AUTHOR QUERIES