

Direction information in multiple object tracking is limited by a graded resource

TODD S. HOROWITZ

Brigham and Women's Hospital, Boston, Massachusetts
and Harvard Medical School, Cambridge, Massachusetts

AND

MICHAEL A. COHEN

Harvard University, Cambridge, Massachusetts

Is multiple object tracking (MOT) limited by a fixed set of structures (slots), a limited but divisible resource, or both? Here, we answer this question by measuring the precision of the direction representation for tracked targets. The signature of a limited resource is a decrease in precision as the square root of the tracking load. The signature of fixed slots is a fixed precision. Hybrid models predict a rapid decrease to asymptotic precision. In two experiments, observers tracked moving disks and reported target motion direction by adjusting a probe arrow. We derived the precision of representation of correctly tracked targets using a mixture distribution analysis. Precision declined with target load according to the square-root law up to six targets. This finding is inconsistent with both pure and hybrid slot models. Instead, directional information in MOT appears to be limited by a continuously divisible resource.

Since we live in a constantly changing environment, we need to update our representations of the world over time. One somewhat artificial example would be air traffic control, in which the controller needs to update his or her knowledge of aircraft positions and flight plans in order to ensure that takeoffs and landings go smoothly. But some version of this basic cognitive task confronts us in our everyday life, whether we are driving on a busy street, walking through a crowd, or chaperoning a child's birthday party.

The primary laboratory tool for studying updating, particularly spatial updating, is the multiple object tracking (MOT) task (Pylyshyn & Storm, 1988). In the MOT task, observers are presented with an array of identical objects. A subset of these objects is designated as targets. The observer must then track these targets while all of the objects move independently for several seconds (or minutes, as in Wolfe, Place, & Horowitz, 2007) and at the end report which items were targets (for recent reviews, see Cavanagh & Alvarez, 2005; Scholl, 2009).

The typical finding in such experiments is that observers can track three to five objects. What is the nature of this capacity limit? In the visual short-term memory (VSTM) literature, a similar limit of three to five items has been observed. Debate in the VSTM field has contrasted two accounts, which we will term the *slot* theory and the *flexible resource* theory. According to the slot theory (Awh, Barton, & Vogel, 2007; Luck & Vogel, 1997), a fixed num-

ber of objects can be stored in VSTM, regardless of their complexity. In contrast, the flexible resource theory argues that the number of objects that can be stored is inversely related to the complexity of the objects because more complex objects take up more of the fixed resource (Alvarez & Cavanagh, 2004; Eng, Chen, & Jiang, 2005). This debate has proven quite productive for the study of VSTM.

The MOT literature was originally dominated by the slot approach. The best known model is Pylyshyn's (1989) FINST ("fingers of instantiation") or visual index model, in which a fixed set of pointers can be assigned to objects for tracking, enumeration, or other visual routines; the phenomenon of MOT was predicted on this account. Here, the limit on tracking is the number of pointers.

Alvarez and Franconeri (2007) recently proposed a flexible resource model. In their model, targets are tracked by flexible indexes (FLEXes). The total number of FLEXes is limited by the availability of a finite resource. This resource determines the spatial resolution of each FLEX, such that when fewer items are tracked, resolution is higher. This account predicts that when spatial resolution is at less of a premium, observers could track more objects by allocating the resource more thinly. Kazanovich and Borisjuk (2006) proposed a somewhat similar approach, in which tracking is accomplished by a limited set of central oscillators, and the number of oscillators is limited by a finite resource. Here, the finite resource is the phase space of the oscillations, such that when there are more

T. S. Horowitz, toddh@search.bwh.harvard.edu

oscillators, it is more difficult to keep them from synchronizing with each other. Thus, tracking is better when fewer independent oscillators are necessary, because they can be farther apart in phase space.

Two major approaches to modeling MOT have not come down on either the slot or flexible resource side of the debate. Oksama and Hyönä (2008) proposed that target-relevant location information is stored in VSTM, meaning that the limits on spatial tracking would derive from the limitations on VSTM itself, which, as we noted above, are currently under debate. Thus, Oksama and Hyönä's model of MOT would be considered a slot model if VSTM turned out to be slot-like, whereas a victory for the flexible resource view of VSTM would turn the Oksama and Hyönä model into a flexible resource account of tracking. Yantis (1992) proposed that observers actually tracked only a single object, a virtual polygon whose vertices were formed by the targets. Capacity limits do not naturally fall out of Yantis's model, although one might infer that the limit is again linked to the resolution of VSTM.

The first evidence to directly address the slot versus resource question comes from Alvarez and Franconeri (2007). They measured the maximum speed at which a given number of targets could be tracked and found that the velocity limit decreased smoothly as a log function of the number of targets, supporting their flexible resource model. A converging experiment showed that when objects were moving faster, close encounters between objects were more disruptive. Further support for the flexible resource model came from a study by Iordanescu, Grabowecky, and Suzuki (2009), who asked observers to click on the position of a missing target. They found that the precision of responses increased for targets under threat from nearby distractors and decreased for targets in isolated areas of the display.

Although these data point toward a flexible resource model, we have two concerns: First, generic tracking accuracy is the wrong measure for testing a resource theory. Second, a reduction in tracking performance with speed rules out only the simplest of fixed-slot architectures.

Although tracking accuracy seems like the common-sense benchmark for testing theories of tracking, in fact it is problematic. Conceptually, there are two ways in which observers could make errors in an MOT task. First, they could simply lose track of a target. In this case, the observer's response would be based on a guess. Second, they could swap a target for a distractor, such that they successfully tracked an object that simply happened to be the wrong object. In the first scenario, the observers would track an object for only part of the trial, whereas in the second scenario, the observers would track an object, although not always the correct object, for the whole trial. We argue that only the first scenario reflects a failure of capacity.¹

Manipulations such as increasing object speed or decreasing the separation between objects are likely to lead to more of the second type of error. This would lead to reduced tracking performance when measured in the conventional way but would not necessarily reflect reduced capacity. For example, consider an observer who tracks

four targets with perfect accuracy at a low speed. At a higher speed, she might still be capable of tracking four objects. However, because of the increased frequency of close encounters, each target is swapped with a distractor, so she ends up with a score of 0 correctly tracked targets. In fact, Franconeri, Lin, Pylyshyn, Fisher, and Enns (2008) argued that these close encounters were the only real limit on tracking performance. Thus, reduced tracking accuracy cannot necessarily be taken as evidence for or against a fixed slot architecture.

Instead, it may make more sense to think about how well targets are being tracked. For example, consider a task in which there are four items: two targets and two distractors. Each item stays in its own quadrant of the visual field throughout a trial. No matter how long the trial lasts or how fast the items move, an observer would always be able to indicate which items were targets. However, if they were asked to click on the position of the lower left target, for example (as in Iordanescu et al., 2009), we would expect to find effects of tracking load, speed, or duration on how precise their responses were. Indeed, Howard and Holcombe (2008) reported just such effects for tracking load on position in MOT.

Zhang and Luck (2008) applied this insight to the study of VSTM. In the standard Luck and Vogel (1997) task, for example, observers are given a set of colored squares to remember and have to detect any change in the colors after a retention interval. The measure of memory in this task is correct detections. Zhang and Luck instead asked observers to indicate the color that they remembered at a given location, using a color wheel (cf. Wilken & Ma, 2004). Their dependent variable was the error between the true color and the reported color, expressed as an angle along the color wheel. They then assumed that these errors came from two distributions: a uniform distribution on trials on which the probe was not remembered and a von Mises (circular Gaussian) distribution on trials on which the probe was remembered. Fitting a simple mixture model allowed them to compute the mixture probability P_r , which indicates how often the probe was remembered, as well as the standard deviation of the von Mises distribution, which indicates the resolution with which the color was represented on trials when it was remembered. Note that these two parameters are independent: One might remember very few items but remember them with high precision or might, conversely, remember many items with low precision.

A straightforward slot model predicts that resolution should be constant as a function of the number of items to be remembered, because each remembered item has its own slot, and forgotten items are factored out by the mixture model. A pure resource model predicts that resolution should decrease indefinitely as more items are loaded into memory, because the resource would be spread more thinly.

If we conceptualize resolution in terms of the standard deviation of the representation, this function should follow a square root law, since the standard deviation of the average of n samples will be proportional to the standard deviation of the individual samples divided by \sqrt{n} (Bonnell & Miller, 1994; Palmer, 1990). Increasing the number of

targets decreases the number of samples, so if the standard deviation of the resolution for one targets is s , the resolution for two targets should be $\sqrt{2}s$, for three targets $\sqrt{3}s$, for four targets should be $2s$, and so forth.

In addition to their methodological innovation, Zhang and Luck (2008) also proposed two theoretical alternatives between a simple fixed-slot architecture and a pure resource account. In the *slots + resources* model, there are a fixed number of slots in memory, but a variable resource can be distributed unequally between these slots. Thus, if the observer only remembers one item, all of the resource is put into that slot, and the memory is a very high resolution one. If two or three items are in memory, the resource is divided among them, leading to lower resolution, following the square-root law. However, only a fixed number of items can be remembered. If the number of items to be remembered exceeds this limit, the excess items are simply not remembered. Thus, resolution declines as more targets are added, but only up to the point at which the number of items equals the observer's capacity. Additional items have no effect on resolution, because they are not remembered and thus contribute uniformly distributed errors to the distribution.

The *slots + averaging* model makes similar predictions from a different logic. In this model, when the number of to-be-remembered items is at or above capacity, each item has a single slot assigned to it. When fewer items are to be remembered, the items can be assigned to more than one slot. This increases resolution, because each slot provides an independent sample of the color. Thus, the model predicts that resolution will be constant when the memory load is above capacity and improve via the square-root law as load drops below capacity.

These two intermediate models thus make the same prediction: The resolution of the representation should decline as the number of items increases, up to the observer's capacity, at which point additional items should not matter. Note that a decline in resolution means an increase in the standard deviation parameter of the fitted von Mises distribution. This is exactly the pattern that Zhang and Luck (2008) observed in short-term color memory. A subsequent experiment allowed them to demonstrate that the *slots + averaging* model was a better way to explain the data than the *slots + resources* model.

Although this technique has not yet been applied to MOT data, Ma and Huang (2009) recently analyzed data from several experiments in which the trajectory deviation paradigm of Tripathy and colleagues (Tripathy & Barrett, 2004; Tripathy, Narasimhan, & Barrett, 2007) was used. In this paradigm, a number of objects move across the display at a constant speed. When they reach the midpoint, some of the objects may change direction subtly; the observer's task is to detect these changes. The number of trajectories that observers can simultaneously monitor is roughly three or four, similar to the MOT limit (Tripathy et al., 2007). Ma and Huang tested Zhang and Luck's (2008) hybrid models against a Bayesian observer limited only by neural noise, rather than by a structurally fixed capacity, and found that the limited-capacity models could not explain the data as well as the Bayesian ob-

server. However, the trajectory deviation paradigm differs from standard MOT in a number of ways. The observers in these experiments tried to detect an event (a change in the direction of one or more moving objects) occurring at a predictable spatiotemporal location (when the objects reach the midpoint of the display). The observers were not required to select targets and ignore distractors; if they confused one item for another, there was no penalty. Thus, although both tasks involve moving objects, we may not be able to generalize from Ma and Huang's findings. In fact, Vul, Frank, Tenenbaum, and Alvarez (2009) argued that an ideal-observer model of MOT cannot account for the reductions in performance with increasing tracking load without introducing a capacity limit.

More to the point, Howard and Holcombe (2008) demonstrated that precision for the position, spatial frequency, and orientation of tracked objects decreases as a function of the number of targets, even at tracking loads below the conventional four-item limit. These data argue against the standard slot model, and Howard and Holcombe argued as much. However, the precision data that they report appear to increase less with tracking load than a pure resource model would predict, suggesting a hybrid model with perhaps three slots. This example points up the importance of analyzing data only from trials on which the observer correctly tracked the object. Although Howard and Holcombe made some effort to remove responses that were likely to be guesses, they did not independently report the precision of the representation for correctly tracked objects.

In this article, we applied Zhang and Luck's (2008) analysis to the MOT paradigm. Instead of measuring the number of targets correctly tracked, we measured the resolution of the representation of tracked targets. In particular, we measured direction information. Since direction is circular, it is somewhat more amenable to the Zhang and Luck analysis technique than is position. Theoretically, an ideal observer would use both position and its derivative (i.e., trajectory) to track items, assuming that trajectories were not too variable, and human observers seem to follow this principle.

There are also empirical reasons to believe that observers encode trajectory information in attentional tracking. Verghese and McKee (2002) showed that extended trajectories engage focused attention, allowing better detection than a series of disconnected brief trajectories (Verghese, McKee, & Grzywacz, 2000; Verghese, Watanianuk, McKee, & Grzywacz, 1999). Holding trajectories in memory disrupts MOT performance (Shen, Makovski, & Jiang, 2006), and repeating trajectories improves performance (Makovski, Vazquez, & Jiang, 2008; Ogawa & Yagi, 2003). Trajectory information is employed in recovering targets that disappear and reappear during MOT. Finally, the use of trajectory information allows us to make at least a qualitative comparison with the results of trajectory deviation tasks (Ma & Huang, 2009; Narasimhan, Tripathy, & Barrett, 2009; Tripathy et al., 2007).

In order to measure direction information, we employed a method-of-adjustment technique (Figure 1) that allows us to measure the distribution of errors around the true trajectory. Following Zhang and Luck (2008), we then de-

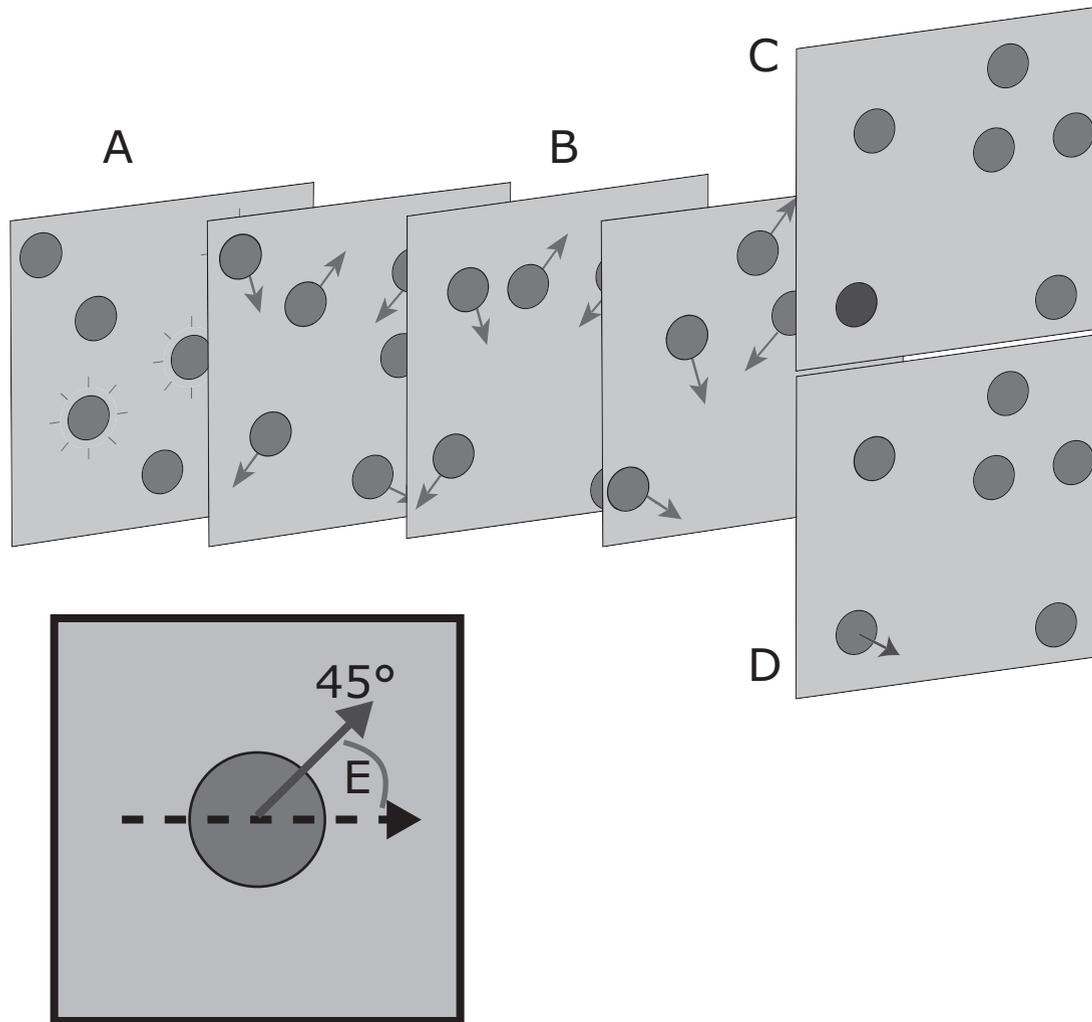


Figure 1. Method. Observers were presented with an array of identical dark gray disks on a light gray background. A variable number of targets were designated at the start of the trial (A). All disks then moved independently for several seconds (B). In Experiment 1, two probes were presented in sequence on the same probe disk at the end of each trial. When the probe disk turned blue (classification probe, C), the observer indicated whether the disk was a target. When a blue arrow was presented on the probe disk (D), the observer had to adjust the arrow to match the probe disk's trajectory. The primary dependent variable was the angular error, the difference between the adjusted angle and the probe's true direction of motion (E). In Experiment 2, observers performed the adjustment task twice on each trial but did not perform the classification task.

rived an estimate of the variance of this distribution as our measure of resolution for correctly tracked targets. An important feature of the Zhang and Luck approach is that it separates guessing responses from noisy responses, theoretically allowing us to measure the resolution of tracked targets independent of guesses on untracked targets.

In Experiment 1, observers tracked 1, 2, or 4 targets among 8 total disks, and we collected both direction information and standard target–distractor classification responses. The results showed that resolution decreased as more targets were tracked, which clearly supports a rejection of the pure slots model. Furthermore, precision was much better for targets than for distractors, and distractor precision decreased as load increased, supporting a flexible resource account. However, tracking 4 targets did not suf-

ficiently exceed capacity for the hybrid models to predict a plateau in resolution. In Experiment 2, we increased the maximum number of targets and the total number of items. Observers tracked 1, 2, 3, or 6 targets among 12 total disks. We used only direction probes on the targets in Experiment 2. We found that the pure resource model captured the data much better than either of the hybrid models.

METHOD

Participants

For each experiment, 12 observers were recruited from the volunteer observer panel of the Visual Attention Laboratory of Brigham and Women's Hospital and compensated for their time at \$10/h. All of the observers passed the Ishihara color-blindness screen, and scored 20/25 or better (with correction) on the Mentor B-VAT acuity test.

Apparatus and Stimuli

The experiments were conducted on Macintosh computers running Mac OS 10.4. The experiments were programmed in MATLAB 7.4 using the Psychophysics Toolbox version 3.08 (Brainard, 1997; Pelli, 1997). The stimuli were presented on 21-in. diagonal CRT monitors, either SuperScan Mc801 RasterOps or Mitsubishi Diamond Pro 91TXM. The monitors were set to a resolution of $1,024 \times 768$ pixels and a refresh rate of 75 Hz. The observers were placed at a viewing distance of approximately 57 cm from the monitor, such that 1 cm on the monitor subtended 1° of visual angle (dva).

The tracking stimuli were dark gray disks subtending 2.1 dva in diameter, including a 0.2 dva black border, presented on a medium gray background.

Procedure

Each trial began with the presentation of 8 (Experiment 1) or 12 (Experiment 2) disks at random positions in a 5×6 grid, with cells spaced 4.2 dva apart. The targets were designated by flashing white four times at 1 Hz. There were 1, 2, or 4 targets in Experiment 1, and 1, 2, 3, or 6 targets in Experiment 2. All of the disks then began moving at 8 dva/sec in random directions. The disks moved smoothly, changing direction only when they approached within 2.1 dva of the edge of the display. The disks that reached this buffer zone changed direction as if bouncing off an invisible boundary. The disks were allowed to occlude one another; the black border was included to help ensure correct disambiguation in this case (Viswanathan & Mingolla, 2002).

After a variable tracking duration (minimum = 5.84 sec, $M = 7.41$ sec, $SD = 6.79$ sec), the disks stopped moving, and one disk was probed. There were two types of probe: classification probes and direction probes. In Experiment 1, the targets and distractors were equally likely to be probed. Both tasks were performed on every trial. The order of the probes was counterbalanced across observers. In Experiment 2, we only employed direction probes, and only targets were probed.

The classification probe consisted of the probe disk turning blue. The observers' task was to indicate whether the probe disk was a target by pressing the up arrow key for *yes* or the down arrow key for *no*.

The direction probe was a blue arrow, extending 2.1 dva from the centroid of the probe disk. The observers' task was to rotate the arrow, using the mouse, until it matched the probe disk's trajectory. The initial angle was selected at random on each trial. In Experiment 2, after this initial response, the probe arrow was replaced at a different random angle, and the observer was asked to adjust the angle a second time. This allowed a more precise estimate of the internal representation.²

The observers were given feedback as to the accuracy of their answers for both types of probe. In the practice trials, the correct angle was illustrated with a red arrow after each trial.

Each block consisted of 10 practice trials, followed by 160 experimental trials; the number of blocks varied across experiments.

Data Analysis

The angular error data were analyzed according to the method devised by Zhang and Luck (2008). We assumed that there were two types of trials: trials on which the observers knew the probed item's trajectory and trials on which the observers did not know the probed item's trajectory. On known-trajectory trials, the responses were assumed to be drawn from a von Mises (circular normal) distribution centered on the actual trajectory, whereas on unknown-trajectory trials, the responses were assumed to be drawn from a uniform distribution. The observed distribution of trajectory errors should therefore be a mixture of these two distributions.

The von Mises distribution has two parameters: the mean μ and concentration parameter κ . For an angle α , the probability density function is given by

$$f_{VM}(\alpha | \mu, \kappa) = \frac{e^{\kappa \cos(\alpha - \mu)}}{2\pi I_0(\kappa)}, \quad (1)$$

where I_0 denotes a Bessel function of order 0.

The uniform distribution is bounded at $-\pi$ and π . The mixture distribution of errors ε is thus given by

$$\varepsilon = P_t f_{VM}(\mu, \kappa) + (1 - P_t) f_u(-\pi, \pi). \quad (2)$$

We assumed that the correct responses would be centered on the true direction (i.e., $\mu = 0$).³ There were thus two free parameters: P_t , which controls the proportion of known trajectory trials, and κ , the concentration parameter of the von Mises distribution. These parameters were fit individually for each observer and each condition using maximum likelihood estimation. Data fitting was implemented in MATLAB using the Statistics Toolbox.

The concentration parameter κ can be converted to a standard deviation σ according to Equation 3:

$$\sigma = \sqrt{2 \left(1 - \frac{I_1(\kappa)}{I_0(\kappa)} \right)}. \quad (3)$$

As an example of the mixture model procedure, Figure 2 illustrates data from 1 observer in the two-target condition of Experiment 2. The bars plot a histogram of the angular errors. The broken line plots the estimated distribution of errors from trials on which the target was tracked (i.e., the von Mises distribution, f_{VM} , with $\sigma = 9.7$), whereas the dotted line plots the estimated distribution of errors on untracked trials (the uniform distribution, f_u). The solid line represents the mixture distribution, $.80 f_{VM} + .20 f_u$.

The obtained P_t and σ values were then treated as descriptive statistics and used to evaluate the four models described below. The models attempt to predict σ , the standard deviation of the directional distribution.

Pure slots model. The pure slot model predicts a fixed resolution, regardless of the number of targets, so any effect of tracking load on σ will disconfirm this model.

The remaining three models use the fitted σ value from one load condition to predict the σ in the remaining load conditions. Therefore, these models do not have any free parameters. Capacity was estimated for each observer in Experiment 1 by multiplying the fit-

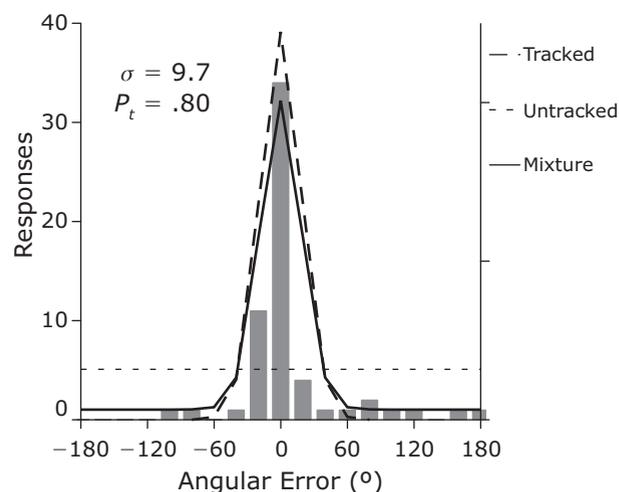


Figure 2. Illustration of the mixture model procedure. Gray bars represent a histogram of the angular errors for 1 observer tracking two targets in Experiment 1, in 20° bins. The broken line (tracked) represents the estimated (von Mises) distribution of errors on trials on which the probed target was tracked, whereas the dotted line (untracked) represents the estimated (uniform) distribution of errors on trials on which the probed target was not tracked. The two distributions were mixed by multiplying the tracked distribution by P_t and the untracked distribution by $(1 - P_t)$. See the Data Analysis section for further details.

ted P_t value from the four-target condition by four. In Experiment 2, we multiplied the fitted P_t value from the three-target condition by three.⁴

Pure resource model. The pure resource model assumes that resolution is maximal when tracking one target and that adding targets reduces resolution. We took σ for the one-target condition and multiplied it by the square root of the number of targets to obtain predictions for the remaining conditions.

Slots + resources model. This model assumes that σ increases as in the pure resource model, up to the observer's capacity, at which point it remains constant.

Slots + averaging model. This model also assumes a fixed number of slots, but when there are fewer targets than slots, more than one slot can be assigned to a given target, leading to a more precise representation of target direction via more samples. Thus, σ at maximum load is taken to indicate the resolution of a single slot. When load drops below capacity, σ is divided by the square root of the number of targets.

Note that the two hybrid models are essentially equivalent, except that the slots + resources model uses σ for the one-target case to predict σ at greater loads, whereas the slots + averaging model works down from σ at the maximum load.

Tracking performance. The proportion of targets tracked can be estimated independently from the classification data and the direction error distribution analysis. Using the classification data, we derived a corrected accuracy value a as $H - F$, where H is the hit

rate and F the false alarm rate. This is intended to correct for guessing (see Cowan, 2001). The distribution fits provide P_t as a direct measure of the proportion of tracked targets.

RESULTS

Experiment 1

Results from Experiment 1 are shown in Figure 3. The left panel of Figure 3 shows the estimated σ values as a function of number of targets. The value of σ for the targets (plotted as circles) increased significantly as the observers were asked to track more targets [$F(2,22) = 43.2, p < .0001$]. Since σ is the inverse of resolution, this demonstrates that resolution declines with tracking load, thus disconfirming the simple slot model. Unsurprisingly, the observers had poor information about direction for the distractors (diamonds). The increase in σ for distractors was not significant [$F(2,22) = 2.2, p = .14$] but may reflect a redistribution of flexible resources away from the distractors as the task became harder.

We cannot discriminate between the remaining models in these data. The pure resolution model predicts that σ should increase with tracking load more or less indefi-

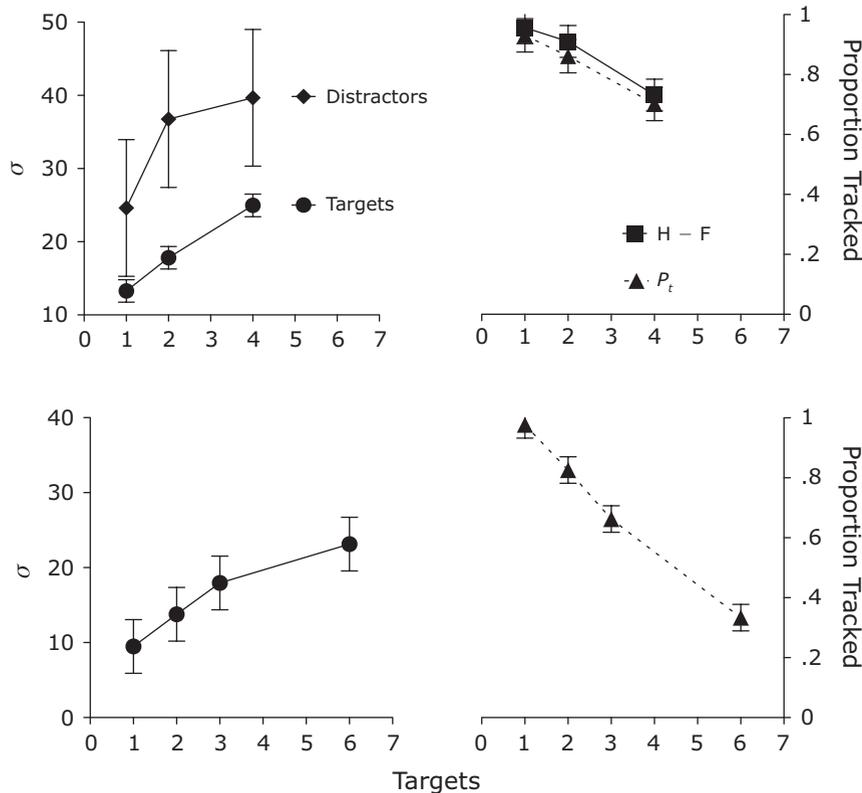


Figure 3. Proportion of targets tracked and distribution parameters as a function of tracking load for Experiment 1 (top panels) and Experiment 2 (bottom panels). The left panels plot σ , the standard deviation of the fitted von Mises distribution, for both targets (circles) and distractors (diamonds). The right panels plot measures of proportion of targets tracked: P_t , the fitted mixture parameter (triangles); and corrected classification accuracy $H - F$ (squares, Experiment 1 only). See the Data Analysis section for details. All error bars represent the 95% within-subjects confidence intervals (Loftus & Masson, 1994; Masson & Loftus, 2003) based on the effect of number of targets.

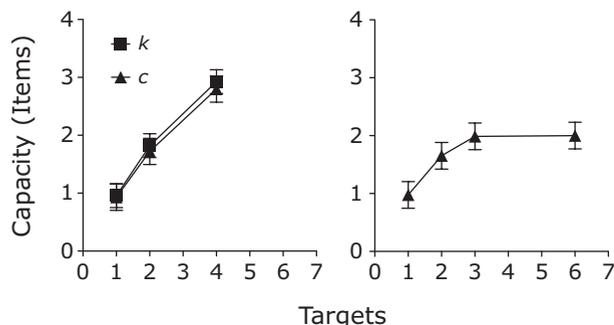


Figure 4. Tracking capacity as a function of tracking load for Experiment 1 (left panel) and Experiment 2 (right panel). The triangles denote capacity derived from the distribution analysis (computed as NP_t), and the squares denote capacity derived from the classification probes [computed as $N(H - F)$]. See the Data Analysis section for details. All error bars represent the 95% within-subjects confidence intervals based on the effect of number of targets.

nately. The hybrid models predict that σ should increase monotonically with tracking load until capacity is reached. The signature of reaching capacity is that the apparent number of items tracked should plateau. The left panel in Figure 4 shows the number of items tracked, according to both classification (k) and distribution (NP_t) analyses. The two measures are positively correlated, but not perfectly (Spearman's $r = .71, .65, \text{ and } .49$ for the one-, two-, and four-target conditions, respectively). The two measures show the same trend, however, which is that there is no apparent plateau in the number of tracked items. Thus, target loads were at or below the observers' capacity in this experiment, and the three remaining models would make similar predictions under these conditions.

Experiment 2

Experiment 2 was designed with two aims. First, we wanted to be able to observe a plateau in tracking capacity, which would allow us to discriminate between pure resource and hybrid models. Second, we wished to eliminate the

dual-task aspect of Experiment 1, in which observers had to make two different kinds of responses (classification and direction) and were sometimes asked about the direction of distractors that they were not supposed to track. Therefore, Experiment 2 was closely modeled on Zhang and Luck's (2008) Experiment 2. Observers tracked 1, 2, 3, or 6 targets among 12 total disks. We probed only targets, and used only direction probes. The left panel of Figure 4 shows that σ again increased significantly with the number of targets [$F(3,33) = 7.6, p = .0005$]. However, as the right panel shows, the increased number of distractors helped to reduce tracking performance (NP_t) such that we now observe a plateau when tracking more than two targets. This allows us to test the three variable resolution models. Since the models were fit separately to each observer's data without any parameters, we compared the models to the data using within-subjects ANOVAs. Figure 5 shows these comparisons separately for each model, so that we can illustrate the within-subjects confidence intervals (for the data \times models interaction) in addition to the RMS error. Note that the pure resource and slots + resource models use the 1-target data to predict the remaining data, so the 1-target data are not included in the ANOVA or RMS error analyses but are shown on the figure for comparisons. The slots + averaging model uses the 6-target condition for this purpose.

DISCUSSION

Directional precision for the tracked targets decreased markedly as the load increased, which disconfirms the pure slot model. Furthermore, we see no evidence of a plateau in resolution, as would be expected from a hybrid model that included a fixed set of slots. Instead, we found that σ , the inverse of resolution, increased according to the square-root law, in accordance with the predictions of the pure resource model.

Combined with recent evidence that attentional resources are dynamically redistributed during MOT (Iordanescu et al., 2009), these properties suggest that performance limitations in MOT are not simply the product

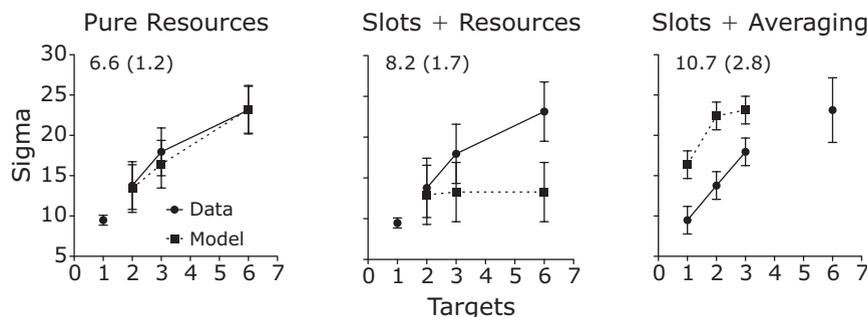


Figure 5. Experiment 2 data as a function of models. The left panel is the pure resources model, the middle panel the slots + resources model, and the right panel the slots + averaging model. Data are plotted as circles and solid lines, model predictions as squares and dotted lines. Error bars represent the within-subjects 95% confidence interval for the data \times model interaction. Unconnected points in each plot are the σ data that the model is based on. Error bars for the base values represent one standard error of the mean. The numbers on each plot indicate the RMS error for that model, with the standard error of the mean in parentheses.

of a limited-capacity system, but a true limited-resource system (Logan, 2002). Instead of a fixed set of structures (e.g., visual indexes), the limit in MOT appears to be a fixed pool of resources that can be divided thinly among several targets or concentrated on a single target (Alvarez & Franconeri, 2007; Kazanovich & Borisyuk, 2006).

The Relationship Between Direction Information and Tracking

One potential objection to our work is that by measuring direction information, we in some sense did not really study tracking per se, which is traditionally associated with position rather than direction (Keane & Pylyshyn, 2006). We have two responses to this objection. The first is that although direction information is our primary dependent variable, the task demands did require the observers to track the targets, as in more traditional MOT studies. In Experiment 1, of course, we directly measured standard tracking performance using the classification responses. In Experiment 2, however, since we only probed targets, the observers had a strong incentive to track the targets, especially given how poor the direction reports were for distractors in Experiment 1.

More conceptually, we argue that the difference between MOT and a spatial memory task is precisely the fact that the objects move, thus requiring the observer to measure not just position but its derivative(s) over time. When we ask the observers whether a particular item is a target, we are asking whether the item at this position is connected by a trajectory to the original position. Any ideal observer performing an MOT task would use both position and direction (as well as speed) to solve the correspondence problem. Indeed, when motion information is reliable, it is used (Fencsik, Urrea, Place, Wolfe, & Horowitz, 2006). Therefore, in tasks where objects move smoothly and predictably, as in our experiments, direction information is an integral part of the tracking task.

In this context, we should note that Ma and Huang (2009) came to a similar conclusion about the Tripathy and colleagues (Tripathy & Barrett, 2004; Tripathy et al., 2007) trajectory deviation detection task. As we noted above, there are important differences between MOT and trajectory deviation detection. However, there is also sufficient overlap to suggest that this pattern of results may be a general property of tasks that are limited by trajectory information. Following Vul et al. (2009), we might speculate that as trajectory information becomes less reliable, a different pattern of results would emerge.

It is important to keep in mind that we do not know whether observers have immediate access to target directions. When Howard and Holcombe (2008) asked observers to report the orientation of tracked gabors, the reported orientations tended to lag up to 40 msec when four targets were tracked (up from essentially a 0-msec lag when one target was tracked). It would be difficult to repeat their analysis with our data, because we deliberately devised the motion algorithm such that the objects changed direction infrequently. This means that an observer could correctly report the direction of a target (that is, report a direction that would be considered as coming from the

tracked distribution) from memory without knowing the target's current direction. We do not see this as a problem for our analysis, for three reasons. First, there is a critical difference between our experiment and the Howard and Holcombe paradigm. In their experiments, targets remained within virtual cages in separate regions of the display, so that it was always completely unambiguous which item was being queried. In this design, an observer who is not tracking a particular target could potentially still report something about it from memory. In our paradigm, however, the observer first had to infer which target was being queried in order to associate it with a particular memory. Thus, correct directions could only be recalled for successfully tracked targets. Second, position information lagged behind orientation information in Howard and Holcombe's experiment, suggesting that the observers might not have immediate access to target positions either. Finally, although there are critical differences between memory and tracking as *tasks* (see below), it seems clear that the *mechanisms* of tracking and memory are closely bound. In our view, tracking is a cycle of using memory to predict target locations and updating memory with the currently perceived position and direction of the targets. Under this view, tracking is memory *and* updating, so there is no real distinction between reporting the direction of a tracked target and reporting the remembered direction of a tracked object. Any discrepancy between remembered and actual properties of the target simply informs us about the temporal properties of the updating process.

Although our experiment was not designed to extract the lag information available in Howard and Holcombe (2008), we can look at the directional errors as a function of the time since the last change in direction (i.e., bounce off the wall). If there was a tendency to respond from memory, the proportion of responses that were close to the prebounce direction should decline over time. A crude analysis of our data indicated that this proportion was roughly constant over a 200-msec time span since the last bounce, suggesting that this was not a factor in our study. However, it would be informative to conduct an experiment that would be able to answer this question more definitively.

There is no reason to think that tracking is limited to position and direction information. As we noted in the introduction, an ideal observer would use all of the available information to help predict target motion. This would include speed and direction and possibly even acceleration and rate of direction change. This information could be extracted, for example, from neurons in the MT area that code direction and speed (Lorteije, van Wezel, & van der Smagt, 2008). However, there may be limits on the amount of information that can be stored, such that when more targets are tracked, less information about each target is available.

For example, a number of MOT studies have shown that observers can successfully recover targets that have disappeared for several hundred milliseconds, even when the objects continue to move while invisible (Alvarez, Horowitz, Arsenio, Dimase, & Wolfe, 2005; Horowitz, Birnkrant, Fencsik, Tran, & Wolfe, 2006; Keane & Pylyshyn,

2006). Fencsik, Klieger, and Horowitz (2007) studied performance on this task with and without motion information. In the motion condition, observers tracked the objects as they moved for 2 sec before disappearing, whereas in the no-motion condition, the objects were stationary for 2 sec before the disappearance. Performance was better in the motion condition, presumably because the observers could better predict where the objects might end up after the blank interval. However, this advantage declined as the number of targets increased, indicating that only a limited amount of motion/direction information could be stored.

What Is the Limited Resource?

What might be the nature of the resource limiting tracking performance? We can conceive of the resource as a fixed set of samples available for representing direction (as well as other spatiotemporal properties, such as position and speed). Selective attention is the act of assigning more samples to targets than to distractors (Prinzmetal, 2005). One possibility is a pool of neurons, for example in the superior parietal lobule or intraparietal sulcus (Culham et al., 1998; Howe, Horowitz, Morocz, Wolfe, & Livingstone, 2009; Imaruoka, Saiki, & Miyauchi, 2005; Jovicich et al., 2001; Seiffert, 2003), that can be assigned to targets. As the load increases, the number of neurons assigned to each target decreases, reducing the precision of the representation. This conception is in agreement with the idea that multiple targets are tracked in parallel (Pylyshyn & Storm, 1988; Yantis, 1992). This approach is supported by recent data from our lab (Howe, Cohen, Pinto, & Horowitz, 2010) showing that tracking is easier when all objects move simultaneously rather than sequentially, as would be predicted by parallel but not serial models (Eriksen & Spencer, 1969; Huang & Pashler, 2007; Shiffrin & Gardner, 1972).

Alternatively, the limit might be temporal. d'Avossa, Shulman, Snyder, and Corbetta (2006) proposed that MOT involves switching a single attentional focus among the targets. Under this account, attention is devoted to each target for less time as the number of targets increases, again reducing the fidelity of the representation. Howard and Holcombe (2008) found evidence for a mixture of serial and parallel updating in their data on the basis of the correlations between multiple reports on the same trial. It may be that some attributes are updated in parallel, some in series. Oksama and Hyönä (2008) proposed a model in which position tracking occurs in parallel but identities are bound to objects in series.

The idea that multiple foci might be sampled periodically has received support from a recent study of divided attention by VanRullen, Carlson, and Cavanagh (2007). According to their analysis, attention samples periodically even when only a single location is attended. This supports the conception of attention as an oscillatory process, as was proposed by dynamic systems theories of attention (Large & Jones, 1999; Martin et al., 2005). Under this approach, the limit on MOT might lie in the natural rhythm of the attentional oscillation, rather than in a fixed number of structures. For example, as was noted above, Kazanovich and Borisyuk (2006) proposed that tracking

is limited by the phase space available to the central oscillators used in tracking. If two independent oscillators are close in phase space, they will be more likely to synchronize with one another, leading to loss of tracking of at least one target.

The Relationship Between Tracking and Memory

Given the close connections between VSTM and MOT (Cavanagh & Alvarez, 2005; Fougne & Marois, 2006), it is surprising that we obtained results different from those of Zhang and Luck (2008). Why should directional precision in MOT decrease indefinitely (or at least up to six items), whereas precision in color memory levels off? One possibility is that color and motion vector information are different, and there are different representations for different types of information. Alternatively, the differences might not reflect differences in the kind of information, but in the processes used to accomplish the task. In a memory task, an observer tries to keep a representation from changing (i.e., degrading) over some time interval, whereas in a tracking task, the observer actively changes the representation to keep up with (and predict) the environment. These contrary task demands might call for different representational strategies; there might be different representations for different purposes.

The neuroanatomical counterpart of Zhang and Luck's (2008) slots + resources hybrid can be found in an article by Xu and Chun (2006), who showed that in a series of VSTM tasks, a slot-like limit of roughly four items could be observed in the activity of the posterior intraparietal sulcus (PIPS), as well as of the lateral occipital complex, whereas continuously increasing activity was observed in the anterior intraparietal sulcus (AIPS). In an MOT task, Howe et al. (2009) found that although tracking stationary objects (i.e., attending to and remembering their positions) primarily involved only PIPS, tracking moving objects recruited enhanced activity from a larger network centered on AIPS, as well as MT+, the superior parietal lobule, and the frontal eye fields. Thus, one hypothesis is that AIPS is used for representations that require frequent updating, whereas PIPS is used for representations that need to be preserved. Under this account, color information in the Zhang and Luck task was stored in PIPS, whereas direction information in our experiments was stored in AIPS. A second hypothesis might be that MT+ represents motion information (e.g., Berman & Colby, 2002; Chawla et al., 1999; Luks & Simpson, 2004), including direction information (Shulman et al., 1999), and the lack of fixed capacity limits might arise there. In support of the idea that there are different representations for different types of information, Scholl, Pylyshyn, and Franconeri (2004) reported that change detection during MOT tasks was much better for spatiotemporal properties, such as position or speed, than for feature information, such as shape or color. However, on a representation-for-purposes account, we would point out that spatiotemporal properties changed over time in these experiments, whereas featural properties did not. Howard and Holcombe's (2008) paradigm might be more fruitful here. In their experiments,

position, orientation, and spatial period varied over time. As was previously noted, they found that precision for all three dimensions decreased with load. Applying the mixture distribution analysis to data from similar experiments should allow us to resolve this conundrum.

CONCLUSIONS

Increasing tracking load in an MOT task leads to decreased resolution for information about target direction. This disconfirms a simple fixed-slot architecture for tracking. Furthermore, resolution decreases in a continuous fashion according to the square-root law up to at least six targets. This suggests that some limited resource, which effectively corresponds to the number of samples, is divided up flexibly among targets. There was no evidence for a role for a fixed number of slots in tracking.

AUTHOR NOTE

This research was funded by NIH Grant MH 75576 to T.S.H. We thank Piers Howe, Steve Luck, Lauri Oksama, and Jeremy Wolfe for useful suggestions, Weiwei Zhang for assistance with the mixture model analysis, and Christina Howard and Yakov Kazanovich for constructive critiques. Correspondence concerning this article should be addressed to T. S. Horowitz, Visual Attention Laboratory, Brigham and Women's Hospital, 64 Sidney Street, Suite 170, Cambridge, MA 02139 (e-mail: toddh@search.bwh.harvard.edu).

REFERENCES

- ALVAREZ, G. A., & CAVANAGH, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106-111.
- ALVAREZ, G. A., & FRANCONERI, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13, Art. 14), 1-10.
- 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 & Performance*, *31*, 643-667.
- AWH, E., BARTON, B., & VOGEL, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*, 622-628.
- BERMAN, R. A., & COLBY, C. L. (2002). Auditory and visual attention modulate motion processing in area MT+. *Cognitive Brain Research*, *14*, 64-74.
- BONNEL, A. M., & MILLER, J. (1994). Attentional effects on concurrent psychophysical discriminations: Investigations of a sample-size model. *Perception & Psychophysics*, *55*, 162-179.
- BRAINARD, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433-436.
- CAVANAGH, P., & ALVAREZ, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*, 349-354.
- CHAWLA, D., BUECHEL, C., EDWARDS, R., HOWSEMAN, A., JOSEPHS, O., ASHBURNER, J., & FRISTON, K. J. (1999). Speed-dependent responses in V5: A replication study. *NeuroImage*, *9*, 508-515.
- COWAN, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, *24*, 87-114.
- CULHAM, J. C., BRANDT, S. A., CAVANAGH, P., KANWISHER, N. G., DALE, A. M., & TOOTELL, R. B. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*, *80*, 2657-2670.
- D'AVOSSA, G., SHULMAN, G. L., SNYDER, A. Z., & CORBETTA, M. (2006). Attentional selection of moving objects by a serial process. *Vision Research*, *46*, 3403-3412.
- ENG, H. Y., CHEN, D., & JIANG, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, *12*, 1127-1133.
- ERIKSEN, C. W., & SPENCER, T. (1969). Rate of information processing in visual perception: Some results and methodological considerations. *Journal of Experimental Psychology*, *79*, 1-16.
- FENCSEK, D. E., KLIEGER, S. B., & HOROWITZ, T. S. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Perception & Psychophysics*, *69*, 567-577.
- FENCSEK, D. E., URREA, J., PLACE, S. S., WOLFE, J. M., & HOROWITZ, T. S. (2006). Velocity cues improve visual search and multiple object tracking. *Visual Cognition*, *14*, 18-21.
- FOUGNIE, D., & MAROIS, R. (2006). Distinct capacity limits for attention and working memory. *Psychological Science*, *17*, 526-534.
- FRANCONERI, S. L., LIN, J. Y., PYLYSHYN, Z. W., FISHER, B., & ENNS, J. T. (2008). Evidence against a speed limit in multiple-object tracking. *Psychonomic Bulletin & Review*, *15*, 802-808.
- HOROWITZ, T. S., BIRNKRANT, R. S., FENCSEK, D. E., TRAN, L., & WOLFE, J. M. (2006). How do we track invisible objects? *Psychonomic Bulletin & Review*, *13*, 516-523.
- HOROWITZ, T. S., KLIEGER, S. B., FENCSEK, D. E., YANG, K. K., ALVAREZ, G. A., & WOLFE, J. M. (2007). Tracking unique objects. *Perception & Psychophysics*, *69*, 172-184.
- HOWARD, C. J., & HOLCOMBE, A. O. (2008). Tracking the changing features of multiple objects: Progressively poorer perceptual precision and progressively greater perceptual lag. *Vision Research*, *48*, 1164-1180.
- HOWE, P. D. L., COHEN, M. A., PINTO, Y., & HOROWITZ, T. S. (2010). Distinguishing between parallel and serial accounts of multiple object tracking. *Journal of Vision*, *10*(8, Art. 11).
- HOWE, P. D., HOROWITZ, T. S., MOROCZ, I. A., WOLFE, J. M., & LIVINGSTONE, M. S. (2009). Using fMRI to distinguish components of the multiple object tracking task. *Journal of Vision*, *9*(4, Art. 10), 1-11.
- HUANG, L., & PASHLER, H. (2007). A Boolean map theory of visual attention. *Psychological Review*, *114*, 599-631.
- IMARUOKA, T., SAIKI, J., & MIYAUCHI, S. (2005). Maintaining coherence of dynamic objects requires coordination of neural systems extended from anterior frontal to posterior parietal brain cortices. *NeuroImage*, *26*, 277-284.
- IORDANESCU, L., GRABOWECKY, M., & SUZUKI, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision*, *9*(4, Art. 1), 1-12.
- JOVICICH, J., PETERS, R. J., KOCH, C., BRAUN, J., CHANG, L., & ERNST, T. (2001). Brain areas specific for attentional load in a motion-tracking task. *Journal of Cognitive Neuroscience*, *13*, 1048-1058.
- KAZANOVICH, Y., & BORISYUK, R. (2006). An oscillatory neural model of multiple object tracking. *Neural Computation*, *18*, 1413-1440.
- KEANE, B. P., & PYLYSHYN, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, nonpredictive function. *Cognitive Psychology*, *52*, 346-368.
- LARGE, E. W., & JONES, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*, 119-159.
- LOFTUS, G. R., & MASSON, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychological Bulletin & Review*, *1*, 476-490.
- LOGAN, G. D. (2002). Parallel and serial processing. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology: Vol. 4: Methodology in experimental psychology* (3rd ed., pp. 271-300). Hoboken, NJ: Wiley.
- LORTEJUE, J. A., VAN WEZEL, R. J., & VAN DER SMAGT, M. J. (2008). Disentangling neural structures for processing of high- and low-speed visual motion. *European Journal of Neuroscience*, *27*, 2341-2353.
- LUCK, S. J., & VOGEL, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279-281.
- LUKS, T. L., & SIMPSON, G. V. (2004). Preparatory deployment of attention to motion activates higher-order motion-processing brain regions. *NeuroImage*, *22*, 1515-1522.
- MA, W. J., & HUANG, W. (2009). No capacity limit in attentional tracking: Evidence for probabilistic inference under a resource constraint. *Journal of Vision*, *9*(11, Art. 11), 1-30.
- MAKOVSKI, T., & JIANG, Y. V. (2009). Feature binding in attentive tracking of distinct objects. *Visual Cognition*, *17*, 180-194.
- MAKOVSKI, T., VAZQUEZ, G. A., & JIANG, Y. V. (2008). Visual learning in multiple-object tracking. *PLoS ONE*, *3*, e2228.

- MARTIN, T., EGLY, R., HOUCK, J. M., BISH, J. P., BARRERA, B. D., LEE, D. C., & TESCHE, C. D. (2005). Chronometric evidence for entrained attention. *Perception & Psychophysics*, **67**, 168-184.
- MASSON, M. E. J., & LOFTUS, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, **57**, 203-220.
- NARASIMHAN, S., TRIPATHY, S. P., & BARRETT, B. T. (2009). Loss of positional information when tracking multiple moving dots: The role of visual memory. *Vision Research*, **49**, 10-27.
- OGAWA, H., & YAGI, A. (2003). Priming effects in multiple object tracking: An implicit encoding based on global spatiotemporal information [Abstract]. *Journal of Vision*, **3**(9), 339a.
- OKSAMA, L., & HYÖNÄ, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology*, **56**, 237-283.
- PALMER, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 332-350.
- PELLI, D. G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442.
- PRINZMETAL, W. (2005). Location perception: The X-Files parable. *Perception & Psychophysics*, **67**, 48-71.
- PYLYSHYN, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, **32**, 65-97.
- PYLYSHYN, Z. W., & STORM, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, **3**, 179-197.
- SCHOLL, B. J. (2009). What have we 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.
- SCHOLL, B. J., PYLYSHYN, Z. W., & FRANCONERI, S. L. (2004). *The relationship between property-encoding and object-based attention: Evidence from multiple object tracking*. Unpublished manuscript.
- SEIFFERT, A. E. (2003). Dissociating neural correlates of attentional tracking and attention to visual motion [Abstract]. *Journal of Vision*, **3**(9), 868a.
- SHEN, Y. J., MAKOVSKI, T., & JIANG, Y. (2006). Short-term visual memory for motion path [Abstract]. *Journal of Vision*, **6**(6), 35a.
- SHIFFRIN, R. M., & GARDNER, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology*, **93**, 72-82.
- SHULMAN, G. L., OLLINGER, J. M., AKBUDAK, E., CONTURO, T. E., SNYDER, A. Z., PETERSEN, S. E., & CORBETTA, M. (1999). Areas involved in encoding and applying directional expectations to moving objects. *Journal of Neuroscience*, **19**, 9480-9496.
- TRIPATHY, S. P., & BARRETT, B. T. (2004). Severe loss of positional information when detecting deviations in multiple trajectories. *Journal of Vision*, **4**, 1020-1043.
- 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, Art. 2), 1-18.
- VANRULLEN, R., CARLSON, T., & CAVANAGH, P. (2007). The blinking spotlight of attention. *Proceedings of the National Academy of Sciences*, **104**, 19204-19209.
- VERGHESE, P., & MCKEE, S. P. (2002). Predicting future motion. *Journal of Vision*, **2**, 413-423.
- VERGHESE, P., MCKEE, S. P., & GRZYWACZ, N. M. (2000). Stimulus configuration determines the detectability of motion signals in noise. *Journal of the Optical Society of America A*, **17**, 1525-1534.
- VERGHESE, P., WATAMANIUK, S. N., MCKEE, S. P., & GRZYWACZ, N. M. (1999). Local motion detectors cannot account for the detectability of an extended trajectory in noise. *Vision Research*, **39**, 19-30.
- VISWANATHAN, L., & MINGOLLA, E. (2002). Dynamics of attention in depth: Evidence from multi-element tracking. *Perception*, **31**, 1415-1437.
- VUL, E., FRANK, M. C., TENENBAUM, J. B., & ALVAREZ, G. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22* (pp. 1955-1963). La Jolla, CA: Neural Information Processing Systems Foundation.
- WILKEN, P., & MA, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, **4**, 1120-1135.
- WOLFE, J. M., PLACE, S. S., & HOROWITZ, T. S. (2007). Multiple object juggling: Changing what is tracked during extended multiple object tracking. *Psychonomic Bulletin & Review*, **14**, 344-349.
- XU, Y., & CHUN, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, **440**, 91-95.
- YANTIS, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, **24**, 295-340.
- ZHANG, W., & LUCK, S. J. (2008). Discrete, fixed-resolution representations in visual working memory. *Nature*, **453**, 233-235.

NOTES

1. The two scenarios are not mutually exclusive. If there are short-term fluctuations in capacity during a trial, an observer might lose a target and then try to recover it using position, direction, or feature information (Horowitz et al., 2007; Makovski & Jiang, 2009). Errors in this recovery process would lead to swaps.

2. This is true only as long as the directional errors are independent for the two reports. If memory for the first report influences the second report, there will be negligible gain for using two reports. The correlation between directional errors for tracking a single target was .49 and increased with the number of targets. We suggest that this increase is due to the fact that as the number of targets increases, so does the proportion of trials on which the observer is merely guessing. On guessing trials, the responses are not drawn from a true representation of direction, so they are likely to be highly correlated. Thus, we infer that the true correlation between directional errors is best estimated by the correlation on the single-target trials. If this is true, the two-response method was partially but not completely successful at improving the precision of our measurements.

3. Relaxing this constraint (i.e., adding μ as a free parameter) did not materially change the results.

4. The three-target condition was selected to match the Zhang and Luck (2008) analysis procedure. We also carried out the same analysis estimating capacity by multiplying P_i from the six-target condition by six, but this did not substantially change the results.

(Manuscript received November 16, 2009;
revision accepted for publication May 12, 2010.)