Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking

Lucica Iordanescu Department of Psychology, Northwestern University, Evanston, IL, USA AQ1

Marcia Grabowecky Department of Psychology, Northwestern University, Evanston, IL, USA, & Interdepartmental Neuroscience Program, Northwestern University, Evanston, IL, USA AQ1

Satoru Suzuki Department of Psychology, Northwestern University, Evanston, IL, USA, & Interdepartmental Neuroscience Program, Northwestern University, Evanston, IL, USA AQ2

The ability to track multiple moving objects with attention has been the focus of much research. However, the literature is relatively inconclusive regarding two key aspects of this ability, (1) whether the distribution of attention among the tracked targets is fixed during a period of tracking or is dynamically adjusted, and (2) whether motion information (direction and/or speed) is used to anticipate target locations even when velocities constantly change due to inter-object collisions. These questions were addressed by analyzing target-localization errors. Targets in crowded situations (i.e., those in danger of being lost) were localized more precisely than were uncrowded targets. Furthermore, the response vector (pointing from the target location to the reported location) was tuned to the direction of target motion, and observers with stronger direction tuning localized targets more precisely. Overall, our results provide evidence that multiple-object tracking mechanisms dynamically adjust the spatial distribution of attention in a demand-based manner (allocating more resources to targets in crowded situations) and utilize motion information (especially direction information) to anticipate target locations.

Keywords: attention, direction, localization, motion, multiple-object tracking, representational momentum, speed

Citation: Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision, 0*(0):1, 1–12, http://journalofvision.org/0/0/1/, doi:10.1167/0.0.1.

Introduction

The ability to track multiple moving objects (multiple-object tracking) is crucial in daily life. For example, while driving a car, one may need to simultaneously keep track of a running dog and child who might suddenly dash into traffic, a nearby car that might unexpectedly swerve, and oncoming traffic that might suddenly turn. Several hypotheses have been proposed to explain the mechanisms underlying this ability. For example, the visual system might use multiple "attention indices[1]" that can be assigned to a limited number of to-be-tracked objects (for reviews, see Cavanagh & Alvarez, 2005; Pylyshyn, 2001; Scholl, 2001). Although the prevalent explanations of multiple-object tracking have postulated the involvement of up to five discrete attention indices, a recent study has suggested that there may be a trade-off between the number of attention indices deployed and the spatial resolution of each attention index (Alvarez & Franconeri, 2007). For example, the visual system might deploy many low-resolution indices or a few high-resolution indices, being constrained by resource limits rather than by a strict number limit. Alternatively, the visual system might rapidly shift a single focus of attention among tracked objects to continually update their locations (e.g., Oksama & Hyönä, 2008). The visual system might also utilize global-pattern processing to track multiple objects as a coherent group, such as tracking target objects as vertices of a deforming polygon (e.g., Yantis, 1992). It is possible that multiple-object tracking is mediated by some combination of these mechanisms.

We used a novel experimental paradigm to address two fundamental questions that were not decisively answered in previous research regarding how the targets are tracked during multiple-object tracking. Answers to these questions will provide important constraints on current models of multiple-object tracking. The first question is whether multiple-object tracking mechanisms dynamically adapt to changing demands. Specifically, we investigated whether attention resources (e.g., spatial resolution of attention indices or frequency of attentional fixation) were shifted

doi: 10.1167/0.0.1 Received March 20, 2008; published 0 0, 2009 ISSN 1534-7362 © ARVO

on-line from targets in uncrowded situations (where low spatial resolution or infrequent attentional fixations would be sufficient to track targets) to targets in crowded situations (where high spatial resolution or frequent attentional fixations would be necessary to avoid losing a target).

The second question is whether multiple-object tracking mechanisms monitor the velocities (as well as the locations) of tracked objects. It is clear that motion information is utilized (and pursuit eye-movements are engaged) when only one object is tracked and the trajectory of its motion is relatively constant (e.g., when playing tennis). In fact, many studies have provided evidence suggesting that the visual system utilizes motion information to anticipate target locations when one or two targets are tracked and the target velocities are constant, nearly constant, or predictably varied (e.g., Fencsik, Klieger, & Horowitz, 2007; Müsseler, Stork, & Kerzel, 2002; Verfaillie & d'Ydewalle, 1991; for a review, see Thornton & Hubbard, 2002). The question we addressed was whether multiple-object tracking mechanisms could utilize motion (direction and/or speed) information even when multiple targets were simultaneously tracked, the trajectories of the tracked targets intermingled with those of multiple moving distractors, and when the velocities of the targets frequently changed due to inter-object collisions.

These questions are difficult to investigate using the conventional multiple-object tracking paradigm where performance is measured in terms of the number of objects that are successfully tracked. Instead, we measured the precision of tracking of individual objects. Our observers tracked three initially flashed target circles moving among seven distractor circles as in a typical multiple-object tracking task. The three targets, however, were labeled with distinct colors, red, green, and yellow (see Figure 1). Because the distractors shared these colors, the colors did not distinguish the targets from distractors, and thus targets had to be attentionally tracked. At the end of a tracking period (6 sec in Experiment 1 and variable in Experiment 2) all circles disappeared, simultaneously with the auditory presentation of a color name. Observers were instructed to precisely indicate with a mouse-click the last-known location of the tracked target of the named color. Because observers did not know in advance which of the three targets would be cued to be localized, they had to track all three targets. We were thus able to measure the localization error (the distance from the final location of the target to the location of the mouse-click) for a randomly selected target while observers tracked all three targets (see the Methods section). This procedure allowed us to evaluate the dynamic distribution of attention resources and the use of motion information during multiple-object tracking.

Regarding the distribution of attention resources, if more resources were dynamically allocated to targets in crowded situations, target-localization error should be inversely related to the degree of crowding (measured as the distance from the target to its nearest distractor) at the time of the display offset. We thus predicted that if the cued target happened to be relatively far from its nearest distractor, multiple-object tracking mechanisms would have allocated a relatively small amount of attention resources (e.g., a low-resolution attention index or non-prioritized allocation of attention), so that it should be localized with a relatively large error. In contrast, if the cued target happened to be close to a distractor, multiple-object tracking mechanisms would have allocated a relatively large amount of attention resources (e.g., a high-resolution attention index or prioritized allocation of attention), so that it should be localized with a relatively small error.

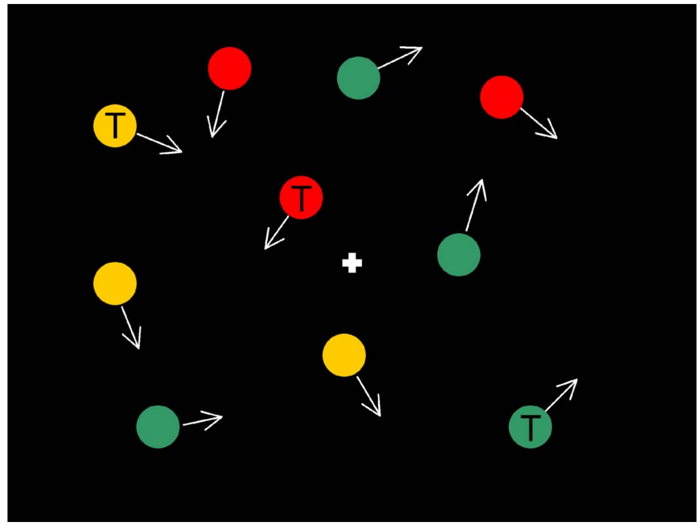

Figure 1. The rectangular tracking region contained ten colored circles that moved independently, bouncing off of one another and the border walls. The three to-be-tracked circles (labeled "T" here) were assigned different colors. The target labels and the arrows were not present in the actual display.

Regarding the use of motion information, in order to intercept a moving target, one must aim ahead of the current location of the target along its motion trajectory. We thus hypothesized that if the targets' motion directions were monitored during multiple-object tracking, observers would systematically mouse-click ahead of the target's actual location at the time of its disappearance along its motion trajectory; that is, the response vector (pointing from the location of the target disappearance to the location of the mouse-click) should be directionally tuned to the target's motion. In addition, if the targets' speeds were also monitored during multiple-object tracking, the amplitudes of forward-location clicking should be positively correlated with the target's speed. Finally, if motion direction and/or speed were not merely monitored but were actually utilized by multiple-object tracking mechanisms, stronger direction tuning (indicative of more reliable encoding of motion direction) and/or stronger correlation between the target's speed and the amplitude of forward-location clicking (indicative of more reliable encoding of speed) should be associated with more precise target localization.

Method for Experiment 1

Observers

Twenty undergraduate students at Northwestern University gave informed consent to participate for partial course credit. They all had normal or corrected-to-normal visual acuity and normal color vision and were tested individually in a dimly lit room.

Stimuli

The tracking display contained ten moving circles (diameter = 0.69°), including the three targets and seven distractors. The moving circles were confined within an 11.8° (horizontal) by 8.9° (vertical) rectangular region. The circles were colored red (CIE = [.623, .347], 19.7 cd/m^2), green (CIE = [.295, .588], 20.6 cd/m^2), and yellow (CIE = [.402, .443], 66.3 cd/m^2) and were presented against a black (0.14 cd/m^2) background. Each of the three targets had a different color, while the distractors were assigned one of the three colors with the constraint that at least two distractors had the same color as each target. All circles initially moved at the same speed (2.43°/sec), but their initial locations and motion directions were randomly determined on each trial. Vertical and horizontal directions were not used as initial directions because motions in these cardinal directions are salient. The circles bounced against one another and against the walls of the rectangular tracking region according to the principle of perfect elastic collision (i.e., conserving both momentum and kinetic energy). Both motion directions and speeds frequently changed due to collisions. At the end of the tracking period, the speeds of the circles had a unimodal distribution with a range of 0.12–6.15°/sec.[2]

The stimuli were displayed on a color CRT monitor (1024 × 768) at 75 Hz, and the experiment was controlled by a Macintosh PowerPC 8600 using Vision Shell software (micro ML, Inc.). A chin rest was used to stabilize the viewing distance at 68 cm.

Procedure

Observers initiated each trial with a button press. The three to-be-tracked targets initially flashed for 1.8 sec. Observers maintained eye fixation at a central cross (0.51° by 0.51°, 31.4 cd/m^2) while attentionally tracking the three target circles for the remaining 4.2 sec. To ensure central eye fixation, a small digit (0.17° by 0.42°) randomly selected between 0 and 9 (inclusive) was flashed for 134 ms replacing the fixation cross at a randomly chosen time on each trial (between 1.6 and 4.6 sec from the trial beginning), and the observer verbally reported the digit when it appeared. The few trials (1.4%) in which the digit was not correctly reported were excluded from the analyses. Although central eye fixation was not required in most prior studies of multiple-object tracking, we enforced it here so that we could measure distributions of attention during tracking without potential confounds from eye movements.

Following the 6-sec tracking period, the display turned blank (except for the central fixation cross), simultaneously with an auditory presentation of a color name. Observers were instructed to mouse-click the location of the target of the indicated color as precisely as possible (the initial location of the mouse cursor was at the fixation cross). Note that mouse-click responses have been shown to be sensitive for revealing encoding of motion information in a localization task, especially for smoothly moving stimuli (e.g., Kerzel, 2003a). When observers lost track of the target, they were instructed to mouse-click a location outside of the rectangular tracking region. The overall probability of losing a target was 9.8% (*SEM* = 1.7%). Note that this probability of target loss would have been much higher if observers adopted the strategy of tracking only one target (with an expected probability of loss = 67%) or two targets (with an expected probability of loss = 33%). It might be the case that observers failed to indicate that they lost the target when they actually did and instead mouse-clicked an arbitrary location. However, this possibility is unlikely as the target localization was overall quite precise for the trials on which observers mouse-clicked within the tracking region; the magnitude of target-localization errors had a unimodal distribution with the mode less than 0.5° and most errors less than 1.0° (Figure 2). These results indicate that observers successfully tracked all three targets on most trials. Each observer

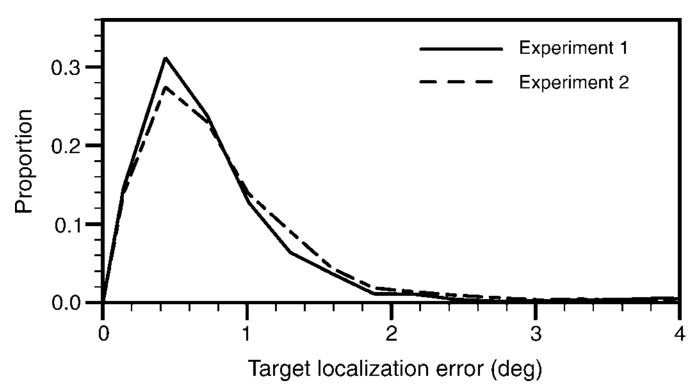

Figure 2. The distributions of target-localization errors (for Experiments 1 and 2) for the trials in which observers did not indicate that they lost the target. Note that most errors were small (less than 1.0°).

performed four blocks of 20 trials; 5 practice trials were given prior to these experimental trials.

Data analyses

Target-localization error was measured as the vector pointing from the location of the center of the target when it disappeared to the location of the mouse-click. Unusually large errors (beyond the 99th percentile in amplitude) were excluded from the analyses because those most likely represented targets that were lost without the observers' knowledge.

The degree of crowding around each target was measured as the distance from the target to its nearest distractor, D_{nearest}. The crowding region was truncated near the borders of the tracking display because distractors could not appear outside of the tracking region. Due to this border effect, targets near the edges of the display would of necessity have been relatively uncrowded, and those uncrowded targets may have produced large localization errors due to their high eccentricity. To resolve this confound, when we analyzed the effect of crowding on the precision of tracking, we recursively removed high-eccentricity targets from the analyses until the average target eccentricity was equivalent across the targets associated with different ranges of D_{nearest}.

Results from Experiment 1

Demand-based dynamic distribution of attention resources

We first discuss evidence indicating that multiple-object tracking mechanisms dynamically allocate attention resources to targets in crowded situations where the chance of losing the targets increases. We used the distance from each target to its nearest distractor (at the time of target disappearance), D_{nearest}, as the measure of crowding, with smaller D_{nearest} values indicating greater crowding. To determine how target-localization error depended on D_{nearest}, we plotted average target-localization error as a cumulative function of D_{nearest} (Figure 3A). Each point shows the average localization error for the targets associated with D_{nearest} equal to or less than the indicated value. An advantage of plotting average localization error as a *cumulative* function of D_{nearest} is that a continuous function could be obtained without dividing D_{nearest} values into artificial bins while preserving the general pattern of dependence of target localization on D_{nearest} (though with progressively greater data smoothing for larger values of D_{nearest}). The target-localization error clearly reduced for smaller values of D_{nearest}, with a significant linear trend, $F(1,19) = 5.762$, $p < 0.027$, $\eta_{\mathrm{p}} = 0.233$ (all of our

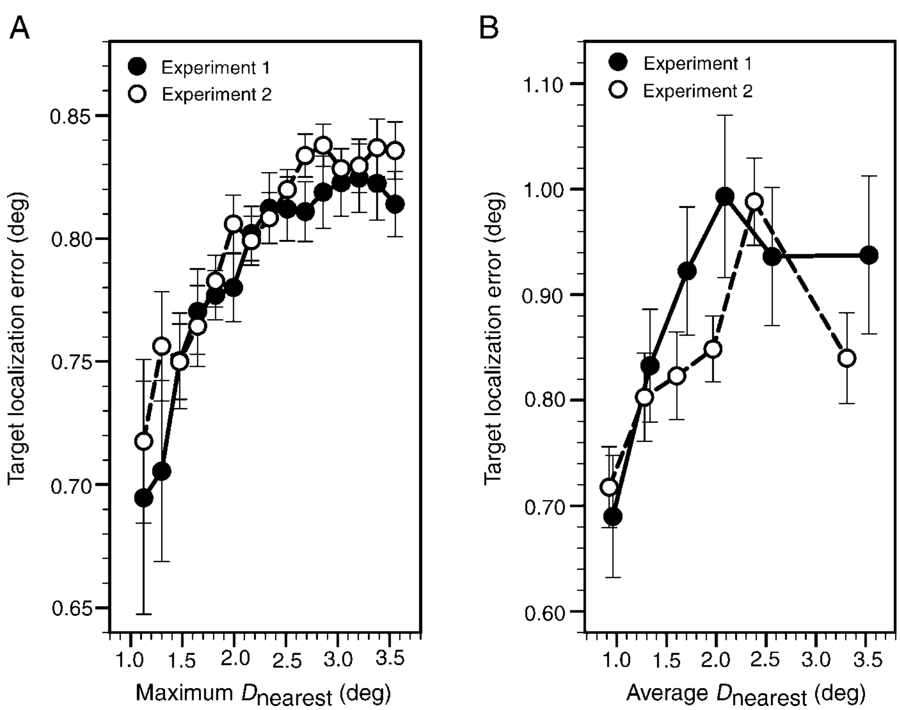

Figure 3. Target-localization error (for Experiments 1 and 2) plotted as a function of the distance between the target and its nearest distractor, D_{nearest}. (A) A cumulative plot where each point shows the average localization error for all D_{nearest} up to the indicated value. (B) A non-cumulative plot where different values of D_{nearest} are divided into six intervals, and the average error for each interval of D_{nearest} is plotted against the average D_{nearest} for that interval. The error bars indicate ±1 *SEM* (with observers as the random effect; the variance due to overall individual differences in localization errors was removed before computing *SEM*).

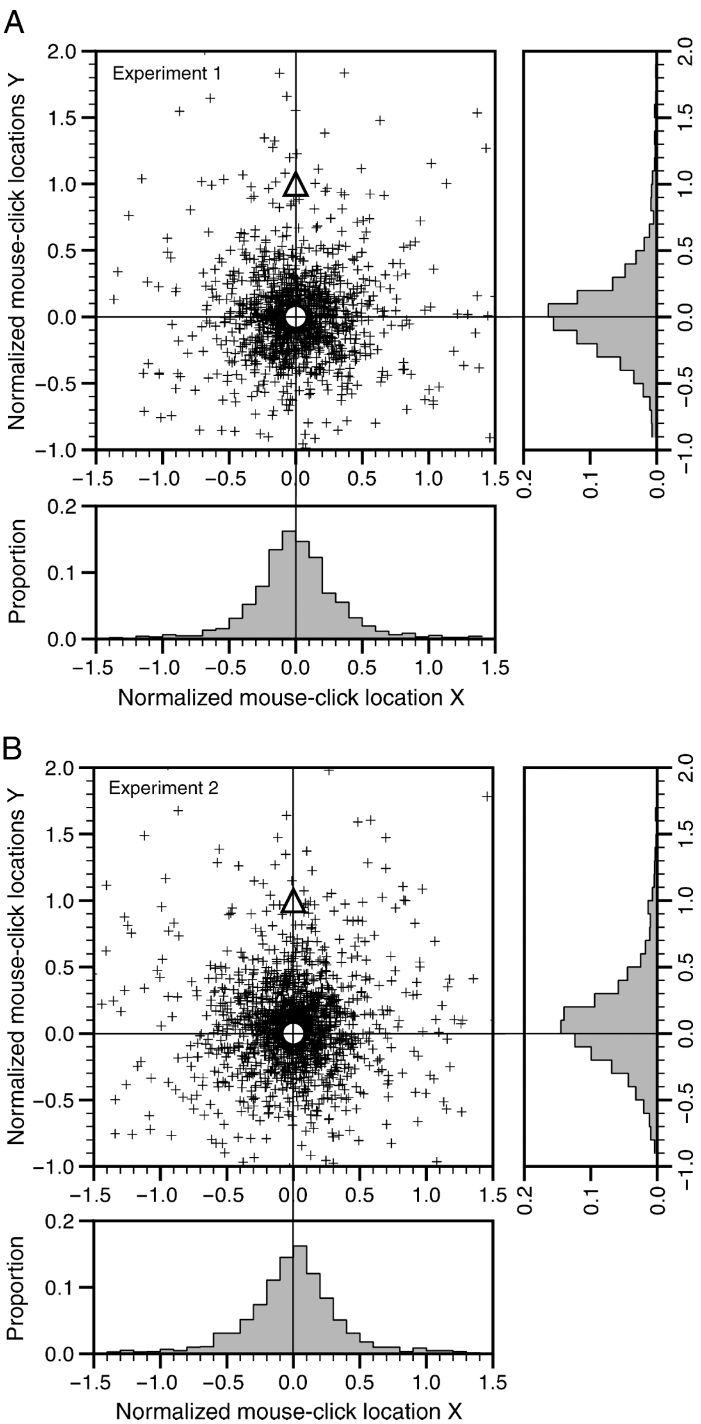

Figure 4. The spatial distribution of mouse-click locations (+ symbols) for Experiment 1 (A) and Experiment 2 (B). The mouse-click locations are normalized with respect to the axis and scale defined by the target (indicated with a circle), its nearest distractor (indicated with a triangle), and the distance between them. The accompanying histograms show the distributions of the normalized X and Y coordinates of mouse-clicks.

linear-contrast analyses are based on a conservative method using a contrast-specific error term). This indicates that the targets were more precisely localized when they were more crowded by the distractors. The same effect is also apparent in the non-cumulative plot (Figure 3B) in which the D_{nearest} values were divided into six nearly even intervals with the constraint that at least five of the six intervals included ten or more data points from each observer. The linear trend was again significant, $F(1, 19) = 7.807$, $p < 0.012$, $\eta_p = 0.291$.

These results are consistent with the idea that greater attention resources are dynamically allocated to crowded targets. Furthermore, this dynamic demand-based allocation of attention did not engage until a distractor closely approached a tracked target as localization error did not begin to improve until D_{nearest} became less than ~3° (Figure 3).

One might argue that the improved target-localization accuracy with close distractors may partially reflect a tendency for observers to group targets with their proximate distractors. Such target-distractor grouping would have been uncommon during the course of tracking because such a strategy would increase the chance of confusing targets and distractors. Nevertheless, it is still possible that at the time of localization the mouse-click response might have been attracted to the center of gravity between the target and its proximate distractor. We evaluated this possibility by plotting all mouse-click locations relative to the axis and scale defined by the target and its nearest distractor. If observers tended to mouse-click the center of gravity between the target and its nearest distractor, the mouse-clicks should be clustered around the mid point between the target and its nearest distractor. In contrast, if observers aimed at the target irrespective of the proximate distractor, the mouse-clicks should be clustered around the target. The data clearly support the latter (Figure 4A).

Taken together, these results demonstrate that the precision of multiple-object tracking increases when a distractor moves close to a tracked target (within ~3° radius), suggesting that multiple-object tracking mechanisms dynamically and adaptively distribute attention resources to targets in more crowded situations where more precise tracking is necessary.

The use of motion information

To determine whether multiple-object tracking mechanisms monitor motion direction information, we examined the direction tuning of the response vector (which points from the location of the target disappearance to the location of the mouse-click). For each trial, we computed the direction of the response vector relative to the direction of the target motion (at the time of its

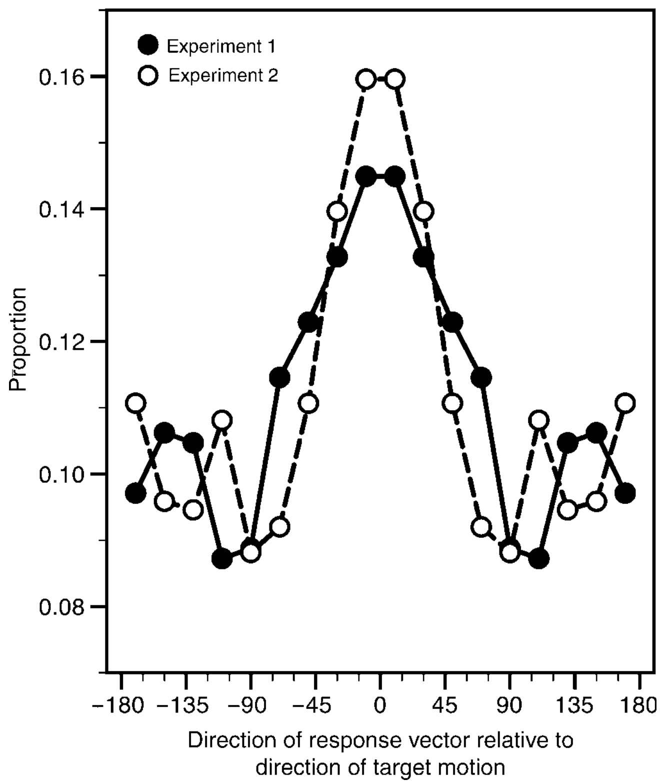

Figure 5. The direction tuning of multiple-object tracking (for Experiments 1 and 2). The distribution of the directions of response vectors (pointing from the target locations to the mouse-click locations) is plotted relative to the direction of target motion (shown as 0°).

disappearance) in terms of angular deviation. We collapsed over the clockwise and counterclockwise deviations while averaging across all directions of target motion. The response vector is clearly tuned to the direction of target motion (Figure 5; the mirror reflection of the data has been added as negative errors to aid visualization of the direction tuning). The flanking troughs around ±90° indicate that the directions orthogonal to target motions are inhibited by multiple-object tracking mechanisms.

To statistically verify this tuning, we computed a direction-tuning index for each observer, defined as,

$$\text{Direction-tuning index} = \frac{[\#\text{ of absolute angular deviations between } 90^\circ \text{ and } 180^\circ]}{[\#\text{ of absolute angular deviations between } 0^\circ \text{ and } 90^\circ]}. \tag{1}$$

A value of zero indicates consistent direction tuning (i.e., all response vectors project positively to target-motion vectors), whereas a value of 1 would indicate no direction tuning (i.e., response vectors project positively and negatively to target-motion vectors with equal probability). The mean direction-tuning index for the 20 observers was significantly below 1 (see Figure 6 for the distribution), verifying that observers tended to click ahead along the target's anticipated motion trajectory; $t(19) = 3.424$, $p < 0.003$, $d = 0.766$ (the ratio-based direction-tuning indices were log-transformed prior to t-test). This provides evidence that the directions of targets' motions were monitored during multiple-object tracking.

Successful monitoring of motion direction information, however, does not necessarily mean that it contributes to tracking performance. If multiple-object tracking mechanisms utilize motion-direction information to increase the precision of tracking, smaller direction-tuning indices (indicative of stronger direction tuning) should be associated with smaller localization errors. This was indeed the case (Figure 6), $r = 0.521$, $t(18) = 2.591$, $p < 0.02$.

We next examined whether the targets' speeds (in addition to directions) were monitored during multiple-object tracking. If speeds were monitored, anticipatory forward shifting of mouse-clicks should be greater when targets were moving faster. To evaluate this relationship, we quantified the amplitude of forward shifting of the mouse-click as the scalar product between the response vector and the unit vector along the target's motion direction; a larger positive value would indicate a greater forward shift, a zero would indicate no shift, and a larger negative value would indicate a greater backward shift. We then computed the linear correlation between this measure of forward shift and the target speed for each observer (outliers outside the 95%

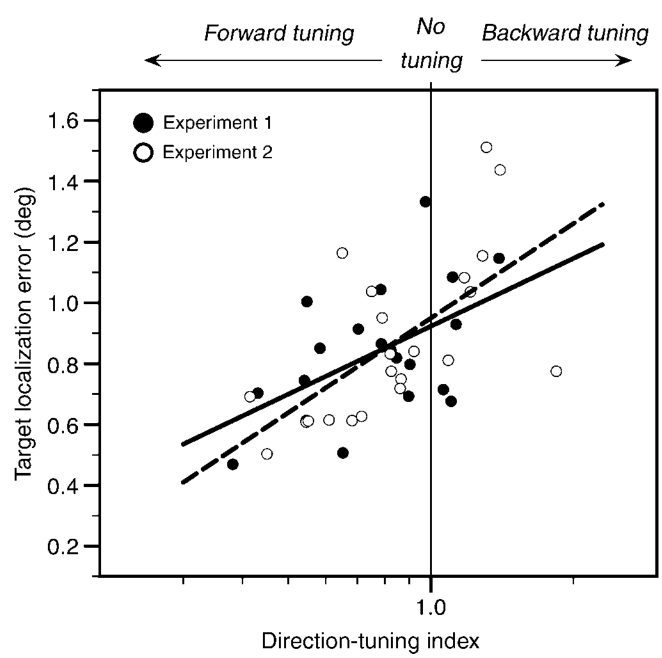

Figure 6. A scatter plot showing the correlation between the direction-tuning index (smaller values indicating stronger direction tuning) and average target-localization error, for Experiments 1 and 2. Each point represents one observer. The solid and dashed lines indicate the linear fits for Experiments 1 and 2, respectively.

confidence ellipse were eliminated before computing each correlation). A larger positive value of the correlation coefficient, r_{speed}, would indicate a more consistent linear relationship between the target's speed and the amplitude of the forward shift of the mouse-click, implying more reliable monitoring of the targets' speeds. Thus, r_{speed} provided an index of speed monitoring.

Although r_{speed} was small (M = 0.115 with SEM = 0.025), it was significantly larger than zero, $t(19) = 4.600$, $p < 0.0002$, $d = 1.029$, suggesting that the targets' speeds (in addition to directions) were monitored during multiple-object tracking. The speed monitoring index, r_{speed}, was uncorrelated with the direction-tuning index (r^2 = 0.002), suggesting that the targets' speeds and directions are monitored as separate parameters during tracking. Although the direction-tuning index was strongly correlated with the precision of target localization (Figure 6), the speed-monitoring index, r_{speed}, was not ($r^2 = 0.000$ for r_{speed}-vs.-localization-error correlation), suggesting that speed information (though monitored) does not substantially contribute to multiple-object tracking. This lack of correlation cannot be due to the possibility that the speed-related forward mouse-clicking substantially contributed as localization errors and canceled out the beneficial effect of speed monitoring. The r_{speed}-vs.-localization-error correlation was still insignificant (r^2 = 0.020) even when we removed the systematic speed-related forward shifting of mouse-clicks (explained by the regression lines) from each response vector prior to computing the correlation. We also note that the amplitude of the speed-related forward shifting was overall very small (i.e., the mean regression slope for the speed-vs.-forward-shifting correlation was 0.074° shift per degrees/sec, indicating that forward shifting increased by only 0.074° per 1°/sec increase in target speed); thus forward shifting minimally contributed to target-localization errors. A null result on the usefulness of speed information, however, must be interpreted with caution. For example, it is possible that, although r_{speed} was an adequate measure to demonstrate that speed information was monitored during multiple-object tracking, it might not have been a sensitive enough measure of the "goodness of speed monitoring" to reveal the contributions of speed information to the precision of target localization. Nevertheless, we have provided clear evidence that (1) the targets' velocities (i.e., both directions and speeds) are monitored during multiple-object tracking, and that (2) at least the motion direction information substantially contributes to the precision of multiple-object tracking.

There were several methodological concerns about the design of Experiment 1. First, the trial duration was always 6 sec, so that observers could have potentially adopted the strategy of only loosely tracking the targets most of the time and focusing attention on the targets only at the end of each trial when they had to report the location of one of the targets. If this were the case, our results would not imply that attention resources are dynamically shifted to crowded targets in the course of continuous tracking. To address this concern, in Experiment 2 we randomly varied the trial duration so that observers did not know when they had to localize a target. A second concern regarded the method we used to ensure central eye fixation. Although the unpredictably flashed probe digits were identified with high accuracy (only 1.4% errors), we cannot definitively rule out eye movements as a factor. Enforcing central eye fixation is crucial in our paradigm so that more precise target localization in crowded situations can be attributable to demand-based allocations of attention resources rather than to eye movements. A third concern was that our results could be specific to the situation where observers perform a concurrent secondary task. In Experiment 2, we thus eliminated the secondary fixation task and monitored eye movements using an eye tracker to ensure central fixation. Finally, because our paradigm is novel, it was important to replicate our original results.

Method for Experiment 2

Observers

Thirty-six undergraduate students at Northwestern University gave informed consent to participate for partial course credit. They all had normal or corrected-to-normal visual acuity and normal color vision and were tested individually in a dimly lit room. Thirteen of them did not complete the experiment because they were unable to maintain central eye fixation on most trials. This may indicate that it is difficult to maintain central eye fixation during multiple-object tracking without a concurrent fixation task, or that many observers in Experiment 1 actually made numerous eye movements. Nevertheless, as shown below, we replicated all of the primary results from Experiment 1 with strict enforcement of central eye fixation, suggesting that our primary results are applicable whether or not central eye fixation is strictly maintained. Of the 23 observers who completed the experiment, the data from one observer was removed from the analyses because of a relatively large number of eye movements (greater than 2.5 SD from the group mean). The remaining 22 observers maintained central eye fixation within 1° from the center on 93.8% of the trials (SD = 4.3%), and the few trials in which their fixation deviated more than 1° were removed from the analyses. Eye movements were monitored using an EyeLink 1000 eye tracker (SR research, with 0.15° resolution).

Stimuli

These were the same as in Experiment 1 except that no digits were flashed at the center, and the trial duration was randomly varied between 4 sec and 8 sec (matched to Experiment 1 in terms of the mean duration of 6 sec); 80 evenly spaced durations were generated within this range, and those durations were randomly assigned to the 80 trials for each observer.

Procedure and data analyses

These were the same as in Experiment 1, except that observers did not perform the secondary task of digit identification.

Results from Experiment 2

The proportion of target loss 3.0% (*SEM* = 0.6%) in this experiment was significantly reduced compared to 9.8% (*SEM* = 1.7%) in Experiment 1, $t(40) = 3.891$, $p < 0.00037$, $d = 1.20$. This improvement is not surprising because observers in this experiment did not perform the concurrent secondary task of digit identification.

As shown in Figures 2, 3, 4, 5, and 6, we replicated all of the primary results from Experiment 1, while we randomly varied trial duration, eliminated the secondary task, and enforced central eye fixation. The overall distribution of target-localization error was nearly identical in the two experiments (Figure 2). Importantly, the target-localization error clearly reduced as the distance to the nearest distractor, D_{nearest}, diminished, with significant linear trends obtained for both the cumulative analysis (Figure 3A), $F(1,21) = 13.687$, $p < 0.0013$, $\eta_p = 0.395$, and the non-cumulative (binned) analysis (Figure 3B), $F(1,21) = 13.392$, $p < 0.0015$, $\eta_p = 0.389$. Note that the dependence of target-localization error on D_{nearest} was similar (Figure 3) and statistically indistinguishable between the two experiments, with all relevant *F*'s (involving experiment as a factor) being less than 1. The normalized spatial distribution of mouse-clicks was also similar for the two experiments (Figure 4).

With regard to the use of motion information, observers in Experiment 2 also tended to click slightly ahead in the direction of the target's motion, yielding a direction tuning similar to that obtained in Experiment 1 (Figure 5). The only difference was that the overall mean direction-tuning index was not as robustly less than 1 in this experiment, $t(21) = 2.020$, $p < 0.056$, $d = 0.463$. This occurred because a few observers yielded direction-tuning indices that were relatively substantially deviated from 1 in the positive direction (indicative of backward clicking), but note that those observers were also poor at localizing targets (see open circles in Figure 6). Importantly, greater direction tuning (i.e., a smaller value of the direction-tuning index indicating a greater tendency for forward clicking) was strongly associated with more precise target localization (i.e., a smaller value of target-localization error), $r = 0.630$, $t(20) = 3.632$, $p < 0.0017$, replicating Experiment 1 with a larger effect size. Thus, anticipating the targets' locations based on their motions increases the precision of multiple-object tracking, whereas the precision suffers when tracking mechanisms fail to anticipate or fall behind. This confirms that motion direction information is utilized by multiple-object tracking mechanisms to increase the precision of target tracking.

The correlation between the amount of forward shifting of mouse-clicks and target speed, r_{speed}, was small ($M = 0.052$ with $SEM = 0.024$) but significant, $t(22) = 2.167$, $p < 0.042$, $d = 0.462$. This replicates Experiment 1 and confirms that target speeds (as well as directions) were encoded during multiple-object tracking. Furthermore, observers who tended to encode targets' speeds with greater reliability (i.e., those with larger positive values of r_{speed}) tended to localize targets more precisely, as reflected in the significant negative correlation between r_{speed} and target-localization error, $r = -0.41$, $t(20) = 2.115$, $p < 0.048$. Note that this correlation was not significant in Experiment 1. It is possible that having no concurrent secondary task and/or the enforcement of central eye fixation in this experiment facilitated the use of speed information.

Finally, as in Experiment 1, the degree to which targets' speeds were monitored (measured by r_{speed}) was not significantly correlated with the degree to which targets' motion directions were monitored (measured by the direction-tuning index), $r = -0.27$, $t(20) = 1.330$, *n.s.*, suggesting that the targets' speeds and directions were monitored as separate parameters during multiple-object tracking. It is, however, possible that a significant correlation might have been obtained had we devised more precise measures of how observers monitored the speeds and directions of target motion, so that this null correlation must be interpreted with caution (although both measures were robust enough to reveal significant correlations with target-localization error).

Overall, the combined results from Experiments 1 and 2 suggest that (1) attention resources are dynamically allocated in a demand-based manner, preferentially to targets in more crowded situations, and that (2) both targets' motion directions and speeds are monitored during multiple-object tracking, with at least the motion direction information substantially contributing to the precision of target localization.

Discussion

People have the ability to simultaneously track multiple moving objects with attention and this ability has been extensively studied. Using a novel paradigm to measure target-localization errors, we investigated two fundamental questions that were not conclusively addressed in previous research. The first question was how attention is distributed among the tracked targets, and the second question was whether motion information (direction and/or speed) is utilized by multiple-object tracking mechanisms. Our results provide important constraints for the current models of multiple-object tracking.

Broadly, there are two classes of models of multiple-object tracking, one postulating that a fixed number of attention indices are available for assignment to tracked objects—"multiple-index" models (for reviews, see Cavanagh & Alvarez, 2005; Pylyshyn, 2001; Scholl, 2001), and the other postulating that a single focus of attention rapidly switches among the tracked objects to update their changing positions—"rapid-switching" models (for a review, see Oksama & Hyönä, 2008). Currently no decisive evidence favors either class of models.

Our primary finding is that the precision of tracking increases for targets in more crowded situations (measured as reduced localization errors for targets in more crowded situations). For the multiple-index type models, this result would suggest that the spatial resolution of an attention index is increased when its assigned target gets more crowded by distractors. This view is consistent with a recent result, suggesting that the resolutions of attention indices are adjustable under the constraint of limited capacity (Alvarez & Franconeri, 2007); for example, many indices might be deployed with coarse spatial resolution, or a few indices might be deployed with fine spatial resolution. Our result would extend this idea in suggesting that even while tracking a fixed number of objects, the spatial resolutions of attention indices can be adaptively adjusted to provide finer resolutions for targets in more crowded situations.

For the rapid-switching type models, our result would suggest that the mechanism that controls the shifting of attention among the tracked targets prioritizes updating of targets in more crowded situations. For either model, our result suggests that increased attention resources (in the form of finer resolution or prioritized updating) are allocated to a target when distractors approach within $\sim 3°$.

Our second major finding is that the motion directions and speeds of the tracked targets are encoded during multiple-object tracking, and that the direction information contributes to the precision of tracking. For the multiple-index type models, this would suggest that the mechanisms by which attention indices follow the targets utilize motion as well as location information. In the rapid-shifting type models, it is postulated that the targets' locations are temporarily stored in visual short-term memory (VSTM) while other targets are being visited and updated; thus, the major source of tracking error according to these models is the discrepancy between the stored location and the actual location of a target when it is re-visited for updating (e.g., Oksama & Hyönä, 2008). Our result would suggest that this discrepancy is reduced by storing the motion (in addition to the location) information in VSTM for each target so that the motion information can be used to appropriately extrapolate a target's location before shifting attention to it (to compensate for its movement while other targets were updated). Note that a recent study suggests that multiple-object tracking mechanisms utilize brief sensory memory (lasting a few hundred milliseconds) of motion trajectories (Narasimhan, Tripathy, & Barrett, 2009). Thus, the motion and location information used for target updating may be extracted from the persisting sensory memory of target trajectories without the use of VSTM.

In summary, any model of multiple-object tracking needs to accommodate our two key findings:

1. attention resources are dynamically and adaptively distributed among the tracked targets so that those in more crowded situations receive more resources than those in less crowded situations, and
2. motion directions (and perhaps also speeds) of the tracked targets are monitored and utilized during multiple-object tracking.

Relation to previous studies

Many prior studies pertain to our finding on the use of motion information during multiple-object tracking. For example, a recent study by Howard and Holcombe (2008) reported a seemingly opposite result; their observers tended to localize a tracked target in its earlier (rather than extrapolated) location. As in our study, Howard and Holcombe measured target-localization errors, but the tracking display they used was somewhat atypical. A square-shaped tracking region was divided into eight sectors of an equal area with each sector containing only one moving item. A subset of these eight items was designated as the targets to be tracked. At the end of each trial, one of the sectors was cued and observers mouse-clicked the last seen location of the target that moved about in the cued sector. Because the targets were always confined within their respective sectors (which were distinct from the sectors that contained the distractors), there was no danger of losing a target as long as observers remembered the target-containing sectors. Thus, rigorous tracking of targets using motion information would not be necessary in their task at least for the purpose of not losing a target. Their observers therefore might have used a less effortful strategy of serially monitoring the target-containing sectors. Such a strategy could result in localization of the target in the cued sector based on its earlier location remembered from the most recent attentional visit to that sector. In contrast, in typical multiple-object tracking tasks (including ours), continuous tracking of all targets is necessary in order not to lose a target because the trajectories of the targets and distractors closely intermingle. In such a demanding case of tracking, encoding the targets' motions to anticipate their future locations would be especially useful. We thus consider the seeming discrepancy between Howard and Holcombe's (2008) results and ours as a reflection of the strategic flexibility of multiple-object tracking mechanisms; a relatively less resource-demanding strategy such as serially monitoring the target-containing regions would be used when such a strategy is sufficient for not losing a target, but all targets are rigorously tracked utilizing most available information (including motion information) when the targets and distractors closely intermingle and the danger of losing a target is high.

In general, prior results on target localization are consistent with the interpretation that the visual system anticipates future locations of objects based on motion-based extrapolation. Notably, the extensive literature on the phenomenon known as representational momentum has demonstrated that when a moving object is suddenly extinguished observers tend to locate it ahead of where it actually disappeared—termed forward displacement, suggesting that the visual system tends to anticipate future locations of moving objects based on their motion trajectories (e.g., Freyd & Finke, 1984; Freyd & Johnson, 1987; Finke & Shyi, 1988; for a review, see Thornton & Hubbard, 2002). The phenomenon of forward displacement has been demonstrated for linear as well as nonlinear motion trajectories such as those with periodic changes in direction and/or speed (e.g., Müsseler et al., 2002; Verfaillie & d'Ydewalle, 1991).

Broadly speaking, forward displacements have been shown to ubiquitously occur when relatively few items are present (typically one or two) in the display, so long as attention is not strongly captured by a distractor at the time of target localization (e.g., Kerzel, 2003b). Forward displacements, however, had not been examined in the context of multiple-object tracking where multiple moving targets are simultaneously tracked in the presence of multiple moving distractors, the targets' and distractors' trajectories intermingle, and when the targets' motions frequently change due to inter-object collisions. We have clearly demonstrated forward-displacement effects in multiple-object tracking, suggesting that the mechanisms underlying the phenomenon of representational momentum are also operational during multiple-object tracking.

The question of whether motion information is utilized by multiple-object tracking mechanisms was previously investigated, but the results were mixed. For example, tracking performance was unaffected whether or not a temporarily occluded target reappeared at the location predicted by the motion trajectory prior to occlusion. Instead, tracking performance was primarily determined by the distance between the points of target disappearance and reappearance (e.g., Franconeri, Pylyshyn, & Scholl, 2006; Keane & Pylyshyn, 2006), suggesting that multiple-object tracking mechanisms relied on proximity rather than motion. In contrast, a recent study showed that, when motion trajectories were nearly constant (except for bouncing at the region boundaries) and only one or two targets were tracked, re-capturing of tracked targets following a 300-ms blank interruption in the tracking display was improved by the presence of motion information prior to the interruption (Fencsik et al., 2007).

Our results strongly support the use of motion information during multiple-object tracking. The consistent forward displacements of mouse-clicks that were positively correlated with the targets' speeds indicated that the targets' motion directions and speeds were both monitored during tracking. The fact that the direction tuning of the response vector robustly correlated with target-localization error in both Experiments 1 and 2 (and r_{speed} correlated with target-localization error in Experiment 2) suggested that multiple-object tracking mechanisms certainly utilize motion direction information (but perhaps also speed information) to increase the precision of tracking. It is possible that this motion-based anticipation of target locations operates only within a short range, both in space and time, shorter than the spatiotemporal dimensions of occluders used in Franconeri et al. (2006) and Keane and Pylyshyn (2006), unless targets' motions are nearly constant as in Fencsik et al. (2007).

Finally, a study investigating how well people detect deviant motion trajectories suggests that the magnitude of deviation that can be detected depends on how many trajectories are simultaneously monitored. It was found that the detectable magnitude of angular deviation was about ±19°, ±38°, or ±76° when the effective number of trajectories tracked by observers was one, two, or four, respectively (Tripathy, Narasimhan, & Barrett, 2007). Our observers tracked three targets, so the expected degree of direction tuning based on Tripathy et al.'s result would be somewhere between ±38° and ±76°. The tuning of the response vector we obtained was about ±70° (Experiment 1) and ±45° (Experiment 2) based on the angular deviations at half amplitude (measured from peak to trough) of the direction-tuning functions shown in Figure 5. Thus, our direction-tuning results are broadly consistent with the result of Tripathy et al. (2007) despite the fact that the two studies used very different experimental paradigms and measures.

Conclusions

We used a novel experimental method to measure target-localization errors while observers tracked multiple targets that independently moved among and collided with distractors. Our results provide evidence that multiple-object tracking mechanisms (1) dynamically distribute attention resources among the tracked targets in a demand-based manner, preferentially allocating resources to targets in more crowded situations (when distractors approach within ~3° of a tracked target), and (2) utilize motion direction (and perhaps also speed) information to anticipate target locations to increase the precision of tracking.

Acknowledgments

We thank Steve Franconeri for help with the eye-tracking aspect of Experiment 2. This research was supported by a National Institutes of Health grant R01 EY018197 and a National Science Foundation grant BCS0643191.

Commercial relationships: none.
Corresponding author: Satoru Suzuki.
Email: satoru@northwestern.edu.
Address: Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208, USA.

Footnotes

[1]According to Pylyshyn and colleagues' original theory (e.g., Pylyshyn, 2003; Pylyshyn & Storm, 1988), these indices that follow their assigned targets operate independently of attention, and they serve as pointers for attention processes to rapidly but serially access the tracked targets. In this sense, their theory is a hybrid between those postulating multiple foci of attention and those postulating rapid switching of a single focus of attention (for reviews of the various models of multiple-object tracking, see Oksama & Hyönä, 2004, 2008, and Cavanagh & Alvarez, 2005).

[2]This range of speed was somewhat slow compared to typical studies examining multiple-object tracking. In most prior studies, the number of successfully tracked targets was the dependent measure, so that the motion needed to be fast enough for most observers to lose track of at least some of the targets. In contrast, our dependent measure was the precision of tracking, so that the motion needed to be slow enough for most observers to successfully track all three targets in most trials. Although this might be considered a potential limitation of our technique, we argue that the speed range we used is relevant on the basis of ecological and neurophysiological considerations. For example, in a typical driving situation (while moving at 35 mi/hr), the stopping distance (under normal road conditions) is about 52 ft. Suppose a driver tracks oncoming traffic (also moving at 35 mi/hr) for potentially dangerous swerves from about twice the stopping distance, 104 (= 52 × 2) ft, to avoid collision. At this distance, the oncoming traffic (assuming a typical 2-lane road) would move toward the retinal periphery at ~7°/sec (assuming the driver is looking forward). The retinal speeds of the cars that the driver is following would typically be in a slower range. A stationary object such as a pedestrian would move toward the periphery at ~3.3°/sec. If a pedestrian was walking (~2.5 mi/hr) or running (~6.3 mi/hr), that would add about ±0° to 2°/sec (if walking) or ±0° to 5°/sec (if running) of retinal speed depending on the direction in which the pedestrian is moving. As can be seen from this example, the range of speeds that we used in our study (0.12–6.15°/sec) covers much of this typical range of retinal speeds of objects that need to be tracked while driving. Furthermore, the average speed tunings of visual neurons in areas V1 and MT are 4.47°/sec and 7.52°/sec, respectively (Priebe, Cassanello, & Lisberger, 2003; Priebe, Lisberger, & Movshon, 2006), within or near the range of speeds included in this study. The speed range we used is thus representative of normal experience on the basis of both ecological and neurophysiological considerations.

References

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

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

Fencsik, 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. [PubMed]

Finke, R. A., & Shyi, G. C. (1988). Mental extrapolation and representational momentum for complex implied motions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 112–120. [PubMed]

Franconeri, S., Pylyshyn, Z. W., & Scholl, B. J. (2006). Spatiotemporal cues for tracking multiple objects through occlusion. *Visual Cognition, 14,* 100–104.

Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 126–132.

Freyd, J. J., & Johnson, J. Q. (1987). Probing the time course of representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 259–268. [PubMed]

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. [PubMed]

Keane, B. P., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cognitive Psychology, 52,* 346–368. [PubMed]

Kerzel, D. (2003a). Attention maintains mental extrapolation of target position: Irrelevant distractors eliminate forward displacement after implied motion. *Cognition, 88,* 109–131. [PubMed]

Kerzel, D. (2003b). Mental extrapolation of target position is strongest with weak motion signals and motor responses. *Vision Research, 43,* 2623–2635. [PubMed]

Müsseler, J., Stork, S., & Kerzel, D. (2002). Comparing mislocalizations with moving stimuli: The Fröhlich effect, the flash-lag, and representational momentum. *Visual Cognition, 9,* 120–138.

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. [PubMed]

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

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

Priebe, N. J., Cassanello, C. R., & Lisberger, S. G. (2003). The neural representation of speed in macaque area MT/V5. *Journal of Neuroscience, 23,* 5650–5661. [PubMed] [Article]

Priebe, N. J., Lisberger, S. G., & Movshon, J. A. (2006). Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *Journal of Neuroscience, 26,* 2941–2950. [PubMed] [Article]

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

Pylyshyn, Z. W. (2003). *Seeing and visualizing*. Cambridge, MA: MIT Press.

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

Thornton, I. M., & Hubbard, T. L. (2002). Representational momentum: New findings, new directions. *Visual Cognition, 9,* 1–7.

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]

Verfaillie, K., & d'Ydewalle, G. (1991). Representational momentum and event course anticipation in the perception of implied periodical motions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 302–313. [PubMed]

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

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1: Please provide e-mail and Web address, if any.
AQ2: Please provide Web address, if any.

END OF AUTHOR QUERIES