Tracking multiple objects is limited only by spatial interference, not speed, time, or capacity

Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M.
Northwestern University

Please address correspondence to:

Steve Franconeri
Northwestern University
2029 Sheridan Rd, Evanston, IL 60208
Phone: 847-491-4994
Fax: 847-491-7859
franconeri@northwestern.edu

RUNNING HEAD: Multiple object tracking
WORD COUNT: 2497

Abstract

In dealing with a dynamic world, we have the ability to maintain selective attention on a subset of moving objects in the environment. Our performance in such tasks is limited by three primary factors - the number of objects that we can track, the speed at which we can track them, and how close together they can be. We argue that a form of this last limit, which we label *spatial interference*, is the root cause of all performance constraints in multiple object tracking tasks. In two experiments, we show that a correlate of spatial interference, the distance that objects travel, can account for performance differences across a wide range of object speeds and tracking task lengths. These results suggest that barring spatial interference, we could reliably track an unlimited number of objects as fast as we could track a single object.

[142 words]

Keywords: Multiple object tracking, MOT, Crowding, Spatial interference, Divided attention

In daily visual life, objects in the world shift drastically across the retina as their positions move relative to our field of view. Despite these dynamic changes, these objects must be continuously selected if they are to be monitored, compared, or encoded in memory. To explore our ability to maintain attention on more than one object at a time, researchers often rely on the Multiple Object Tracking (MOT) task. This task requires observers to mentally track a set of target objects moving among featurally identical distractor objects (Pylyshyn & Storm, 1988), similar to the street magician who places an object under one of several quickly moving inverted cups.

Performance in this task reveals several limits on tracking abilities. First, there is a limit on capacity, the number of objects that can be tracked concurrently. Initially, many results suggested that this limit was four objects (Intriligator & Cavanagh, 2001; Pylyshyn & Storm, 1988; Yantis, 1992), but later work demonstrated that with some methodological changes tracking capacity could reach 8 or 9 objects (Alvarez & Franconeri, 2007). Second, there is an interaction of capacity with object speed. As more objects are tracked, they must move more slowly in order to maintain the same level of accuracy (Alvarez & Franconeri, 2007). Third, there is a limit from object proximity, where objects that come too close to other objects are more likely to be lost (Franconeri, et. al., 2008; Intrilligator & Cavanagh, 2001; Pylyshyn, 2004; Shim, Alvarez, & Jiang, 2008; Tombu & Seiffert, 2008).

These limits constrain the potential underlying mechanisms that our visual system might use to concurrently track multiple objects. Past accounts have explained these limits by positing a set number of trackers (Pylyshyn & Storm, 1988), a variable number of trackers (Alvarez & Franconeri, 2007), or memory for the global shape created by the targets (Yantis, 1992). Here we propose a more parsimonious account that predicts the limits on capacity, the interaction of capacity with speed, and object proximity limits using a single known limitation of the visual system. Under this *spatial interference* account, there is no limit on the number of 'trackers', and no limit per se on tracking capacity. Instead, tracking is accomplished in parallel for an unlimited number of objects at once. One implementation would be local and independent neural circuits that maintain a local activation peak for a tracked object (Koch & Ullman, 1985; Pylyshyn.

2000). While there would be a speed limit for how quickly this peak could shift, the critical speed would be the same for one object as for N objects, and there should be no interaction of tracking capacity with speed. The important limit for this mechanism would be spatial interference among tracked targets. Because a locus of spatial attention is known to have a suppressive surround (Bahcall & Kowler, 1999; Cutzu & Tsotsos, 2003; Hopf et. al., 2006; Tsotsos et. al., 1995), multiple loci of spatial attention should inhibit each other if they are within a critical distance, creating noise in the selection region.

Spatial interference can explain capacity limits in MOT, because the capacity for independently selecting *static* locations is also 8-9 objects, with capacities diminishing rapidly as the selected objects are placed closer together (Franconeri et. al., 2007). Such object proximity limits would stem from two sources. When targets move closer to each other, the interference from suppressive surrounds of the selection regions would increase, creating noise in the target locations representations. Recent studies using MOT tasks have shown that decreased target-target distance impairs MOT performance (Pylyshyn, 2004; Shim, Alvarez, & Jiang, 2008). Moving targets closer to distractors would increase the likelihood that representations of the target locations would fail to exclude nearby distractors (Intrilligator & Cavanagh, 2001; Pylyshyn, 2004).

However, one prediction of the spatial interference account is not as intuitive. If each object is tracked independently, there should be a constant upper limit on speed for each object, but that speed limit should not interact with the number of tracked objects. So why should increased speed lead to lower tracking capacity? We suggest that increasing speed increases the number of close interactions among objects. One previous study supports this possibility (Franconeri et. al., 2008). Participants tracked a set of objects in a small tracking display, and in a display magnified by fourfold, which increased all speeds by fourfold. If speed were the limiting factor, then performance should have dropped dramatically. However, because the degree of spatial interference should not change with screen magnification (e.g., Toet & Levi, 1992), there should be no difference in the distribution of close interactions. Accuracy levels were highly similar across the two conditions, suggesting that object speed only affected MOT performance through its impact on the distribution of spatial interactions among the objects. However, because it is possible that the display scaling manipulation altered other aspects of the tracking display, such as the spatial frequency profile of the moving objects, more evidence is needed to support the spatial interference account.

The present study provides direct evidence that spatial interference is the root cause of limits of MOT performance. If the critical factor limiting performance is the number of times that objects pass too close to one another, then performance should be primarily limited not by object speed, but by the distance the objects travel. If objects moved at 10 degrees/second for 10 seconds, then if the same animation were played in 'fast forward', showing the same animation in half or one quarter of the time, the distance covered by each object would be identical, making the distribution of object distances across the animation identical. The spatial interference account predicts that performance across these conditions should be identical, despite large changes in object speed. In Experiments 1a and 1b we test MOT performance under a variety of speed and time combinations, chosen to have multiple instances of the same distance but with widely varied speeds.

Experiment 1

Experiment 1a tests four combinations of speed and time (see x-axis of Figure 2a for details) including a variety of cumulative distances. Experiment 1b replicates and extends this result using six combinations (Figure 2b).

Methods

Participants

Twenty-three observers participated in Experiment 1a, and twenty-four in 1b, in exchange for course credit or payment. Some participants were removed from the analysis (3 in 1a, 5 in 1b) because they were not able to track objects with at least 75% accuracy in the easiest short distance condition.

Stimuli

All experiments were run on Intel Macintosh computers using MATLAB 7.6 and Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997). Figure 1 illustrates the stimuli display. A 15" Viewsonic monitor ran at 640x480 resolution at 120 Hz approximately 50cm from the participant. Distances are reported in pixels (~18 pixels/degree). Twelve black circles (0 cd/m^2) (diameter 8 pixels) were present on a white (approximately 70 cd/m^2) background. Targets and distractors were paired (intervening diameter of 55 or 110 pixels for the center or corner pairs) always remaining 180 degrees apart as each pair orbited an imaginary center point. The centerpoints for the four outer pairs were on the corners of an imaginary square 300 pixels wide, and centerpoints of the two middle pairs were 60 pixels above and below the fixation point.

Object pairs rotated around their centerpoint in a clockwise or counter-clockwise pattern, always at a set speed with instantaneous transitions in direction. Rotation speed was between 0.167 and 1.6 rotations per second, and tracking was between 1.5 and 6 seconds. Objects in each pair randomly and independently changed directions, traveling at least 0.1 and at most 2 revolutions before changing direction, randomly chosen from a rectangular distribution.

Procedure

Subjects were given strict fixation instructions. At the start of a trial, all 12 circles appeared on the screen and began moving, with targets cued in red for 2 seconds. Targets would then turn black, and participants tracked the six targets for the designated time period. In Exp 1a, objects would slow exponentially over the final half second of the time period, while in Exp 1b the objects would abruptly stop. The participant then heard a voice cue to click the targets within the "top" or "bottom" three pairs of objects. This partial report was used after participants reported forgetting known targets during pilot experiments while clicking on all 6 objects. After selecting all known or guessed targets pressing the space bar registered 50% (chance) accuracy on all remaining targets. Object speeds and tracking times are listed in Figures 2a and 2b. Each of these conditions was blocked with 20 (1a) or 18 (1b) trials per block, with block order randomized across subjects. Each subject received 5 practice trials using the shortest distance condition. Exp 1a lasted approximately 40 minutes, and 1b 45 minutes.

Results & Discussion

Figure 2a depicts accuracy rates for Experiment 1a. Accuracy values were submitted to an ANOVA (some with Greenhouse-Geisser corrections), revealing a main effect of condition, $F(3,57)=88$, $p<.001$, $\eta_p^2=0.822$. From left to right in Figure 2a (in order of increasing speed), accuracy was highest for the shortest distance condition (M= 92.2%) compared to the two medium distance conditions (M= 74.5%, 74.2%), both $t(19)>8.2$, $p<.001$, $d>2.20$. The two medium distance conditions were equal, $t(19)<1$, but were both higher than the long distance condition (M= 61.2%), both $t(19)>6.9$, $p<.001$, $d>1.49$.

Figure 2b depicts accuracy rates for Experiment 1b. An ANOVA revealed a main effect of condition, $F(5,90)=40$, $p<.001$, $\eta_p^2=0.688$. From left to right in Figure 2b (in order of increasing speed), accuracy was highest for the shortest distance condition (M=86.4%), relative to all other conditions, all $t(18)>4.6$, $p<.001$, $d>1.67$. There were no significant differences among the second (M=71.5%), third (M=74.5%), and fourth (M=71.7%) conditions. The fifth condition had lower accuracy (M=58.4%) than all other conditions, all $t(18)>4.9$, $p<.001$, $d>1.11$. The sixth and highest speed condition showed a medium accuracy level (M=66.0%) that was slightly lower than the second and third conditions of equivalent distance, both $t(18)>2.7$, $p<.02$, $d>0.71$, but not the fourth condition, $t(18)=1.8$, $p=.09$, $d=0.62$. Accuracy was again best captured by differences in distance traveled, not speed or time.

One result in Experiment 1b seems at first incongruous with the spatial interference account. The last condition (6th bar in Figure 2b) has a distance equal to the three medium distances, yet has an accuracy level that is slightly (M=6.6%) lower. Is this evidence for an interaction between tracking capacity and speed even when distance is controlled? We think not. Speed can have an effect for two reasons. First, even if all objects are tracked in parallel with independent speed limits, there is still an independent speed limit. This limit might be due to loss of faster targets that move farther during moments of inattention, sudden sounds nearby, or eyeblinks. See Norman & Bobrow (1975) for a related dissociation, between *data-limited* and *resource-limited* processes (also discussed in Alvarez & Franconeri, 2007). This type of speed limit should not interact with the number of objects tracked.

Therefore, in a separate control experiment (N=8) using identical displays, participants tracked only 2 objects. We used 2 tracked objects instead of 1 to prevent participants from using eye movements. This set size should reflect the same performance levels as tracking 1 object if the objects are always located in separate visual hemifields (Alvarez & Cavanagh, 2005). The targets were always drawn from the 2 diagonally opposite corner pairs, and both targets were reported. There were two speed conditions, slow (0.4 r/s) but long (6 sec) and fast (1.6r/s) but short (1.5 sec), with equal distances. Performance was higher in the slow condition (M=93.7%) compared to the fast condition (82.7%), $t(7)=3.6$, $p=0.017$, suggesting that the speed impairment of the last condition in Experiment 1b was not due to an interaction of capacity and speed, but to a main effect of speed.

Figure 3 depicts the results of both experiments (accuracy levels for 10 total conditions) arranged by distance, speed, and time. The distance panel shows that distance best accounts for the variance in tracking accuracy. The relationship appears to be logarithmic, and indeed when the x-axis is log-transformed the function becomes highly linear, with R^2 values of 0.85 including all data form both experiments, 0.92 excluding the single speed-limited point, and 0.98 when further excluding E1a (to compare only within a single group of subjects). This logarithmic relationship (seen also in Alvarez & Franconeri, 2007) is likely due to the diminishing impact of greater distance on accuracy. If object proximity results in an unrecoverable target loss, the impact of proximity (or any other factor that impairs tracking) should be greater when more targets are still tracked.

In contrast to the distance panel, the speed panel shows roughly constant accuracy levels across a fourfold difference in object speed. Note that the few points in this panel that seem to indicate a relationship (the two highest and two lowest) are all confounded with low and high distance, and can be seen in similar positions within the distance panel. The time panel shows no relationship to accuracy.

Conclusions

Across 10 combinations of object speed and tracking time, we found that *distance traveled* was by far the best predictor of tracking accuracy. This result is consistent with an account where all limits on multiple object tracking have their origin in spatial interference among the tracked objects. This single parsimonious explanation can predict a large swath of previous results in MOT tasks. It correctly predicts that when interference is blocked by placing tracked objects in separate hemifields or quadrants where spatial interactions are eliminated or reduced (Chakravarthi & Cavanagh, 2009; Liu, Jiang, Sun, & He, 2009), tracking performance would be fully or partially independent for those objects (Alvarez & Cavanagh, 2005; Carlson et. al., 2007). It also predicts that if the speed of all objects in a tracking display is increased by also translating the display as a whole, which will not change relative inter-object distances, performance would be unaffected (Liu et. al., 2005). Asking participants to track for longer periods of time should increase total distance traveled, which would impair accuracy (Oksama & Hyona, 2004). Constraining the global virtual polygon created by the target objects to remain convex should serve to keep targets farther apart, and should lead to better performance (Yantis, 1992). Video game training that tends to improve MOT performance (Green & Bevelier, 2003) could do so by improving participant's spatial resolution for object interactions (Green & Bevelier, 2007). Observers should be equally successful at tracking many moving objects as they are at tracking mixed collections of moving objects and static locations (Howe, Cohen, Pinto, & Horowitz, 2009).

The spatial interference account provides a concrete mechanism and moves beyond redescriptions of tracking limits that label tracking as 'resource dependent' or 'requiring attention'. Instead, it presents a simple and falsifiable hypothesis of the limits underlying our ability to track multiple objects at once. We hope that future work will challenge whether this explanation alone can account for all limits in our ability to maintain our selection of multiple objects in the environment.

Figure captions

Figure 1: A sample tracking display for Experiments 1a and 1b.

Figure 2: (a) Tracking accuracy for Experiment 1a, and (b) 1b. R stands for Revolutions, and R/S stands for Revolutions/Second. Note that one of the medium distance combinations in Experiment 1b also employs a speed that is significantly faster than the speed used in the longest distance condition.

Figure 3: Tracking accuracy for the 10 total conditions in Experiments 1a (gray ovals) & 1b (black ovals), organized by distance, speed, and time.

References

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637-6.

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.

Bahcall, D. O. & Kowler, E. (1999). Attentional interference at small spatial separations. *Vision Research 39*, 71-86.

Brainard, D. H. (1997) The Psychophysics Toolbox, *Spatial Vision, 10*, 433-436.

Carlson, T. A., Alvarez, G. A., & Cavanagh, P. C. (2007). Quadrantic deficit reveals anatomical constraints on selection. *Proceedings of the National Academy of Sciences, 104*(33), 13496-13500.

Chakravarthi, R. & Cavanagh, C. (2009). Bilateral field advantage in visual crowding. *Vision Research, 49*, 1638-1646.

Cutzu, F. & Tsotsos, J. K. (2003). The selective tuning model of visual attention: Testing the predictions arising from the inhibitory surround mechanism. *Vision Research, 43*, 205–219.

Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can you select? *Journal of Experimental Psychology: Human Perception & Performance, 33*(5), 1003-1012.

Franconeri, S. L., Lin, J., Pylyshyn, Z. W., Fisher, B., & Enns, J. T. (2008). Multiple object tracking is limited by crowding, but not speed. *Psychonomic Bulletin & Review. 15*(4), 802-808.

Green, C.S. & Bavelier, D. (2003). Action video games modify visual selective attention. *Nature, 423*, 534-537.

Green, C.S. & Bavelier, D. (2007). Action video game experience alters the spatial resolution of attention. *Psychological Science, 18*(1), 88-94.

Hopf, J. M., Boehler, C. N. Luck, S. J., Tsotsos, J. K., Heinze, H. J., & Schoenfeld, A. M. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proceedings of the National Academy of Sciences*, *103*, 1053-1058.

Howe P.D., Cohen M., Yair P., & Horowitz T.S. (2009). Distinguishing between parallel and serial accounts of multiple object tracking. *Vision Sciences 2009*, Sarasota, FL.

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

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4*, 219-227.

Liu, G., Austen, E.L., Booth, K.S., Fisher, B., Rempel, M.I., & Enns, J. T. (2005). Multiple object tracking is based on scene not retinal coordinates. *Journal of Experimental Psychology: Human Perception & Performance, 31*, 235-247.

Liu T., Jiang Y., Sun X. & He S. (2009). Reduction of the Crowding Effect in Spatially Adjacent but Cortically Remote Visual Stimuli. *Current Biology, 19*(2), 127-132.

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

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

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

Pylyshyn, Z. W. (2000). Situating vision in the world. *Trends in Cognitive Sciences, 4*(5), 197-207

Pylyshyn, Z.W. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition, 11*, 801-822.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 1-19.

Shim, W.M., Alvarez, G.A., & Jiang, Y.V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychonomic Bulletin & Review, 15*(2), 390-397.

Toet, A. & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research, 32(*7). 1349-1357.

Tombu, M. & Seiffert, A. E. (2008). Attentional costs in multiple object tracking. *Cognition, 108*(1), 1-25.

Tsotsos, J. K., Culhane, S. W., Wai, Y. L., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence, 78*, 507–547.

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

Figure 1

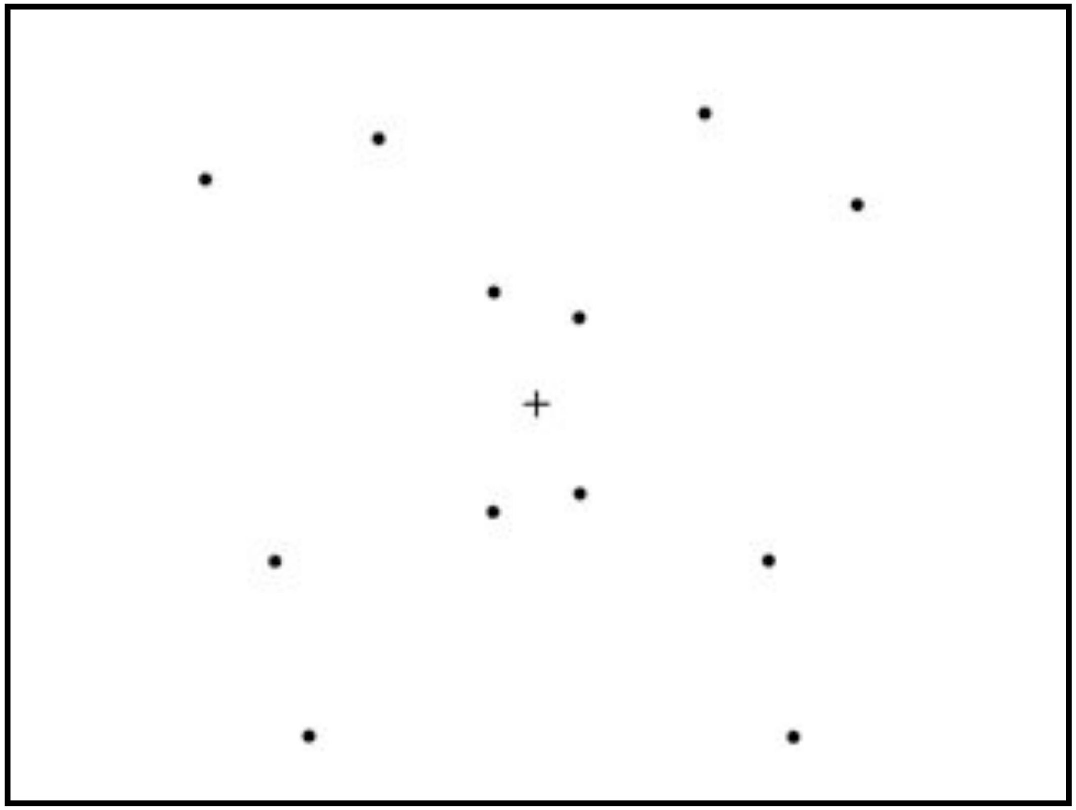
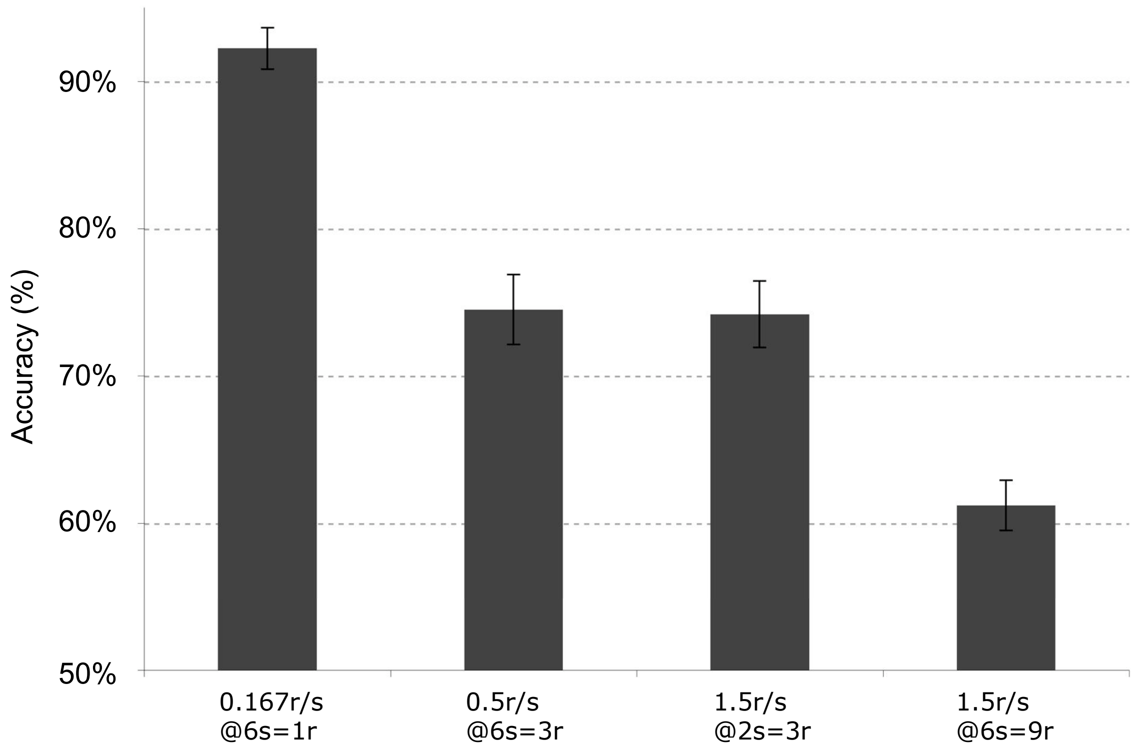


Figure 2a



Accuracy (%)
90%
80%
70%
60%
50%
0.167r/s @6s=1r
0.5r/s @6s=3r
1.5r/s @2s=3r
1.5r/s @6s=9r

Figure 2b

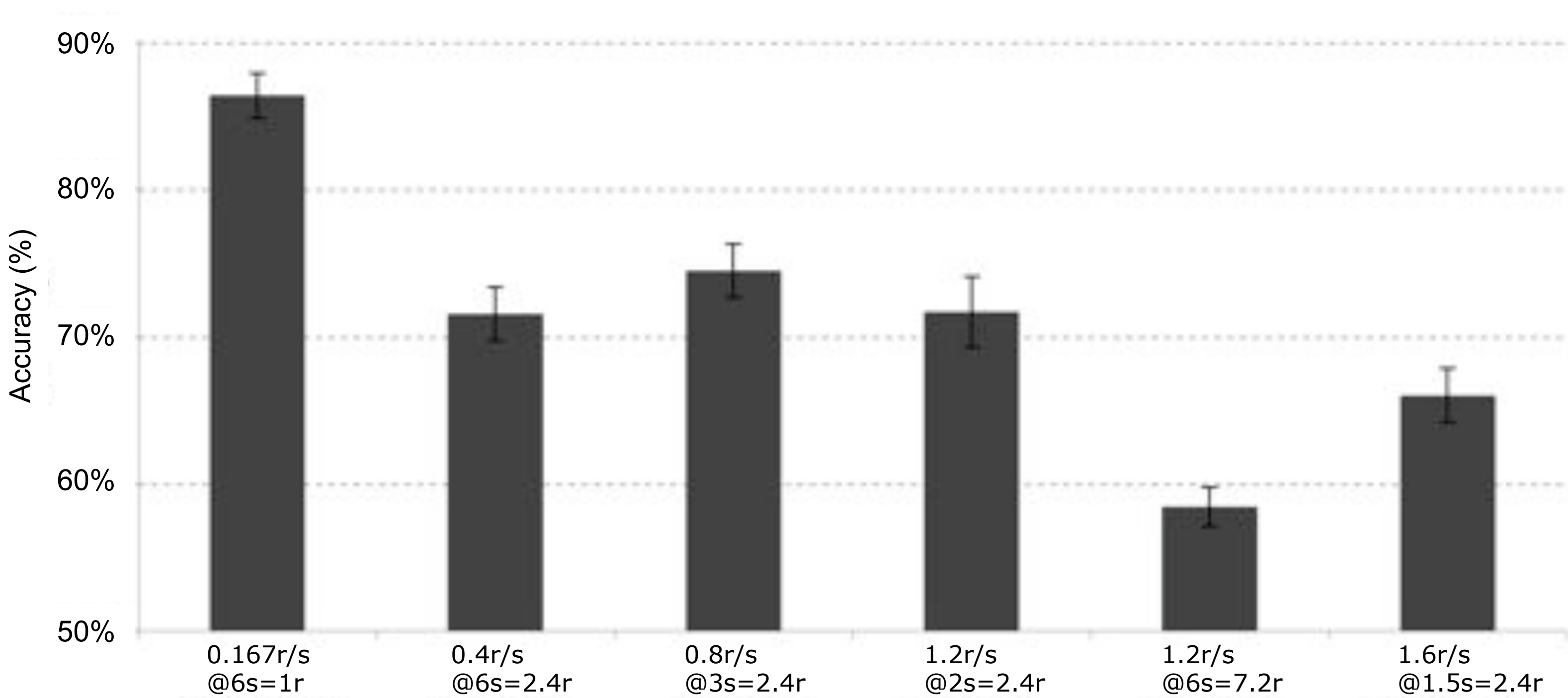
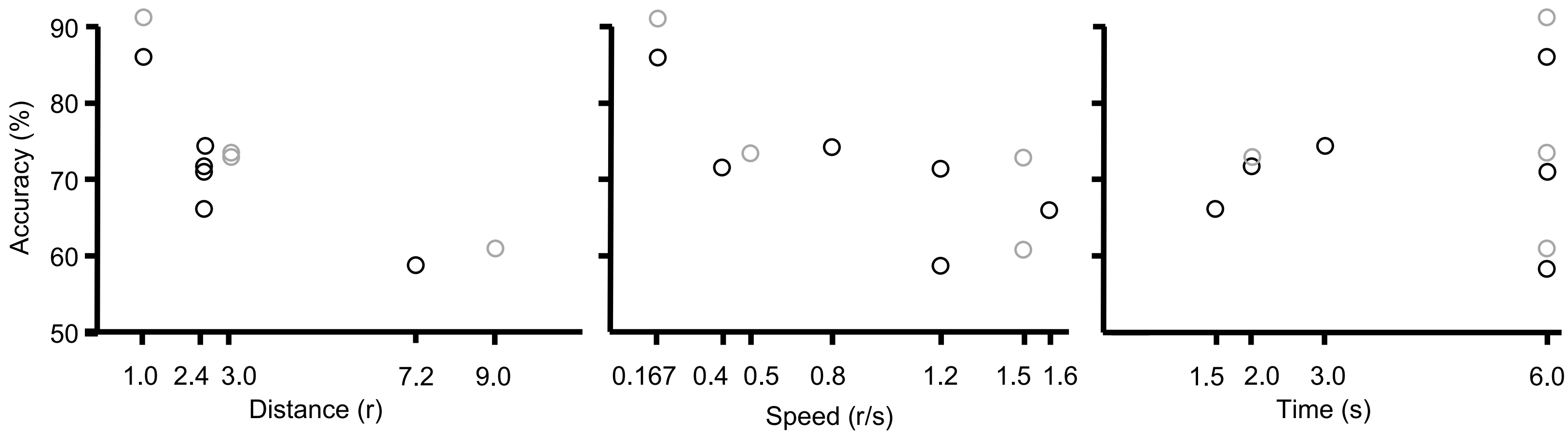

90%
80%
70%
60%
50%
Accuracy (%)
0.167r/s @6s=1r
0.4r/s @6s=2.4r
0.8r/s @3s=2.4r
1.2r/s @2s=2.4r
1.2r/s @6s=7.2r
1.6r/s @1.5s=2.4r

Figure 3


Accuracy (%)
90
80
70
60
50
1.0
2.4
3.0
7.2
9.0
Distance (r)
0.167
0.4
0.5
0.8
1.2
1.5
1.6
Speed (r/s)
1.5
2.0
3.0
6.0
Time (s)

Supplemental material: Franconeri, Jonathan, & Scimeca

The results of Experiments 1a and 1b clearly show that distance is a critical factor in multiple object tracking performance. We therefore conducted an additional analysis to show that targets that passed closer to other targets were more likely to be lost (see also Pylyshyn, 2004; Shim et. al., 2008). Using results from the four equal-distance conditions from Experiment 1b, we calculated tracking accuracy for each target based on how much time it spent relatively close to other targets in the display. Specifically, we measured accuracy for corner objects based on their distance to horizontally matched targets as well as vertically matched targets, and we measured accuracy for center targets based on their distance to the other center target. These measurements were then collapsed across screen locations into target types (corner or center), to provide accuracy levels for corner objects with respect to their vertical or horizontal counterparts, and for center targets with respect to their vertical target counterparts. The distance measurement was the percentage of animation frames that the target spent more than 300 pixels (the distance between the centerpoints of the groups) from the opposing target (120 pixels for the center objects). These distance measurements were then binned into 10% increments, and only the 25-35%, 35-45%, 45-55%, and 55-65% bins contained a sufficient number of trials for analysis.

The resulting accuracy data (see Supplemental Figure) were submitted to a 3 (distance type: vertical for corner objects (grey line), horizontal for corner objects (dotted line), and vertical for center objects (dashed line) x 4 (time spent; 4 percentage bins on x-axis) repeated measures ANOVA. There were no main effects of distance type or time spent, both $F<1$, but there was an interaction between these factors, $F(6,108)=4.2$, $p<.01$, $\eta_p^2=0.188$, suggesting that the different types of target-target distance affected tracking performance in different ways. Individual ANOVAs examining time spent for each distance type showed that vertical distance was helpful for corner objects, $F(3,54)=5.0$, $p<.01$, $\eta_p^2=0.219$ but horizontal distance did not affect accuracy $F<1.2$. For the central objects, an opposite trend emerged where performance was better when objects were *closer* to each other, $F(3,54)=4.0$, $p<.025$, $\eta_p^2=0.182$. Further ANOVA analysis using each combination of 2 distance types, across time spent bins, showed that there was no significant difference between the effect of vertical and horizontal distance for corner objects. That is, the dotted and dashed lines in Figure 4 do not significantly interact. However, the effect of time spent on center objects (grey line) was significantly different than for vertical distance for corner objects (dashed line), $F(3,54)=6.0$, $p<.005$, $\eta_p^2=0.250$, and also different than for horizontal distance for corner objects (dotted line), $F(3,54)=4.6$, $p<.025$, $\eta_p^2=0.205$.

For corner objects, performance was worse when targets were closer together. There was a trend for this effect to be stronger for vertical distance, consistent with other results showing spatial interactions to be stronger across vertical distances within a visual hemifield relative to horizontal distances across visual hemifields (Carlson, Alvarez, & Cavanagh, 2007). While these vertical and horizontal distances were only loosely related to distance from fixation, this was not true for center objects where target proximity related strongly to eccentricity. These objects showed the reverse pattern, where closer targets were easier to track. This reversal is likely due to the scope of spatial interactions becoming rapidly smaller as objects move closer to the fovea (Toet & Levi, 1992).

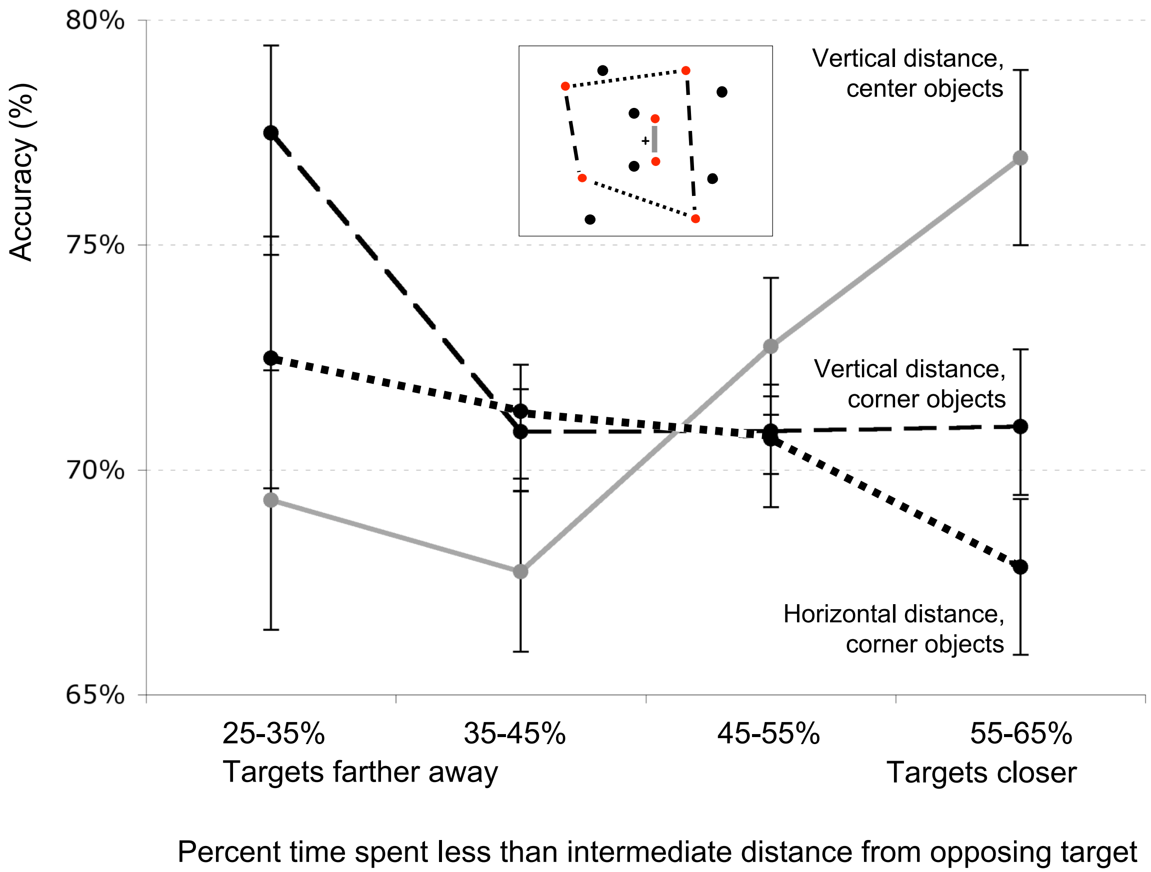
References

Pylyshyn, Z.W. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition, 11*, 801-822.

Shim, W.M., Alvarez, G.A., & Jiang, Y.V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychonomic Bulletin & Review, 15*(2), 390-397.

Toet, A. & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research, 32(*7). 1349-1357.

Carlson, T. A., Alvarez, G. A., & Cavanagh, P. C. (2007). Quadrantic deficit reveals anatomical constraints on selection. *Proceedings of the National Academy of Sciences, 104*(33), 13496-13500.



Supplemental Figure: Accuracy for a given target type, according to the percent of total trial time it spent less than 300 pixels (the distance between pair centerpoints) from the opposing target. The inset tracking display depicts the three types of distance relationships.