

Evidence against a speed limit in multiple-object tracking

S. L. FRANCONERI

Northwestern University, Evanston, Illinois

J. Y. LIN

University of British Columbia, Vancouver, British Columbia, Canada

Z. W. PYLYSHYN

Rutgers University, New Brunswick, New Jersey

B. FISHER

Simon Fraser University, Vancouver, British Columbia, Canada

AND

J. T. ENNS

University of British Columbia, Vancouver, British Columbia, Canada

Everyday tasks often require us to keep track of multiple objects in dynamic scenes. Past studies show that tracking becomes more difficult as objects move faster. In the present study, we show that this trade-off may not be due to increased speed itself but may, instead, be due to the increased crowding that usually accompanies increases in speed. Here, we isolate changes in speed from variations in crowding, by projecting a tracking display either onto a small area at the center of a hemispheric projection dome or onto the entire dome. Use of the larger display increased retinal image size and object speed by a factor of 4 but did not increase interobject crowding. Results showed that tracking accuracy was equally good in the large-display condition, even when the objects traveled far into the visual periphery. Accuracy was also not reduced when we tested object speeds that limited performance in the small-display condition. These results, along with a reinterpretation of past studies, suggest that we might be able to track multiple moving objects as fast as we can a single moving object, once the effect of object crowding is eliminated.

The visual system takes a 2-D array of light and creates representations that support useful action in a 3-D world. One of these representations focuses on selecting and monitoring objects over time (Pylyshyn & Storm, 1988; Scholl, 2001). In a typical laboratory task designed to study these representations, a set of identical shapes appear on a computer display, and a subset of them flash to indicate that they are target objects. All of the objects then move randomly about the screen for several seconds. At the end of the trial, the observer reports which objects were the original targets.

Performance in these multiple-object tracking tasks has salient limits, with accuracy dropping precipitously when observers try to track too many objects or when objects move too quickly (Alvarez & Franconeri, 2007; Liu et al., 2005; Pylyshyn & Storm, 1988). Various mechanisms have been proposed to explain these limits. For example, capacity limits may reflect a fundamental architectural limit of the visual system (Pylyshyn & Storm, 1988) or a limit on

attentional resources (Alvarez & Franconeri, 2007; Cavanagh & Alvarez, 2005). The decrease in accuracy with faster object speeds could also reflect an increased reliance on motion-sensitive cells with larger receptive fields, which provide coarser positional information about each target object's location (Alvarez & Franconeri, 2007).

Here, we explore the possibility that the limits on tracking capacity and object speed are both rooted in the visual system's limited *attentional resolution*—the ability to individuate and select an individual object from its neighbors. This failure of individuation due to proximity (*crowding*) is of particular relevance when tracked objects and their neighbors are identical (Intriligator & Cavanagh, 2001), because the maintenance of a distinct object identity then depends entirely on consistent individuation over time.

How might crowding explain limits on both capacity and speed in tracking? Capacity limits may arise because objects are necessarily pushed closer together as more objects are added to a tracking display. Once interobject

S. L. Franconeri, franconeri@northwestern.edu

distances become small enough that objects lose their unique representations, tracking will suffer as well. The effect of speed on crowding is more indirect. If two objects pass within a threshold distance, observers may have more difficulty maintaining selection of the target. Once a target object is lost, it cannot be recovered, because it is identical in appearance to the distractor objects. The more frequently the threshold distance is crossed, the higher the probability that a target will be lost. Thus, crowding could account not only for the impairment caused by increasing the number of objects in a display,¹ but also for the impairment caused by increased object speed.

There is already evidence that crowding can account for the performance drop caused by adding more objects to a display. When the number of objects in a display is kept constant, but the objects are spaced more densely, tracking performance drops sharply (Alvarez & Franconeri, 2007). One study found that the targets that pass close to distractors are the most likely to be lost (Pylyshyn, 2004), but even pushing two targets closer to one another can impair tracking, possibly because attended targets inhibit each other (Shim, Alvarez, & Jiang, 2008).

It is more difficult to disentangle object speed from the number of instances of crowding that occur per unit time. Here, we exploit a technique that allows us to manipulate speed while keeping crowding constant. As objects move farther from the fovea, the interobject distance required for individual location representation increases proportionally (Pelli, Palomares, & Majaj, 2004; Toet & Levi, 1992)—meaning that when a pair of objects doubles its distance from the fovea, it doubles its “mandatory separation zone” as well. Because of this proportional relationship, it is possible to scale the size of a tracking display linearly without changing the interobject crowding. At the same time, a display that is scaled to a larger size contains objects that move faster across the screen and, therefore, across the retina.

We used this scale manipulation in two experiments to show that when crowding is held constant, tracking accuracy is unchanged, even when displays are four times as

large (Experiment 1). Tracking accuracy is not impaired even when larger display sizes also lead to a fourfold speed increase (Experiment 2). We then demonstrate that the use of a large display nonetheless leaves tracking accuracy just as vulnerable to crowding (Experiment 3). These data suggest that object speed per se may not be a limiting factor in multiple-object tracking, but that speed limits tracking through its effect on crowding.

EXPERIMENT 1 Capacity Limits in Tracking in Large and Small Fields

In Experiment 1, we compared tracking on a small display typical of past studies and on a large display, scaled to be four times larger. Although object crowding should remain equal across the two display-size conditions, objects will move at four times the speed on the larger display. If crowding, and not speed, is the limiting factor on tracking, then performance should be equal across the two display-size conditions.

Method

Participants. Twelve undergraduate students participated in exchange for course credit. One participant, who could not track a single object with at least 80% accuracy in both display conditions, was removed from the analysis.

Stimuli. Figure 1 illustrates the display dome and the stimuli. Participants were seated 2.70 m from an Elumens VS3 hemispheric projection dome, which was 3.47 m wide \times 2.37 m high \times 1.47 m deep. Displays were projected at 85 Hz through a hemispherical lens to prevent image distortion, and Elumens software was used to transform the image, in order to ensure that pixels were evenly spaced across the inner surface of the dome. The tracking of displays and the input of responses were controlled by a Macintosh Powerbook G3, using the VisionShell libraries (www.visionshell.com). The large display was 88° wide \times 44° high (750 \times 360 pixels) and filled most of the projection dome. All aspects of the small tracking display, including all object sizes and spatial relationships, were reduced by a factor of four, relative to the large display (22° wide \times 10.6° high [188 \times 90 pixels]).

Objects were 16 small black discs (16 pixels in diameter in the large-display condition, 4 pixels in the small-display condition) on a

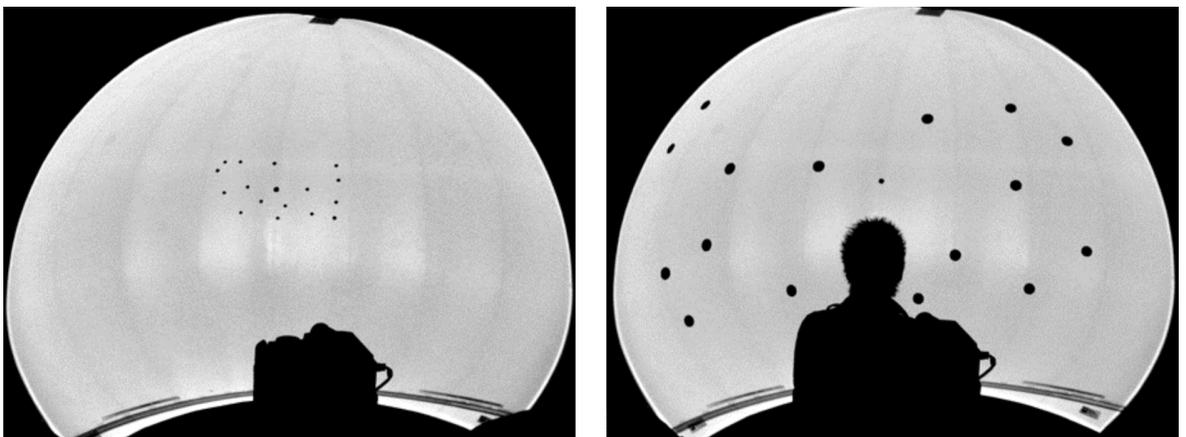


Figure 1. Photographs of typical displays in Experiment 1. The small (left) and large (right) displays differ in size by a factor of four. As shown at right, participants were seated directly behind the projector.

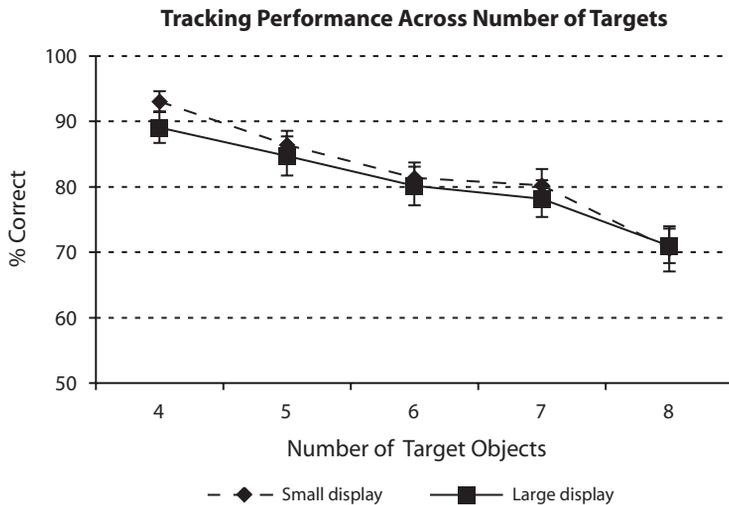


Figure 2. Mean accuracy in the small- and large-display conditions in Experiment 1. Error bars represent standard errors.

white background. Each object's motion path was determined 45% by an initially randomly generated independent motion direction, and 55% by repulsion from other objects, which began when objects were within 100 pixels (25 pixels in the small-display condition) of each other (all reported interobject distances are center to center), and increased with shorter distances, proportional to the inverse of the square of the distance between the items. The repulsion changed only the direction of the object's motion, not its speed. At the end of each video frame, each object's independent motion direction was changed by an angle from 0° to 1.2° . This same angular change was maintained in the next video frame as a form of angle-change inertia, and this inertia value could be randomly altered by a maximum of 0.3° per frame. In addition, if an object's nearest neighbor was within 80 pixels (20 pixels in the small-display condition), and the objects were headed toward each other, each object's internal angle of motion was altered by 8.5° in the mutually opposite direction. Objects bounced off of the four invisible walls of the display, and moved at a constant rate of $9.94^\circ/\text{sec}$ in the large-display condition and $2.49^\circ/\text{sec}$ in the small-display condition. Demonstration videos are available at www.psych.northwestern.edu/~franconeri/lab/projects/.

Procedure. Participants were instructed to maintain fixation on a dot at the screen center. All 16 black discs would appear, distributed through the display, for 1 sec. Then, 4, 5, 6, 7, or 8 of the discs turned blue with yellow centers for 2.5 sec, in order to indicate their status as targets. All objects changed back to black, and then moved around the screen for the next 6 sec. When all of the objects had stopped moving, one object turned blue. The participant was told to press the "T" key if the blue object was a target, or the "D" key if it was a distractor. The probed object was either a randomly chosen target or a target's closest distractor. After responding to the probe, participants received auditory feedback and were then moved on to the next trial automatically.

Each participant received 60 practice trials, using four to eight targets on both display sizes, and then completed a total of 480 trials over approximately 90 min. Each randomly ordered block consisted of 6 trials of each of the five set sizes. For half of the participants, the first eight blocks were shown on the small display and the second eight blocks were on the large display, and the other half received the opposite ordering.

Results and Discussion

Mean tracking accuracy for the two display sizes is shown in Figure 2. These data show the expected large reduction in accuracy with an increasing number of targets

($M = 91\%$ for four targets, $M = 71\%$ for eight targets), but they do not indicate any differences between the two display sizes ($M = 82\%$ for small, $M = 81\%$ for large). A 2×5 repeated measures ANOVA revealed a significant main effect of number of targets [$F(4,36) = 33, p < .001; \eta_p^2 = .84$], with no effect of display size ($F < 1$) and no interaction between the two factors ($F < 1$).

Across the small- and large-display conditions, tracking accuracy was equal at several target set sizes. Because crowding strongly impairs object tracking, the equivalent performance in the two conditions is consistent with the idea that the influence of crowding was similar in the two display-size conditions. In contrast, object speed, as measured by speed across the display, or speed across the retina, was four times faster on the large display, and yet did not impair performance. At first glance, this result suggests that speed per se does not affect tracking performance. However, because we did not systematically manipulate speed in Experiment 1, perhaps object speed was not high enough to affect performance in either display condition. In Experiment 2, we directly manipulated object speed, and again asked participants to track objects on both large and small displays.

EXPERIMENT 2 Speed Limits in Tracking in Large and Small Fields

Method

Participants. Twelve undergraduate students participated in exchange for course credit. One participant, who could not track a single object at the lowest speed with at least 80% accuracy in both display conditions, was removed from the analysis.

Stimuli. Displays were identical to those used in Experiment 1, with two exceptions. Displays did not extend as far into the periphery (the large display was 82° wide \times 36.4° high [700×310 pixels], and the small display was 20.5° wide \times 9.1° high [175×78 pixels]). Although both displays were slightly smaller, there were also fewer objects (12), leaving object density similar to that in Experiment 1. The average minimum distance between an object and its nearest

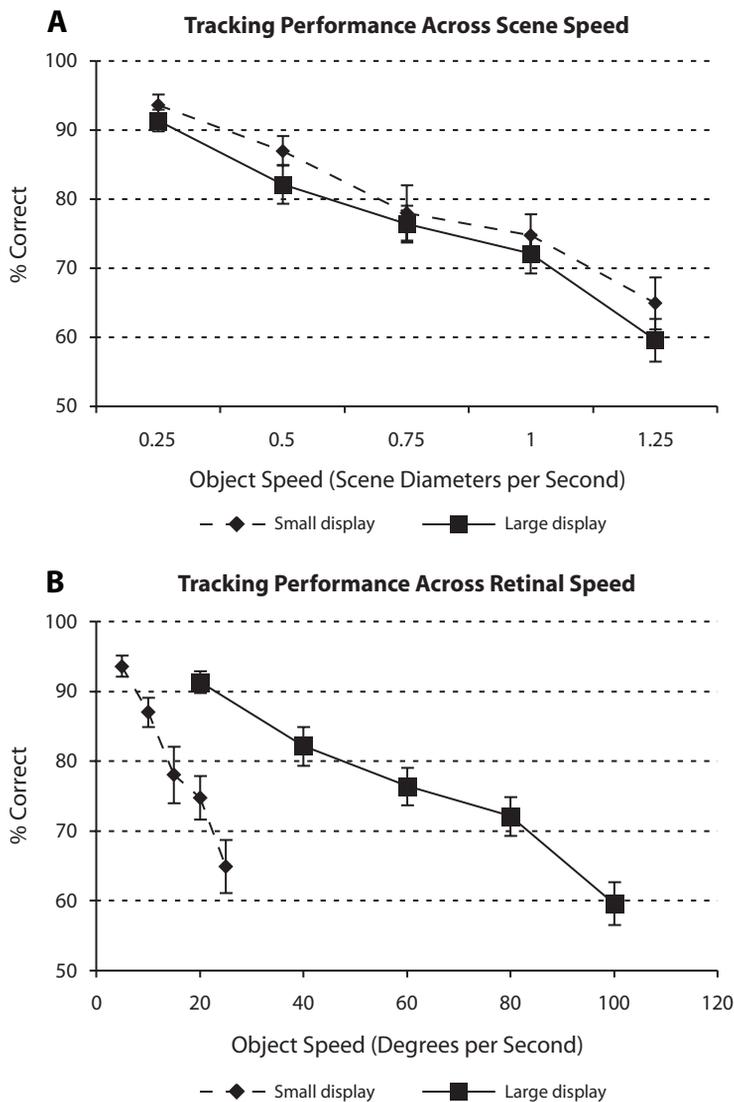


Figure 3. Mean accuracy in Experiment 2. (A) Object speed relative to the scale of the scene (scene widths per second). (B) Object speed relative to the retina (degrees of visual angle per second). Error bars represent standard errors.

neighbor was 11.3° in the large-display condition, similar to 11.7° in the large-display condition of Experiment 1.

Procedure. Displays always contained 4 target objects among 12 total objects. The relative speed of the objects was varied between trials, at 0.24, 0.49, 0.73, 0.97, or 1.22 scene widths per second. The actual retinal speeds for the large-display condition were 19.9° , 39.9° , 59.8° , 79.7° , or $99.6^\circ/\text{sec}$, and 5.0° , 10.0° , 14.9° , 19.9° , or $24.9^\circ/\text{sec}$ in the small-display condition. Trial sequence and block design were identical to those in Experiment 1, except that speed, instead of the number of targets, was varied.

Results and Discussion

Mean tracking accuracy is shown in Figure 3, plotted both as a function of speed relative to the scale of the scene (Figure 3A) and as a function of speed on the retina (Figure 3B). The data show the expected reduction in tracking accuracy with an increase in speed, but, most remarkably, there is little difference in accuracy between the small-

and large-display conditions when speed is expressed in scene-relative speed. In contrast, when object speed is expressed in terms of displacement on the retina, higher performance is observed in the large-display condition for the same object speed. For example, when an object moves at $20^\circ/\text{sec}$ in the small-display condition, tracking accuracy is at only 75% (see Figure 3B). But when objects move at $20^\circ/\text{sec}$ in the large-display condition, accuracy increases to 91% [$t(10) = 6.4$, $d = 2.03$]. Note also that when display size is changed by a factor of 4 from the small to the large display, retinal speed is increased by the same factor, to $80^\circ/\text{sec}$, yet accuracy only drops slightly, from 75% to 72% ($t < 1$).

These observations for scene-relative speed were confirmed by a 2×5 ANOVA showing that accuracy decreased from the slowest object speed ($M = 93\%$) to the fastest speed ($M = 62\%$) [$F(4,40) = 62$, $p < .001$; $\eta_p^2 =$

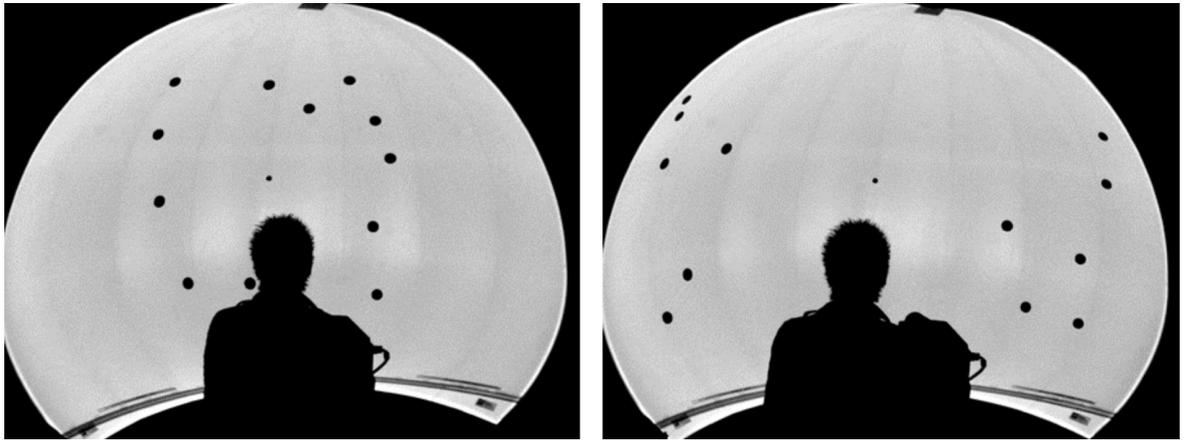


Figure 4. Photographs of typical central (left) and peripheral (right) display in Experiment 3.

.86]. Accuracy was also slightly higher for the small-display ($M = 80\%$) than for the large-display ($M = 76\%$) condition, and this difference did not reach traditional statistical significance [$F(1,10) = 3.7, p = .085; \eta_p^2 = .27$]. There was no interaction between display size and object speed ($F < 0.4$).

EXPERIMENT 3 Object Crowding Limits

Past studies have shown that crowding limits tracking on small displays. In Experiment 3, we showed that crowding also limits performance in our large-display condition. We used more densely spaced objects, and constricted them to either a central area, closer to the fovea, or a more peripheral area. Because attentional resolution is lower in the periphery (Intriligator & Cavanagh, 2001), we expected lower tracking accuracy in the peripheral condition.

Method

Participants. Fifteen undergraduate students participated in exchange for course credit. The data from 3 participants were not included, because they were unable to track in the periphery with at least 80% accuracy in both of the lowest load conditions (two targets).

Stimuli. Stimuli were identical to those in the large-display condition of Experiment 1, with the following exceptions. In each trial, the tracking task was presented within two invisible, independent rectangles. The two rectangles, each 21.7° wide \times 42.2° high (185×360 pixels), either flanked the fixation point, with no separation between them (central display), or were pushed outward into the periphery (peripheral display) by 20° (170 pixels) each. In both conditions, half of the targets and half of the distractors were constrained to the left rectangle and the other half of the right rectangle. See Figure 4 for photographs of the two display types. Because the total tracking area was reduced by 50%, relative to Experiment 1, the total number of objects was reduced to 12, to prevent overcrowding. Object density, defined by the average distance between an object and its nearest neighbor, was higher (74 pixels) than that in the large-display condition of Experiment 1 (100 pixels).

Procedure. Participants tracked 1, 2, or 3 targets in each display rectangle simultaneously, for a total tracking load of 2, 4, or 6 targets out of 12. There were 80 trials of each combination of condition and tracking load, for a total of 480 trials, presented in a random order.

Results and Discussion

Mean tracking accuracy is shown in Figure 5. These data show the expected reduction in accuracy with an increased number of targets, but they also show that this reduction is greater for peripherally than for centrally presented displays. A 2×3 repeated measures ANOVA revealed a significant main effect for number of targets [$F(2, 22) = 21, p < .001; \eta_p^2 = .65$] and generally higher accuracy for central ($M = 95\%$) than for peripheral ($M = 83\%$) displays [$F(1, 11) = 50, p < .001; \eta_p^2 = .82$]. There was also an interaction between the factors [$F(2, 22) = 10, p = .001; \eta_p^2 = .48$]; however, we caution that this interaction could be due to the near-ceiling performance in the low target load central display condition. These results confirm that tracking performance is limited by crowding in our large-display conditions.

GENERAL DISCUSSION

Multiple-object tracking is limited by the number of target objects, the speed at which the objects move, and interobject crowding. Is there a unitary explanation for

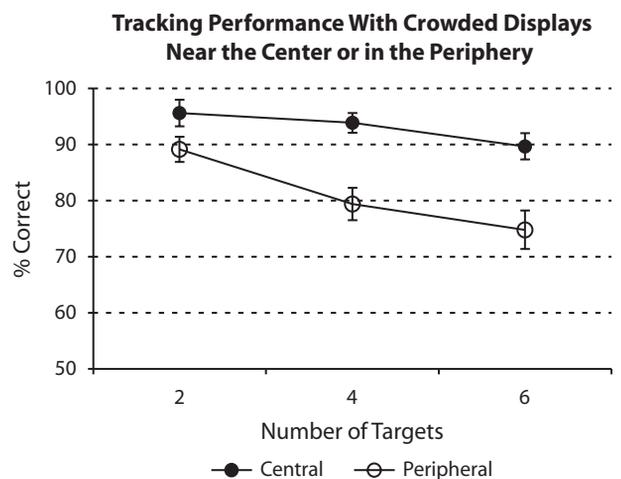


Figure 5. Mean accuracy in Experiment 3. Error bars represent standard errors.

all three of these limits? Past evidence suggests that the limit on the number of tracked objects comes from crowding (Intriligator & Cavanagh, 2001) or interobject inhibition (Franconeri, Alvarez, & Enns, 2007; Shim et al., 2008). Here, we argue that crowding can also explain the limit on object speed. In Experiment 1, we showed that increasing the scale of the tracking display did not, in itself, reduce the number of successfully tracked objects, even though it dramatically increased the object speeds. Experiment 2 showed identical performance over a range of display-relative speeds in both small- and large-display conditions, despite the fact that the retinal speeds were four times as fast in the large-display conditions. Experiment 3 confirmed that crowding still limited tracking in our large-display conditions.

In the most peripheral regions of our displays (where crowding should produce the greatest impairment), the average distance between an object and its nearest neighbor is similar to a past estimate of the minimum interobject spacing required to individuate a peripheral object. Extrapolating from estimates reported in Intriligator and Cavanagh (2001), we expected the minimum spacing to be roughly 10.4° (averaging across lower and upper visual fields) at the 40° eccentricity of our large display (or 2.6° at the 10° eccentricity of the small display). Our actual average nearest neighbor distances were similar: 11.7° for Experiment 1 and 11.3° for Experiment 2 for the large display, and 2.8° and 2.9° , respectively, for the small display. In the peripheral-display condition of Experiment 3, at 40° eccentricity, spacing was much tighter (8.7°) than in Experiments 1 and 2. But in the central-display condition, spacing (8.7°) for objects only 20° from fixation was, on average, comfortably larger than the estimated requirement of 5.2° . Note that these distances reflect averages, meaning that half of the time the nearest neighbor is actually closer. For example, in the large-display condition of Experiment 1, although the average distance was 11.7° , at the 10th closest percentile in the distribution, the distance was 8.6° , and at the 1st percentile, it was 6.1° . In Experiment 3, at the 10th percentile, the distance was 5.8° and at the 1st percentile, it was 3.9° .

Although we claim that making displays four times larger kept crowding roughly constant while increasing each object's retinal speed, an alternative view is that object speed is not defined in retinal coordinates. If the relevant measure were angular speed normalized to the scale of the tracking display, speed would be the same across our small- and large-display conditions. The literature on speed perception suggests that speed judgments can be based either on estimates of objective speed or on angular speed across the retina (McKee & Smallman, 1998). But neither of these ways of measuring speed yields equal measures of speed across our display conditions; both objective and angular speeds were faster in our large-display conditions, because the large display is objectively larger and subtends a larger visual angle.

Our crowding account suggests a reinterpretation of several past findings. First, one study reported tracking capacities of eight objects by undergraduate students (Alvarez & Franconeri, 2007), which is far above the typical four

or five found in past studies using a similar population. The higher limit in that study likely stems both from the large interobject distances used and from the slow speeds that kept interobject distances to a minimum. Second, Alvarez and Franconeri found that fewer targets could be tracked reliably over a 6-sec interval as their speed increased. But, according to the present account, this may have been caused by the increased probability that a target would pass within a critical distance of a distractor. In another study, tracking was impaired by increases in object speed within a 3-D scene, but not by the increased movement of the 3-D scene containing the objects (Liu et al., 2005). We also reinterpret those data as consistent with an increase in the number of crowding instances when object speed increased, but not when the scene itself moved.

Our account also predicts that if crowding could be eliminated, speed should no longer limit tracking accuracy. One way to accomplish this is to place the tracked targets in separate visual hemifields, where interobject crowding does not appear to strongly cross the hemifield boundary (Alvarez & Cavanagh, 2005; Carlson, Alvarez, & Cavanagh, 2007; see also Sereno & Kosslyn, 1991). A previous study showed precisely this result: Tracking capacity was independent when displays were separated into hemifields, but they dropped sharply when the display was in a single hemifield (Alvarez & Cavanagh, 2005).

The present findings, along with a reinterpretation of past findings, point to the possibility that speed per se may not limit multiple-object tracking. Instead, increased speed might limit performance solely because it increases the number of "close encounters" per unit time. This implies a bold conclusion that demands further empirical validation. Within the limits of interobject crowding, if you can track one object at a given speed, you can track an unlimited number of objects at that speed.

AUTHOR NOTE

S.L.F. was supported by an Isaac Walter Killam Postdoctoral Fellowship. J.T.E. was supported by a Discovery Grant from NSERC (Canada). We thank George Alvarez, Anne Hillstrom, Steve Luck, and Satoru Suzuki for helpful comments. Correspondence concerning this article should be addressed to S. L. Franconeri, Northwestern University, 2029 Sheridan Rd., Evanston, IL 60208 (e-mail: franconeri@northwestern.edu).

REFERENCES

- ALVAREZ, G. A., & CAVANAGH, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*, 637-643.
- 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.
- 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*, 13496-13500.
- CAVANAGH, P., & ALVAREZ, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*, 349-354.
- FRANCONERI, S. L., ALVAREZ, G. A., & ENNS, J. T. (2007). How many locations can you select? *Journal of Experimental Psychology: Human Perception & Performance*, *33*, 1003-1012.
- HOPF, J.-M., BOEHLER, C. N., LUCK, S. J., TSOTOS, 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.

- INTRILIGATOR, J., & CAVANAGH, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, **43**, 171-216.
- LIU, G., AUSTEN, E. L., BOOTH, K. S., FISHER, B. D., ARGUE, R., 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.
- MCKEE, S. P., & SMALLMAN, H. S. (1998). Size and speed constancy. In V. Walsh & J. J. Kulikowski (Eds.), *Perceptual constancies: Why things look as they do* (pp. 373-408). New York: Cambridge University Press.
- PELLI, D. G., PALOMARES, M., & MAJAJ, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, **4**, 1136-1169.
- 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**, 179-197.
- SCHOLL, B. J. (2001). Objects and attention: The state of the art. *Cognition*, **80**, 1-46.
- SERENO, A. B., & KOSSLYN, S. M. (1991). Discrimination within and between hemifields: A new constraint on theories of attention. *Neuropsychologia*, **29**, 659-675.
- 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**, 390-397.
- TOET, A., & LEVI, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, **32**, 1349-1357.

NOTE

1. Since errors typically occur when the identity of a target and a nontarget are switched (Pylyshyn, 2004), the number of target-nontarget pairs can modulate the effect of crowding. This may explain why, when the number of targets (T) increases while the total number of objects (A) remains constant, performance declines, even though crowding remains constant (see, e.g., Alvarez & Franconeri, 2007). Since the number of target-nontarget pairs in this case is $T(A - T)$ this number increases with increasing T when $T < \frac{1}{2}A$. Another suggestion for explaining this result is that as the number of targets increases, an inhibitory region may surround each attended target (Hopf et al., 2006), which could increasingly interfere with other tracked targets (Shim, Alvarez, & Jiang, 2008; see also Franconeri, Alvarez, & Enns, 2007, for a similar example using static displays).

(Manuscript received July 5, 2007;
revision accepted for publication February 6, 2008.)