

## Research Article

# The Representation of Simple Ensemble Visual Features Outside the Focus of Attention

George A. Alvarez and Aude Oliva

*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology*

---

**ABSTRACT**—*The representation of visual information inside the focus of attention is more precise than the representation of information outside the focus of attention. We found that the visual system can compensate for the cost of withdrawing attention by pooling noisy local features and computing summary statistics. The location of an individual object is a local feature, whereas the center of mass of several objects (centroid) is a summary feature representing the mean object location. Three experiments showed that withdrawing attention degraded the representation of individual positions more than the representation of the centroid. It appears that information outside the focus of attention can be represented at an abstract level that lacks local detail, but nevertheless carries a precise statistical summary of the scene. The term ensemble features refers to a broad class of statistical summary features that we propose collectively make up the representation of information outside the focus of attention.*

---

As people go about their daily lives, they seem to effortlessly manage the extremely rich and detailed stream of information entering their eyes. For the most part, people successfully navigate through busy intersections; find items of interest, such as food or friends; and understand complex social situations—all just by the simple act of looking. However, despite these many successes, there are also countless demonstrations that people fail to notice potentially important visual events, particularly when their attention is focused elsewhere. For example, traffic accidents often involve drivers “not seeing” clearly visible obstacles (e.g., McLay, Anderson, Sidaway, & Wilder, 1997). Such occurrences are typically interpreted as attentional

lapses: It seems that people have a severely limited ability to perceive, understand, and act upon information that falls outside the current focus of attention.

Even when attention is focused intensely on a particular object, however, people do not experience “blindness” for all other visual information. The purpose of the current study was to probe what type of representation can be maintained outside the focus of attention. In so doing, we emphasized the distinction between local visual features and statistical summary features. Local visual features are properties that describe an individual item, independently of other items. For example, the size and the location of an individual object are local visual features. In contrast, there are a variety of statistical summary features that represent information at a more abstract level, collapsing across local details (Ariely, 2001; Chong & Treisman, 2003, 2005b). For the present study, we focused on relatively simple summary features, such as the center of mass of a collection of objects (henceforth, the “centroid”), which is essentially the mean position of the group. Specifically, we tested the hypothesis that withdrawing attention impairs perception of local features more than it impairs the perception of summary features.

Object location was used as a test case for investigating whether summary features can be represented more robustly than local visual features outside the focus of attention. We used a multiple-object-tracking task (Pylyshyn & Storm, 1988) in which eight objects moved around the display. The primary task was to attentively track a subset of four target objects while ignoring four distractor objects. This attentionally demanding tracking task drew focal attention toward the targets and away from the distractors (Intriligator & Cavanagh, 2001; Sears & Pylyshyn, 2000). At a random time during each trial, all items disappeared briefly (200 ms), and then all but one or four randomly chosen targets or distractors reappeared. Participants had to localize either the single missing item (*individual test*) or the centroid of the missing group of items (*centroid test*). Because distractors receive less attention than targets, we predicted that

---

Address correspondence to George Alvarez, Department of Brain and Cognitive Sciences, 46-4078, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, e-mail: alvarez@mit.edu.

localizing missing distractors would be more difficult than localizing missing targets. However, of principal interest was whether the distractor centroid would be represented more robustly than the individual distractor positions, indicating a relative sparing of summary features outside the attentional focus.

Across three experiments, we varied the extent to which subjects could selectively attend to targets. In Experiment 1, targets and distractors were physically identical and moved among each other, making selective target processing most difficult. In Experiment 2, targets were white and distractors were black, and this separation of targets from distractors in feature space facilitated target selection. In Experiment 3, targets and distractors were again identical, but target selection was facilitated by spatially separating targets from distractors, such that the focus of attention was far removed from the distractors. We found that selective processing of the targets improved in Experiments 2 and 3 relative to Experiment 1, to the extent that in Experiments 2 and 3, participants performed at chance level when asked to localize a single missing distractor. In contrast, we found in all experiments that participants could accurately localize the distractor centroid, even when individual distractors were localized at chance levels. This finding suggests that the cost of withdrawing attention from distractors can be compensated for by pooling together noisy local signals and computing summary statistics.

## GENERAL METHOD

### Participants

Each experiment had a separate group of 8 participants who were between the ages of 18 and 35, gave informed consent, and were paid \$10/hr for their participation.

### Apparatus

The experiments were run using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The display was  $35^\circ \times 28^\circ$ , viewed from 57 cm.

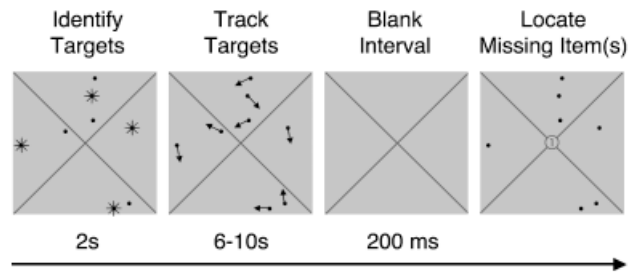
### Stimuli

The stimuli were eight circles (radius =  $0.35^\circ$ ) that moved at a constant rate of  $4^\circ/s$  within a central region of the screen, marked by a black, square outline ( $24.5^\circ \times 24.5^\circ$ , line thickness =  $0.1^\circ$ ; see Fig. 1). Two diagonal red lines connected the corners of the square (line thickness =  $0.1^\circ$ ), and the background was gray. The circles' direction of motion was constrained such that items appeared to avoid one another, while otherwise moving randomly about the display.

### Procedure

#### Test Phase

Participants performed a multiple-object-tracking task followed by a missing-item localization task (see Fig. 1). At the start of



**Fig. 1.** Illustration of the experimental procedure. At the beginning of each trial, eight items appeared; four of the items were identified as targets by flashing on and off. Then all items moved about the display, and participants counted the total number of times a target circle touched a red line. At an unpredictable time between 6 and 10 s from the onset of motion, there was a 200-ms blank interval, followed by a final test display consisting of stationary items. Either one item was missing (a randomly selected target or distractor) or four items were missing (all the targets or all the distractors). A single digit at the center of the display informed participants of how many items were missing. Participants indicated where the missing item or items were located, clicking directly on the location of a single missing item or on the centroid of four missing items.

each trial, eight circles appeared, and four of these items were identified as tracking targets by flashing off and on for 2 s. Next, all items moved for a random duration between 6 and 10 s. The primary task was to attentively track the targets, counting the number of times a target item touched or crossed one of the red lines. Participants kept one running count collapsed across all targets. This task was attentionally demanding and ensured that participants were continuously focusing their attention on the target items.

At the end of each trial, all the circles disappeared for 200 ms, and then some reappeared. In the *individual-test* condition, either a single target or a single distractor was missing from the final display. In the *centroid-test* condition, either all four targets or all four distractors were missing. A cue that appeared at the center of the screen informed participants of how many items were missing (“1” for individual tests, “4” for centroid tests). Participants used the mouse to move a crosshair and clicked either on the location of the single missing item or on the location corresponding to the centroid of the four missing items. After clicking on the selected position, participants typed in the number of times targets had touched the red lines during the trial. Although participants entered this number second, they were instructed that the counting task was their primary task, and that they should not sacrifice accuracy on the counting task to perform the localization task. There were 80 trials, with conditions randomly intermixed.

#### Guessing Phase

After the test phase, participants completed a guessing phase, in which they were not required to track any targets. They were simply shown test displays with either one or four items missing, and were asked to guess where they thought the missing item or items were located. These displays were generated in the same

way as in the test phase, but participants were shown only the final frame. The data from this phase provided an estimate of how well participants could guess where the missing items were located on the basis of the configuration of the items present in the test display, providing an empirical estimate of chance performance.

### EXPERIMENT 1: WITHDRAWING ATTENTION WITH MULTIPLE-OBJECT TRACKING

Experiment 1 tested how well participants could judge the location of individual items or the centroid of a group outside the focus of attention. We hypothesized that the tracking task would draw attention away from distractors, resulting in better localization accuracy for targets than distractors, but that ability to localize distractors would be better on centroid tests than on individual tests.

#### Method

In Experiment 1, the targets and distractors were all black, so that target selection was difficult.

#### Results and Discussion

Overall participants accurately performed the primary counting task, typically missing one or two touches. Counting errors did not vary systematically with the type of localization test (individual vs. centroid),  $F(1, 7) = 2.06, p = .194, \eta_p^2 = .23$ , or with the type of item that was missing (target vs. distractor),  $F < 1$ . Error in the guessing phase, which provides an empirical estimate of chance performance, averaged  $10.0^\circ$  ( $SEM = 0.5^\circ$ ) for individual tests and  $5.4^\circ$  ( $SEM = 0.1^\circ$ ) for centroid tests (see the dashed lines in Fig. 2).

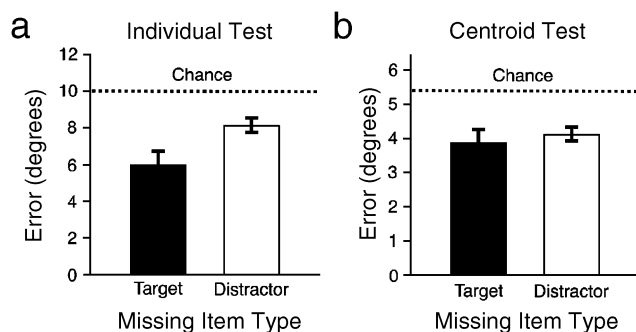
Localization accuracy revealed an important difference between individual tests and centroid tests. As Figure 2a illustrates, error in reporting the location of a single missing item was significantly lower for targets than for distractors,  $t(7) = 3.69, p = .008, r^2 = .66$ , and performance was better than chance for

both targets,  $t(7) = 4.04, p = .005, r^2 = .70$ , and distractors,  $t(7) = 3.20, p = .015, r^2 = .59$ . Although performance was better for targets than for distractors, which suggests that attention was more focused on the targets, the ability to identify distractor locations better than chance suggests that target selection was imperfect, and that some attention may have been paid to distractors. However, the amount of attention was not enough to localize individual distractors as accurately as individual targets.

In contrast, there was no significant difference in localization error for targets and distractors in the centroid-test condition,  $t(7) = 1.07, p = .321, r^2 = .14$  (see Fig. 2b). Thus, despite noisy individual representations of distractor locations—as evidenced by error on the individual tests—participants could determine the location of the distractor centroid as well as the location of the target centroid. Performance was again better than chance for both targets,  $t(7) = 4.08, p = .005, r^2 = .70$ , and distractors,  $t(7) = 5.48, p = .001, r^2 = .81$ .

Could this remarkable level of accuracy in localizing distractor centroids be achieved by sampling just one or two distractors and making an informed guess about where the centroid would be located? To assess this possibility, we ran Monte Carlo simulations to determine how well participants could judge the location of the distractor centroid by pooling one, two, three, or four noisy individual samples. We assumed that estimates of individual item positions are noisy and estimated the amount of noise from performance on the individual tests. The distribution of errors was approximately normal, and so the model assumed that each individual item position was represented with independent, normally distributed noise with a standard deviation estimated separately for each observer. For example, if each distractor's position could be estimated within  $4^\circ \pm 1.5^\circ$  on average for a given observer, we could simulate how accurately the centroid of all four distractors could be determined if one, two, three, or four estimates were averaged. The results of this simulation suggested that, given how noisy the individual estimates appeared to be, participants would have had to pool signals from all the distractors to achieve the level of accuracy we observed. Guessing any four random positions on the screen and pooling those estimates would not yield this level of performance; only if these very noisy individual estimates were centered on the actual distractor positions would pooling them enable participants to localize the centroids as accurately as they did.<sup>1</sup>

It appears that observers can make accurate judgments about distractors as a group by pooling information from all the individual distractors, even when the individual details of the distractors are not represented accurately. This suggests that observers maintain some awareness of the summary features of items appearing outside the focus of attention, even when local



**Fig. 2.** Results of Experiment 1: participants' error (in degrees) in localizing (a) missing individual targets and distractors and (b) the centroid of groups of missing targets and distractors. Error bars represent standard errors of the means.

<sup>1</sup>A supplementary appendix providing the details of all simulations reported in this article can be obtained directly from the first author or downloaded online at <http://cvcl.mit.edu/george/Publications.htm>.

information is represented inaccurately. However, because the targets and distractors were physically identical in Experiment 1, it is possible that participants could not completely filter out the distractor items, or that distractors were occasionally confused for targets, and therefore that distractors received enough attention to improve localization of the distractor centroid. Indeed, participants could localize distractors at better than chance levels, which suggests that some attention “spilled over” to the distractors. Experiments 2 and 3 tested how improving selective processing of targets affects the localization of individuals and their centroid.

## EXPERIMENT 2: INCREASED TARGET SELECTION USING A SALIENT FEATURE DIFFERENCE

Previous work has demonstrated that what people see depends on how they tune their attention—their attentional set—such that irrelevant information is more likely to be noticed or processed if it matches the physical properties of attended items (Most, Scholl, Clifford, & Simons, 2005; Most et al., 2001). However, there appears to be little to no perception of information that falls outside the attentional set. For instance, the appearance of an irrelevant black item will go undetected when participants attend to white items, especially if the distractors are black (Most et al., 2001). In Experiment 2, we increased the degree to which targets could be selectively attended by making the targets white and the distractors black. Continuous tracking of the targets was still required, because the primary task was to count the number of times a target touched a red line in the display. Of principal interest was whether participants would still be capable of accurately judging the location of the distractor centroid, even when the distractors fall outside the attentional set.

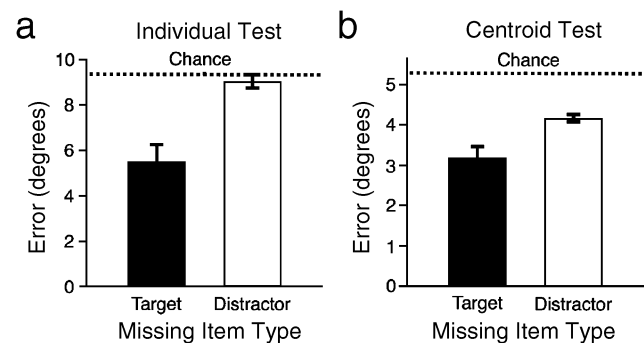
### Method

All aspects of the stimuli and procedure were the same as in Experiment 1, except that the targets were white instead of black.

### Results and Discussion

Counting errors did not vary with the type of localization test (individual vs. centroid),  $F < 1$ , or with the type of item that was missing (target vs. distractor),  $F < 1$ . Error in the guessing phase averaged  $9.4^\circ$  ( $SEM = 0.1^\circ$ ) for individual tests and  $5.3^\circ$  ( $SEM = 0.1^\circ$ ) for centroid tests (see the dashed lines in Fig. 3).

As in Experiment 1, localization accuracy showed different patterns in the individual-test and centroid-test conditions. As Figure 3a illustrates, error in reporting the location of a single missing item was significantly lower for targets than for distractors,  $t(7) = 4.44$ ,  $p = .003$ ,  $r^2 = .74$ . Error was below the empirical chance estimate for targets,  $t(7) = 5.18$ ,  $p = .001$ ,  $r^2 = .79$ , but, as expected, was not better than chance for distractors,  $t(7) = 1.00$ ,  $p = .350$ ,  $r^2 = .13$ . This indicates that the salient



**Fig. 3.** Results of Experiment 2: participants’ error (in degrees) in localizing (a) missing individual targets and distractors and (b) the centroid of groups of missing targets and distractors. Error bars represent standard errors of the means.

feature difference was effective in enhancing selective target processing.

As Figure 3b illustrates, in the centroid condition, there was a significant difference in localization error for targets and distractors,  $t(7) = 3.37$ ,  $p = .012$ ,  $r^2 = .62$ , but most important, performance was better than chance for both targets,  $t(7) = 8.32$ ,  $p < .001$ ,  $r^2 = .91$ , and distractors,  $t(7) = 7.47$ ,  $p < .001$ ,  $r^2 = .89$ . Thus, even though the individual position of any single distractor was so poorly represented that performance was at chance for individual judgments, participants could determine the location of the distractor centroid well above chance level.

Simulation results again suggested that participants would have had to pool signals from all the distractors to achieve the level of accuracy we observed. Thus, it appears that an accurate representation of the distractor centroid can be attained by pooling noisy local signals, even when target selection is facilitated by a salient feature difference between targets and distractors.

Previous work suggests that attention is focally allocated to targets in a multiple-object-tracking task and does not spread over the space between targets (Intriligator & Cavanagh, 2001; Sears & Pylyshyn, 2000), as if there were multiple, independent foci of attention (Cavanagh & Alvarez, 2005). Nevertheless, the targets in Experiments 1 and 2 were often distributed across the display such that the “virtual polygon,” or convex hull, formed by the targets encompassed several of the distractors. It is possible that distractors frequently received diffuse attention in these displays, and that this attention was insufficient for the local computations necessary to accurately judge the individual distractor locations, but sufficient for accurate judgments about the centroid of the distractors (Chong & Treisman, 2005a). In Experiment 3, we investigated this possibility.

## EXPERIMENT 3: WITHDRAWING ATTENTION FROM THE DISTRACTOR REGION

In Experiment 3, we attempted to withdraw participants’ attention from the region of space containing distractors by con-

straining the targets and distractors to move in opposite halves of the display. Targets were randomly constrained to move within the top, bottom, left, or right half of the screen, and distractors were constrained to move within the opposite half of the screen. As in the previous experiments, continuous tracking of the targets was required, because the primary task was to count the number of times targets touched a red line in the display. However, the spatial separation between the targets and distractors ensured that distractors never fell within the convex hull formed by the targets. Of principal interest was whether this manipulation would eliminate participants' ability to make accurate judgments about the distractor centroid.

## Method

All aspects of the stimuli and procedure were the same as in Experiment 1, except that targets remained spatially separate from the distractors for the entirety of each trial.

## Results and Discussion

Counting errors did not vary systematically with the type of localization test (individual vs. centroid),  $F(1, 7) = 2.53$ ,  $p = .156$ ,  $\eta_p^2 = .27$ , or with the type of item that was missing (target vs. distractor),  $F < 1$ . Error in the guessing phase was on average  $7.9^\circ$  ( $SEM = 0.5^\circ$ ) for individual tests and  $4.5^\circ$  ( $SEM = 0.3^\circ$ ) for centroid tests (see the dashed lines in Fig. 4).

As Figure 4a illustrates, error in reporting the location of a single missing item was significantly lower for targets than for distractors,  $t(7) = 3.28$ ,  $p = .013$ ,  $r^2 = .61$ . Error was below the empirical chance estimate for targets,  $t(7) = 4.26$ ,  $p = .004$ ,  $r^2 = .72$ , but not for distractors,  $t < 1$ . As Figure 4b illustrates, localization error in the centroid condition was significantly lower for targets than for distractors,  $t(7) = 3.17$ ,  $p = .016$ ,  $r^2 = .59$ , but centroid localization was better than chance for both targets,  $t(7) = 8.32$ ,  $p < .001$ ,  $r^2 = .88$ , and distractors,  $t(7) = 2.76$ ,  $p = .028$ ,  $r^2 = .52$ . Thus, as in Experiment 2, the individual position of any single distractor was so poorly represented

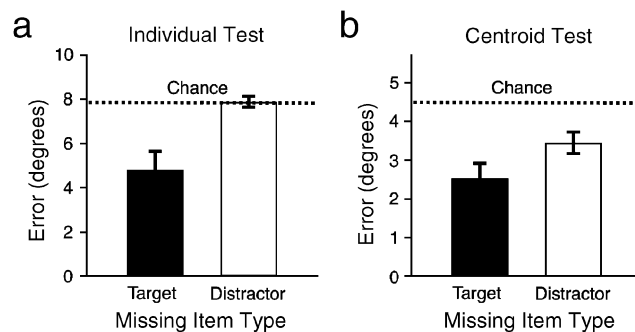
that performance was at chance for individual judgments, yet participants could localize the distractor centroid well above chance level.

Spatially separating the targets from the distractors ensured that distractors never fell within the convex hull formed by the targets, and therefore prevented distractors from receiving continuous diffuse attention. Nevertheless, the results again showed that participants could accurately judge the location of the centroid of distractors with high accuracy, with performance again well better than chance. Simulations suggested that this level of accuracy could have been attained only if (a) noisy individual estimates from all the distractors were pooled together, and (b) the noisy individual estimates were centered around the actual positions of the distractors, and were not truly random guesses. Thus, the distractor centroid can be represented accurately even when targets and distractors are spatially separated, such that target selection is facilitated and distractors are kept far from the focus of attention.

## GENERAL DISCUSSION

Visual information can be represented at multiple levels of abstraction, from local details to abstract features that summarize the local details. We used object location as a test case to explore the representation of local versus summary visual features outside the focus of attention. The location of an individual object is a local detail, whereas the centroid of a collection of objects is a simple summary feature that represents the objects as a group. In an adapted multiple-object-tracking task, participants were required to attentively track a set of moving targets while ignoring a set of moving distractors. During the tracking task, all of the items disappeared briefly (200 ms), and then all but one or four randomly chosen items reappeared; the secondary task was to localize the missing items. The results suggested that although participants knew very little about the local details of individual distractors, they could accurately report the centroid of the distractors.

These findings are related to previous findings concerning inattention blindness and change blindness. Inattention blindness studies have shown that without attention, there is little or no consciously accessible representation of a scene (Mack & Rock, 1998; Neisser & Becklen, 1975). These studies typically aim for participants to completely withdraw attention from the tested items, and sometimes even actively inhibit information outside of the attentional set (Most et al., 2001). In contrast, observers in our task were attempting to monitor all information, but the primary task required them to focus attention on a particular subset of that information. Thus, we assume that our observers were aware of, and paying some attention to, all information in the display. For this reason, our study is more related to change-blindness studies, in which displays consist of two alternating scenes that differ in one aspect (e.g., a single item changes color). In such studies, ob-



**Fig. 4.** Results of Experiment 3: participants' error (in degrees) in localizing (a) missing individual targets and distractors and (b) the centroid of groups of missing targets and distractors. Error bars represent standard errors of the means.

servers often fail to notice substantial differences between the scenes, and this finding has been interpreted as reflecting a failure to represent information outside the focus of attention (Rensink, O'Regan, & Clark, 1997). However, these studies have typically manipulated local features, such as the color or orientation of an individual object. Less is known about how well summary visual features are represented outside the focus of attention in the change-detection paradigm. The current results suggest that changes to local feature information outside the focus of attention are unlikely to be noticed, unless the changes alter the summary statistics of the scene.

Previous work suggests that processing of summary statistics is improved when observers spread their attention diffusely (Chong & Treisman, 2005a), and Treisman (2006) has argued that summary statistics are computed automatically when attention is spread diffusely. The current results suggest that even when attention is focally allocated to a subset of items, summary features can be computed outside the focus of attention. Most important, the representation of these summary features is more robust to the withdrawal of attention than is the representation of local visual features. Surprisingly, our experiments show that even when local features are so poorly represented that they are identified at chance levels, it is possible to pool estimates of those local details and attain an accurate representation of the group outside the focus of attention. However, whether the centroid position of distractors is computed only when required by the task demands, or whether it is computed automatically, remains an open question to be addressed by future research.

Although object location was particularly well suited for an initial investigation into the representation of summary features outside the focus of attention, it is important to recognize that there are many other types of summary features. In general, other researchers have referred to these features as global features (Navon, 1977; Oliva & Torralba, 2001), holistic features (Kimchi, 1992), or sets (Ariely, 2001; Chong & Treisman, 2003, 2005b). We refer to these types of features under the umbrella term *ensemble visual features*. We use the term *ensemble* because other terms carry certain connotations that do not accurately represent our view of what counts as a summary statistic. Specifically, the terms *global* and *holistic* are often used interchangeably with *low spatial frequency*, and the term *set* is often used to refer to collections of discrete objects. But *ensemble* refers to any summary statistic that collapses across individual image details, whether or not those details are contained within a specific spatial-frequency band, and whether those details are attached to discrete objects, parts, or a location in space. Moreover, an ensemble can include relatively simple features, such as the mean size or the centroid of a collection of objects, or more complex features, such as particular combinations of local orientation and spatial-frequency information (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Torralba, Oliva, Castelano, & Henderson, 2006).

## CONCLUSION

This study shows that information outside the focus of attention remains consciously accessible in the form of an ensemble representation that lacks local detail, but nevertheless carries a precise statistical summary of the visual scene. Future work will be necessary to determine which classes of ensemble features are represented robustly outside the focus of attention. Of particular interest are ensemble features that capture the statistics of the natural world and are likely to play a vital role in everyday perception.

**Acknowledgments**—For helpful conversation and comments on earlier drafts, we thank Tim Brady, Talia Konkle, Ruth Rosenholtz, Antonio Torralba, and two anonymous reviewers. G.A.A. was supported by the National Institutes of Health, National Eye Institute (Fellowship F32EY016982). A.O. was supported by the National Science Foundation (CAREER Award 0546262).

## REFERENCES

- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*, 157–162.
- Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.
- Cavanagh, P., & Alvarez, G.A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences, 9*, 349–354.
- Chong, S.C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*, 393–404.
- Chong, S.C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics, 67*, 1–13.
- Chong, S.C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45*, 891–900.
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*, 171–216.
- Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin, 112*, 24–38.
- Mack, A., & Rock, I. (1998). *Inattention blindness*. Cambridge, MA: MIT Press.
- McLay, R.W., Anderson, D.J., Sidaway, B., & Wilder, D.G. (1997). Motorcycle accident reconstruction under Daubert. *Journal of the National Academy of Forensic Engineering, 14*, 1–18.
- Most, S.B., Scholl, B.J., Clifford, E.R., & Simons, D.J. (2005). What you see is what you set: Sustained inattention blindness and the capture of awareness. *Psychological Review, 112*, 217–242.
- Most, S.B., Simons, D.J., Scholl, B.J., Jimenez, R., Clifford, E., & Chabris, C.F. (2001). How not to be seen: The contribution of similarity and selective ignoring to sustained inattention blindness. *Psychological Science, 12*, 9–17.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353–383.
- Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology, 7*, 480–494.

- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal in Computer Vision*, *42*, 145–175.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J.A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), 739–744.
- Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Pylyshyn, Z.W., & Storm, R.W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.
- Rensink, R.A., O'Regan, J.K., & Clark, J.J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.
- Sears, C.R., & Pylyshyn, Z.W. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology*, *54*, 1–14.
- Torralba, A., Oliva, A., Castelano, M.S., & Henderson, J.M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*, 766–786.
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*, 411–443.

(RECEIVED 5/25/07; REVISION ACCEPTED 9/29/07)

# Supplementary Appendix

---

Alvarez, G. A., and Oliva, A. (2008). The representation of simple ensemble features outside the focus of attention. *Psychological Science*, 19(4), 392-398.

## APPENDIX

Simulations were conducted to determine whether the observed error in localizing the target centroid or the distractor centroid could be achieved by sampling a subset of the individual locations, and making an informed guess about the centroid position. The most important assumption of this model is that estimates of individual item positions are noisy, and that this noise can be estimated from participants performance on individual item tests. In each experiment, participants were required to localize a single missing target or distractor on some trials. The distribution of errors was well approximated by a normal distribution, and the model assumes that each individual item position was represented with independent, normally distributed noise of this magnitude and standard deviation.

### *Simulation Methods*

For a particular participant, individual estimates have an average error (meanErr), and a standard deviation of errors (stdErr). The accuracy of guessing the mean by pooling different numbers of individual estimates was simulated as follows.

1. A random set of 8 xy-coordinates were generated (4 targets and 4 distractors). These coordinates are referred to here as actualX and actualY. The actualCentroidX and actualCentroidY values were calculated by taking the mean of the actualX and actualY values, respectively.

2. Noisy estimates of these individual locations were generated as follows (modeling the participants' internal representations of the individual items).

```
errAngle = rand(0,359)
errMagnitude = meanErr + randn * (stdErr)
predictedX = actualX + magnitude*cos(angle/180*pi)
predictedY = actualY + magnitude*sin(angle/180*pi)
```

Where errAngle is a random integer between 0 and 359, corresponding to the direction in which the error occurs relative to the actual xy position, errMagnitude is the size of the error (normally distributed with an average of

meanErr and a standard deviation of stdErr), predictedX is the predicted x position, and predictedY is the predicted y position.

3. Predicted centroid locations were generated by averaging 1, 2, 3, or 4 of these noisy local estimates.

```
predictedCentroidX=average(predictedX1,predictedX2...)
predictedCentroidY=average(predictedY1,predictedY2...)
```

4. The error in these centroid estimates was calculated

```
errX = predictedCentroidX-actualCentroidX
errY = predictedCentroidY-actualCentroidY
predErrMagnitude = sqrt(errX*errX + errY*errY)
```

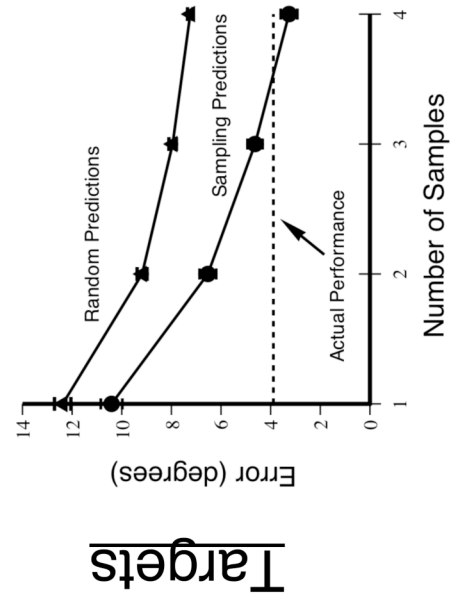
Where errX is the error in the X position, errY is the error in the Y position, and predErrMagnitude corresponds to the magnitude of the prediction error.

### *Simulation Results*

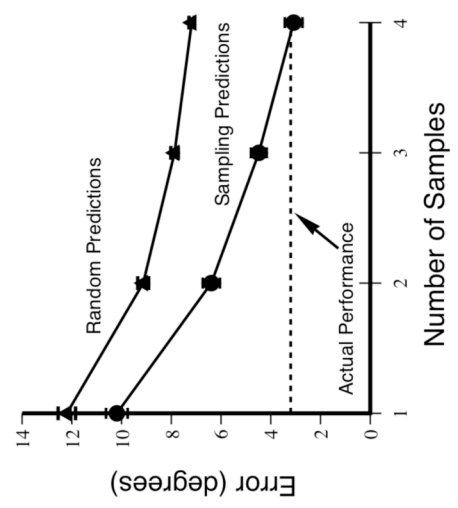
Figure A1 shows the simulation results for each experiment. The y-axis shows the predicted error in degrees, and the x-axis shows the number of individual positions sampled to predict the centroid position. For reference, the dashed lines show actual error in centroid localization averaged across participants. Each panel shows two prediction functions: circles show predictions calculated from noisy estimates centered around actual item positions, as described above, whereas triangles show predictions calculated from noisy estimates centered around completely random locations. This latter curve was generated to demonstrate that it is not possible to localize the centroid as accurately as participants did just by randomly guessing. The top panels show the results of the model simulation for target items. As the number of individual targets sampled increases, error in predicting the centroid decreases. However, only when sampling all 4 targets does predicted performance reach the actual performance level observed in each experiment. This suggests that information from all 4 targets is required to estimate the centroid as accurately as participants actually did, given the noise observed in estimates of the individual locations. The results were qualitatively identical for target simulations and distractor simulations.



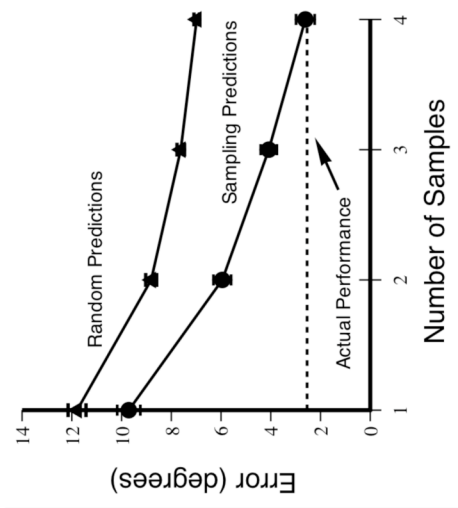
## Experiment 1



## Experiment 2



## Experiment 3



## Distractors

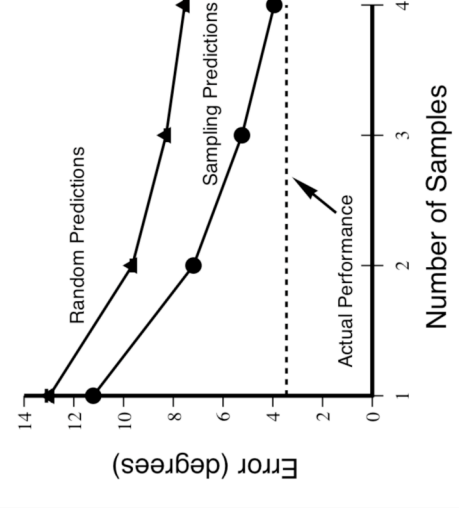
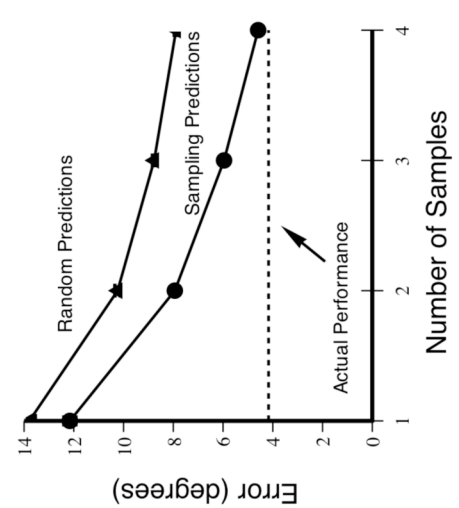
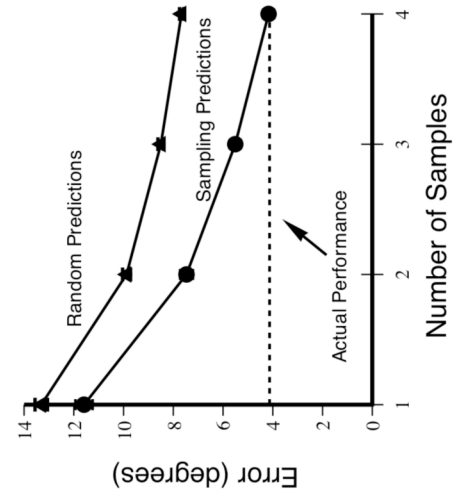


Figure A1. Predicted centroid localization error in each Experiment. The y-axis shows the predicted error in degrees, and the x-axis shows the number of individual positions sampled to predict the centroid position. Error bars are presented where greater than the data symbols and represent 1 s.e.m. For reference, the dashed lines shows actual performance averaged across participants. Circles show predictions based on noisy samples of individual items (sampling predictions). Only when sampling all 4 items does predicted performance reach the actual performance level (dashed line) for each simulation shown. Triangles show predictions based on noisy samples centered around random locations (random predictions), which never reach the actual performance level. This suggests that information from all 4 targets is required to estimate the centroid as accurately as observed. The results were similar for target and distractor simulations.