

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

The representation of visual information inside the focus of attention is precise relative to the representation of information outside the focus of attention. We find 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, while 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 which 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, which we propose collectively comprise 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 we fail to notice potentially important visual events, particularly when our attention is focused elsewhere. For example, traffic accident reports often include accounts of drivers “not seeing” clearly visible obstacles (e.g., McLay, Anderson, Sidaway, & Wilder, 1997). Such occurrences are typically interpreted as attentional lapses: it seems that we have a severely limited ability to perceive, understand, and act upon information that falls outside the current focus of attention.

Importantly, even when attention is focused intensely on a particular object, we 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, with an emphasis on the distinction between local visual features, and statistical summary features. Local visual features are properties that describe an individual item, independent of other items. For ex-

ample, the size or the location of an individual object would be a local visual feature. 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; Chong & Treisman, 2005b). For the present study, we will focus 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 will impair perception of local features more than it will impair the perception of summary features.

Object location was used as a case study for investigating whether summary features can be represented more robustly than local visual features outside the focus of attention. In a multiple object-tracking task (Pylyshyn & Storm, 1988), 8 objects moved around the display and the primary task was to attentively track a subset of 4 target objects while ignoring 4 distractor objects. The attentionally demanding tracking task drew focal attention to the targets, and away from the distractors (Intriligator & Cavanagh, 2001; Sears & Pylyshyn, 2000). At a random time during each trial, some targets or distractors would randomly disappear, and participants had to localize either a single missing item (individual test), or the centroid of a missing group of items (centroid test). Because distractors receive less attention than targets, we predicted that 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 experiments, we varied the extent to which subjects could selectively attend to targets. In Experiment 1, targets and distractors were physically identical and moved amongst each other, making selective target processing most

For helpful conversation and/or comments on earlier drafts, we thank Tim Brady, Talia Konkle, Ruth Rosenholtz, Antonio Torralba and two anonymous reviewers. GAA was supported by NIH/NEI fellowship #F32EY016982. AO was supported by a NSF CAREER award #0546262. Address correspondence and reprint requests to: George A. Alvarez, Department of Brain and Cognitive Sciences, 46-4078, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139; Email: alvarez@mit.edu; Web: <http://cvcl.mit.edu/george/home.htm>; Phone: 617-258-9675; Fax: 617-258-8654.

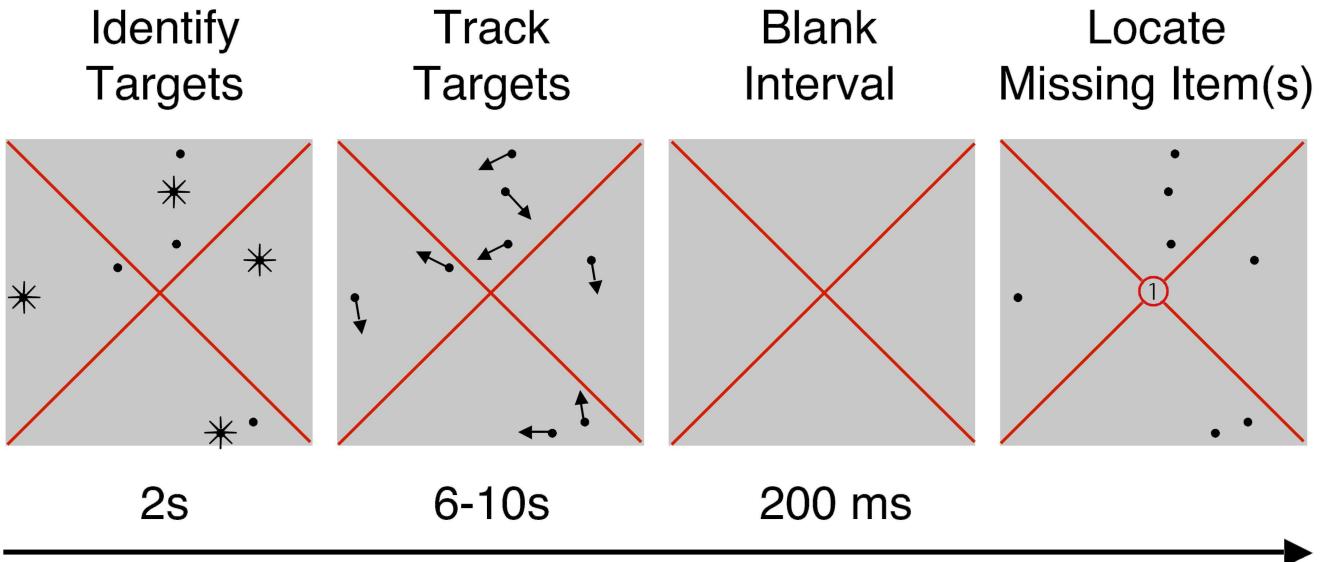


Figure 1. Multiple object tracking plus missing item localization. At the beginning of each trial 4 items blinked off and on to identify them as targets. Then all items moved about the display and participants counted the total number of times a target circle touched a red line. At an unpredictable time between 6 and 10s from the onset of motion, there was a brief 200ms blank interval, followed by a final test display consisting of stationary items. Either 1 item was missing (a random target or distractor) or 4 items were missing (all of the targets or all of the distractors). A single digit at the center of the display informed participants of how many items were missing. The task was to indicate where the missing items were located, clicking directly on the location of a single missing item, or on the “centroid” of 4 missing items.

difficult. In Experiment 2, targets were white, while distractors were black, facilitating target selection by separating targets from distractors in feature space. 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 find 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 were at chance 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 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 eight participants who were between the ages of 18 and 35, gave informed consent, and were paid \$10/hour for their participation.

Apparatus

Experiments were run using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), on a 35° by 28° display, viewed from 57 cm.

Stimuli

As shown in Figure 1, eight black circles (radius=.35°) moved at a constant rate of 4°/s, within a central region of the screen marked by a black, square outline (24.5° x 24.5°, line thickness = .1°). Two diagonal red lines connected the corners of the square (line thickness = .1°), and the background was gray. The motion direction of the circles 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 Figure 1). Each trial began with 4 items flashing off and on for two seconds to designate them as tracking targets. Then each item moved for a random duration between 6 and 10 seconds. 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 insured that participants were continuously focusing their attention on the target items.

At the end of each trial all circles disappeared for 200 ms, and then some of them reappeared. In the individual test condition, either a single target or a single distractor was missing. In the centroid test condition, either all 4 targets or all 4 distractors were missing. A cue appeared at the center of the screen to inform participants of how many items were missing (the number “1” for individual tests, and “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 4 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 the number of touches 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. In this phase of the task, participants were not required to track any targets. They were simply shown the test screen with either 1 or 4 items missing, and were asked to guess where they thought the missing items were located. These displays were generated in the same way as in the test phase, but participants were only shown the final frame. This provided an estimate of how well participants could guess where the missing items were located based on the configuration of the items that were present in the test display, providing an empirical estimate of chance performance.

EXPERIMENT 1: Withdrawing Attention with Multiple Object Tracking

This experiment tested how well participants could judge the location of individual items or the centroid of a group outside the focus of attention. We hypothesize that the tracking task would draw attention away from distractors, resulting in better accuracy for targets than distractors, but that performance would be more robust to the withdrawal of attention on centroid tests than individual tests.

Method

See General Method.

Results & Discussion

Overall participants accurately performed the primary counting task, typically missing 1 or 2 touches. Counting errors did not vary systematically with the type of localization test (individual vs. centroid, $F(1,7) = 2.06, p = .194, r^2 = 0.23$), or with the type of item that disappeared (target vs. distractor, $F < 1$). Error in the guessing phase gives us an empirical estimate of chance performance (see dashed lines, Figure 2), averaging 10.0° ($SEM = .5^\circ$) for individual tests, and 5.4° ($SEM = 0.1^\circ$) for centroid tests.

Localization accuracy revealed an important difference between individual tests and centroid tests. Figure 2a shows that 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 distractors, suggesting that attention was relatively 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, even this 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 ($t(7) = 1.07, p = .321, r^2 = .14$). 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 target centroid. Performance was also 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 for 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 1, 2, 3, or 4 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 assumes 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, we could simulate how accurately the centroid of all 4 distractors could be determined if 1, 2, 3, or 4 estimates were averaged. The results of this simulation suggest that, given how noisy the individual estimates appear to be, participants would have to pool signals from all of the distractors to achieve the level of accuracy we observed. Critically, 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 centroid as accurately as the did¹.

It appears that participants can make accurate judgments about distractors as a group by pooling information from all of the individual distractors, even when the individual details of the distractors are not accurately represented. This suggests that participants maintain some awareness of the summary features of items appearing outside the focus of attention, even when local information is represented inaccurately. However, because the targets and distractors were physically identical in Experiment 1, it was possible that participants could not completely filter out the distractor items, or that distractors were occasionally confused for targets, and therefore distractors received enough attention to improve localization of the distractor centroid. Indeed, participants

¹A Supplementary Appendix reporting the details of all simulations reported in this paper can be obtained directly from author G.A., or downloaded online at <http://cvcl.mit.edu/george/Publications.htm>.

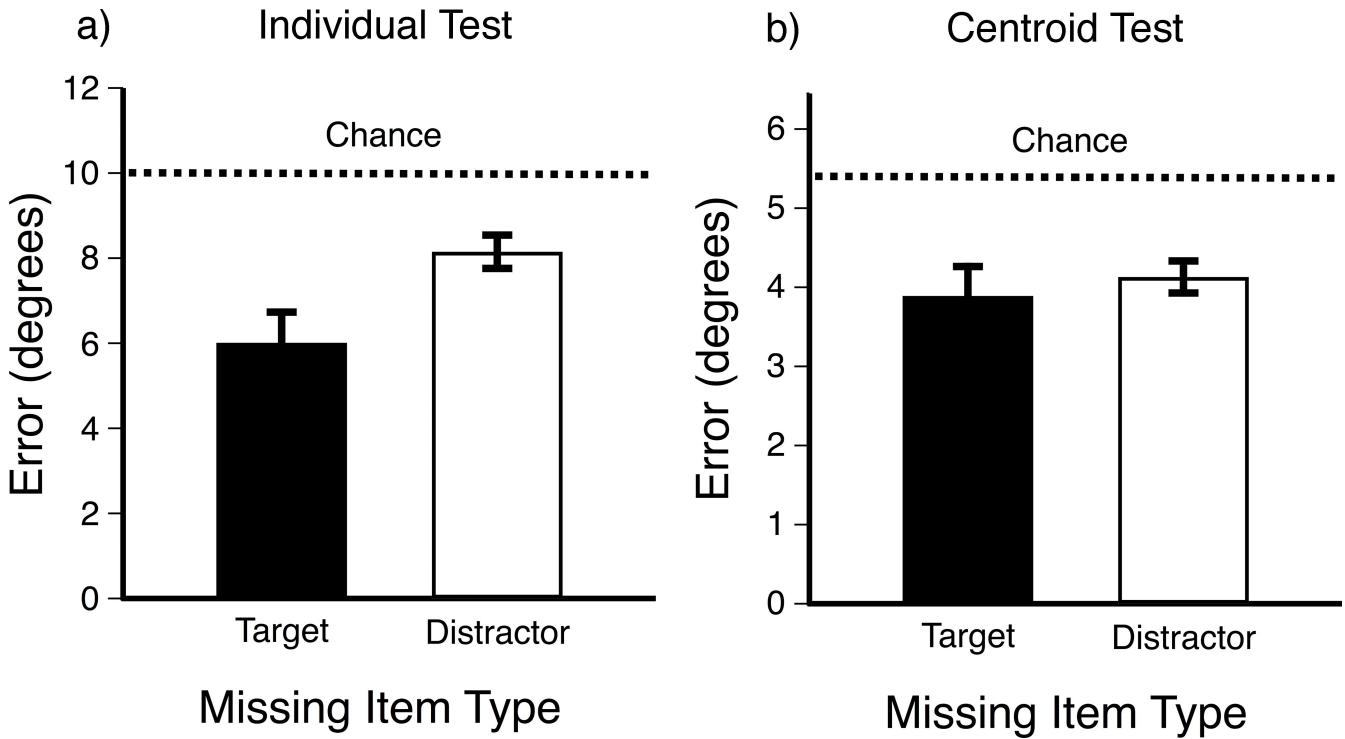


Figure 2. Results of Experiment 1. Error in degrees for localizing missing items. Error bars represent 1 s.e.m. a) Error in localizing individual targets and distractors. Performance was better for targets than distractors, and was better than chance for both targets and distractors. b) Error in localizing the centroid of targets and distractors. Performance was not significantly different for targets and distractors, and was better than chance for both targets and distractors.

could localize distractors at better than chance levels, suggesting that some attention “spilled over” to the distractors. Experiments 2 and 3 test the effect of improving selective processing of targets on the localization of individuals and their centroid.

EXPERIMENT 2: Increased Target Selection Using a Salient Feature Difference

Previous work has argued that what we see depends on how we tune our attention - our 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). On the other hand, there appears to be little to no perception of information that falls outside of 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). The current experiment 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 of the attentional set.

Method

All aspects of the stimuli and procedure were identical to Experiment 1, except that the targets were white instead of black.

Results & Discussion

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

As in Experiment 1, localization accuracy showed different patterns in the individual test and centroid test conditions. As shown in Figure 3a, 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 better than the empirical chance estimate for targets ($t(7) = 5.18$, $p = .001$, $r^2 = .79$), but as expected error was not better than chance for distractors ($t(7) = 1.00$, $p = .350$, $r^2 = .13$). This indicates that the salient feature difference was effective in enhancing selective target processing.

Figure 3b shows that in the centroid condition, there was a significant difference in localization error for targets and dis-

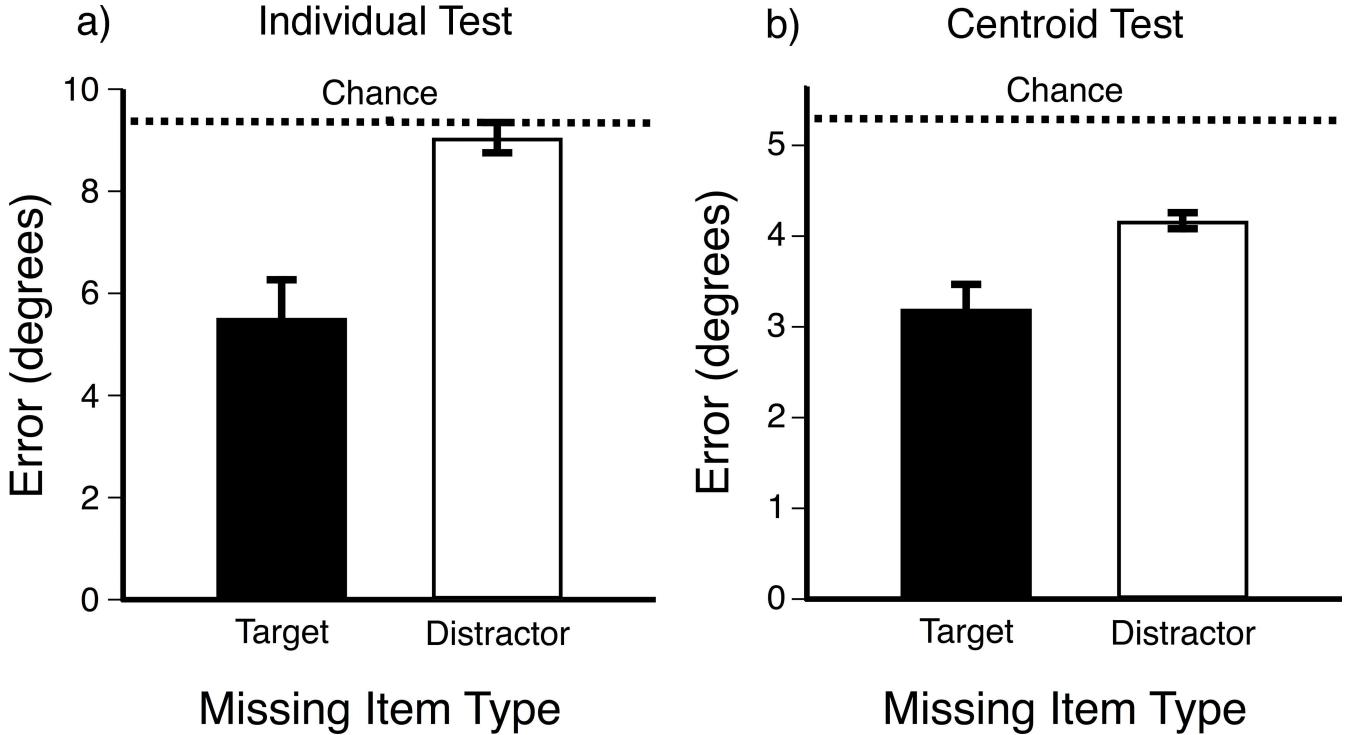


Figure 3. Results of Experiment 2. Error in degrees for localizing missing items. Error bars represent 1 s.e.m. a) Error in localizing individual targets and distractors. Performance was better for targets than distractors, and was better than chance for targets, but not for distractors. b) Error in localizing the centroid of targets and distractors. Performance was significantly better for targets than distractors, and was better than chance for both targets and distractors.

tractors ($t(7) = 3.37, p = .012, r^2 = .62$), but most importantly, 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.

Simulation results (see Note 1) again suggest that participants would have to pool signals from all of 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 would often encompass several of the distractors. It is possible that distractors frequently received diffuse attention in these displays, which was insufficient to do the local computations necessary to accurately judge the individual distractor loca-

tions, but sufficient to accurately make judgments about the centroid of the distractors (Chong & Treisman, 2005a). Experiment 3 investigates this possibility.

EXPERIMENT 3: Withdrawing Attention from the Distractor Region

In the current study, we attempted to withdraw attention from the region of space containing distractors by constraining the targets and distractors to move in opposite halves of the display. Targets were randomly constrained to move within one half of the screen (top, bottom, left, or right), and distractors were constrained to move within the opposite half of the screen. As in the previous experiments, continuous tracking of the targets was still 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 insured that distractors never fell within the convex hull formed by the targets. Of principal interest was whether this manipulation would eliminate the ability to make accurate judgments about the distractor centroid.

Method

All aspects of the stimuli and procedure were identical to Experiment 1, except that targets remained spatially separate

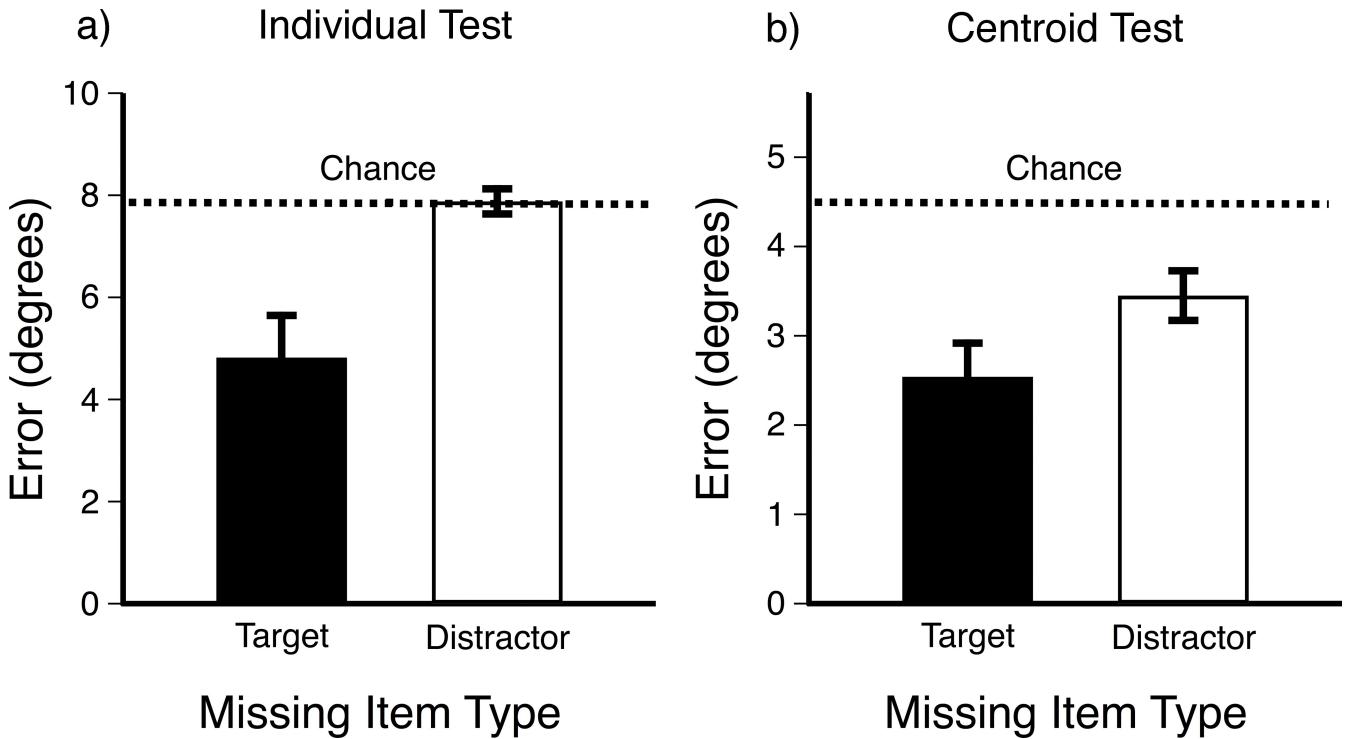


Figure 4. Results of Experiment 3. Error in degrees for localizing missing items. Error bars represent 1 s.e.m. a) Error in localizing individual targets and distractors. Performance was better for targets than distractors, and was better than chance for targets, but not for distractors. b) Error in localizing the centroid of targets and distractors. Performance was significantly better for targets than distractors, and was better than chance for both targets and distractors.

from the distractors for the entire trial.

Results & Discussion

Counting errors did not vary systematically with the type of localization test (individual vs. centroid, $F(1,7) = 2.53, p = .156, r^2 = 0.27$), or with the type of item that disappeared (target vs. distractor, $F < 1$). Error in the guessing phase (see dashed lines, Figure 4), was on average 7.9° ($SEM = .5^\circ$) for individual tests, and 4.5° ($SEM = 0.3^\circ$) for centroid tests.

As shown in Figure 4a, 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 better than the empirical chance estimates for targets ($t(7) = 4.26, p = .004, r^2 = .72$), but not for distractors ($t < 1$). Figure 4b shows that localization error in the centroid condition was significantly lower for targets than 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, yet participants could localize the distractor centroid well above chance.

Spatially separating the targets from distractors insured that distractors never fell within the convex hull formed by the targets, and therefore prevented distractors from receiv-

ing 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 (see Note 1) also suggested that this level of accuracy could only be attained if (a) noisy individual estimates from all of the distractors were pooled together, and (b) the noisy individual estimates were centered around the actual positions of the distractors, and not truly random guesses. Thus, an accurate representation of the distractor centroid can be attained even when targets and distractors are spatially separated, facilitating target selection and keeping distractors far from the focus of attention.

GENERAL DISCUSSION

Visual information can be represented at multiple levels of abstraction, from local details to abstract features which summarize the local details. We used object location as a “case study” 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, while 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 some items would

randomly disappear, and the secondary task was to localize the missing items. The results suggested that, while participants knew very little about the local details of individual distractors, they could accurately report the centroid of the distractors.

The current findings are related to previous work on inattentional blindness and change blindness. Inattentional 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 to completely withdraw attention from the tested items, and may also include active inhibition of 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 observers were aware of, and paying some attention to, all information in the display. In this way, the current paradigm is more related to change blindness studies, in which two alternating scenes are presented, with one difference between the scenes (e.g., a single item changing color). Observers often fail to notice substantial changes in such tasks, which has been interpreted as 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 summary statistics is improved when we spread our attention diffusely (Chong & Treisman, 2005a), and Treisman (2006) has argued that summary statistics are computed automatically under diffuse attention. The current results suggest that, even when our attention is focally allocated to a subset of items, summary features can be computed outside the focus of attention. Most importantly, the representation of these summary features is more robust to the withdrawal of attention than the representation of local visual features. Surprisingly, the current studies 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, others have referred to these features as global features (Navon, 1977; Oliva & Torralba, 2001), holistic features (Kimchi, 1992), or sets (Ariely, 2001; Chong & Treisman, 2003; Chong & Treisman, 2005b). We refer to each

of 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 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 segmented objects. But “ensemble” refers to any summary statistic which collapses across individual image details, whether those details are contained within a specific spatial frequency band, and whether those details are attached to segmented 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 (Torralba, Oliva, Castelhano, & Henderson, 2006).

Conclusion

The present study shows that information outside the focus of attention remains consciously accessible in the form of an ensemble representation which 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 robustly represented outside the focus of attention. Of particular interest are ensemble features that capture the statistics of the natural world, which are likely to play a vital role in our every day perception.

References

- Ariely, D. (2001). Seeing sets: representation by statistical properties. *Psychol Sci*, 12(2), 157-62.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat Vis*, 10(4), 433-6.
- Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends Cogn Sci*, 9(7), 349-54.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Res*, 43(4), 393-404.
- Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Percept Psychophys*, 67(1), 1-13.
- Chong, S. C., & Treisman, A. (2005b). Statistical processing: computing the average size in perceptual groups. *Vision Res*, 45(7), 891-900.
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognit Psychol*, 43(3), 171-216.
- Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: a critical review. *Psychol Bull*, 112(1), 24-38.
- Mack, A., & Rock, I. (1998). *Inattentional blindness*. The 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 inattentional blindness and the capture of awareness. *Psychol Rev*, 112(1), 217-42.

- 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 inattentional blindness. *Psychol Sci, 12*(1), 9-17.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *9*(3), 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.
- Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis, 10*(4), 437-42.
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat Vis, 3*(3), 179-97.
- 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. *8*(5), 368-373.
- Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processing. *Can J Exp Psychol, 54*(1), 1-14.
- Torralba, A., Oliva, A., Castelhano, 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. *Psychol Rev, 113*(4), 766-86.
- Treisman, A. (2006). How the deployment of attention determines what we see. *Vis cogn, 14*(4-8), 411-443.

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).*

$$\text{errAngle} = \text{rand}(0,359)$$

$$\text{errMagnitude} = \text{meanErr} + \text{randn} * (\text{stdErr})$$

$$\text{predictedX} = \text{actualX} + \text{magnitude} * \cos(\text{angle}/180*\pi)$$

$$\text{predictedY} = \text{actualY} + \text{magnitude} * \sin(\text{angle}/180*\pi)$$

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.*

$$\text{predictedCentroidX} = \text{average}(\text{predictedX1}, \text{predictedX2}...)$$

$$\text{predictedCentroidY} = \text{average}(\text{predictedY1}, \text{predictedY2}...)$$

4. *The error in these centroid estimates was calculated*

$$\text{errX} = \text{predictedCentroidX} - \text{actualCentroidX}$$

$$\text{errY} = \text{predictedCentroidY} - \text{actualCentroidY}$$

$$\text{predErrMagnitude} = \sqrt{\text{errX}^2 + \text{errY}^2}$$

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.

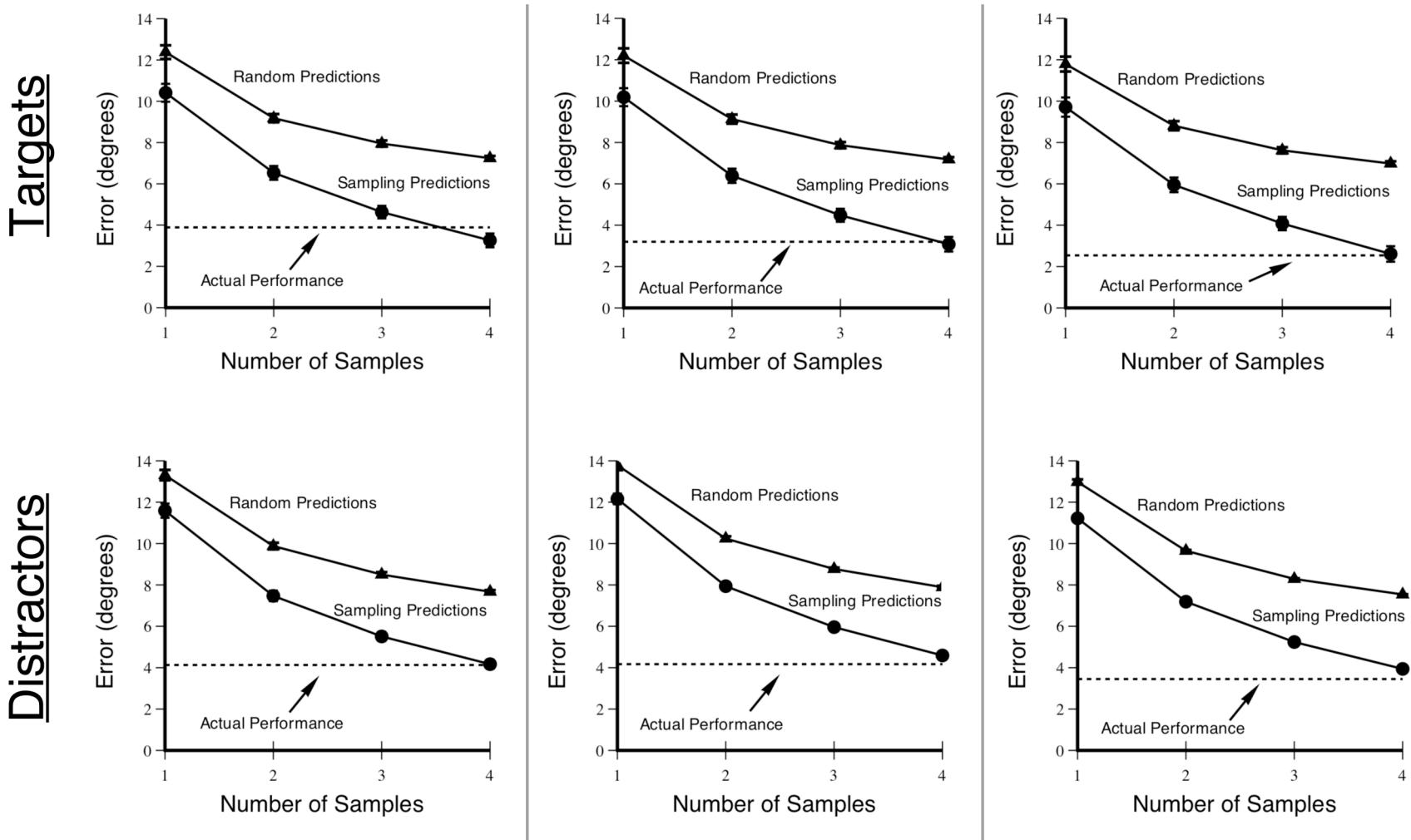


Figure A1. Predicted centroid localization error in each Experiment. The y-axis shows the predicted error in degrees, and the x-axis shows the number of individual positions sampled to predict the centroid position. Error bars are presented where greater than the data symbols and represent 1 s.e.m. For reference, the dashed lines show 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.