Spatial Ensemble Statistics Are Efficient Codes That Can Be Represented With Reduced Attention

Spatial ensemble statistics are efficient codes that can be represented with reduced attention

George A. Alvarez[a,1] and Aude Oliva[b]

[a]Department of Psychology, Harvard University, Cambridge, MA 02138; and [b]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Anne Treisman, Princeton University, Princeton, NJ, and approved March 13, 2009 (received for review September 10, 2008)

There is a great deal of structural regularity in the natural environment, and such regularities confer an opportunity to form compressed, efficient representations. Although this concept has been extensively studied within the domain of low-level sensory coding, there has been limited focus on efficient coding in the field of visual attention. Here we show that spatial patterns of orientation information (''spatial ensemble statistics'') can be efficiently encoded under conditions of reduced attention. In our task, observers monitored for changes to the spatial pattern of background elements while they were attentively tracking moving objects in the foreground. By using stimuli that enable us to dissociate changes in local structure from changes in the ensemble structure, we found that observers were more sensitive to changes to the background that altered the ensemble structure than to changes that did not alter the ensemble structure. We propose that reducing attention to the background increases the amount of noise in local feature representations, but that spatial ensemble statistics capitalize on structural regularities to overcome this noise by pooling across local measurements, gaining precision in the representation of the ensemble.

efficient coding | natural image statistics | summary statistics

The visual system encounters a continuous flow of information that must be parsed and grouped into meaningful features, surfaces, objects, and events to enable visual recognition and action in the world. Although visual analysis is a computational challenge, there is a tremendous amount of redundancy in natural images (1, 2) upon which the visual system can capitalize to form a more efficient coding scheme (3, 4). For example, natural images have a great deal of regularity in contrast and intensity distributions (5, 6), chromatic structure (7–10), reflectance spectra (11, 12), and spatial structure (2, 3, 13–15). Statistical regularities and correlations are a form of redundancy, and where there is redundancy, there is an opportunity to compress information (16, 17).

Since Shannon's theory of information (17, 18), psychology research has focused on the idea that the visual system calculates efficient codes for representing information (19–21). At the computational level, efficient coding uses the statistics of the image ensemble to derive a compact code that maximally reduces the redundancy in the patterns with minimal loss of information (22). In the past decades, efficient coding theory has provided an explanation for a wide range of properties of the visual and auditory system (23–29). However, there has been limited focus on efficient coding in the field of attention, where capacity limitations are pronounced and the ability to use compressed codes is therefore essential to compensate for the limited resources available for high-level visual cognition. Compressed representations would enable us to represent more about the environment when few resources are available.

The current study takes a step toward integrating these issues of attentional resource limitations, information redundancy, compression, and the efficient coding of visual information. Recently there has been increased interest in the ability of human observers to perceive statistical properties of a visual scene, which can be seen as a form of efficient coding in high-level vision. For instance, observers can quickly and efficiently perceive the number of objects (30), the mean size of objects (31, 32), the centroid of a collection of objects (33), or the average facial emotion or gender in a crowd (34). Critically, these statistical properties represent a single, compact statistical summary of all of the information in a scene. However, many of the regularities in natural images involve the layout of spatial frequency and orientation information across the visual field (35, 36). Here we demonstrate that the visual system can also represent such spatial layout statistics, even under conditions of reduced attention. Thus, statistical summary features that resemble the statistical structure of natural images are efficient codes, which are robust to noise and can be represented in the human visual system under conditions of reduced attention.

The Costs of Reduced Attention and the Benefits of Computing Ensemble Statistics

Visual attention enables us to manage the overwhelming flow of input information by selecting a subset of the incoming information for further, prioritized processing. However, such selective processing comes at a great cost: the less attention we pay to objects, the less precisely their features are represented (37–39). For example, directing attention away from an object reduces the perceived clarity (40), contrast (41), and high-frequency response for the object (42, 43).

One way the visual system can compensate for the costs associated with reduced attention to objects and regions in the visual field is to pool local features to form a statistical summary. Such summary statistics are a compressed code that is less rich than the full distribution of local details yet provides an accurate representation of group features. The increased precision of ensemble statistics is because independent sources of noise cancel out when averaged together. For example, if the task were to judge the size of 2 discs, and their mean size (e.g., ref. 31), judgments of the mean size would tend to be more accurate than judgments of the individual sizes, because noise in the individual estimates cancels out when averaged together (assuming independent sources of noise for each individual). The degree of benefit from averaging will be proportional to the amount of noise or uncertainty in the representation of the individual elements. Previous research has demonstrated that judgments of mean size (31, 44) and mean location (33) are more accurate than judgments of these features for individual items. Judgments of mean size have also been shown to be quite robust, remaining accurate with 50-ms presentation durations (44) for dense arrays (32) and for a wide range distributions (44).

The representation of the centroid location of a set of objects is also quite robust, remaining accurate even when attention is

Author contributions: G.A.A. and A.O. designed research; G.A.A. performed research; G.A.A. analyzed data; and G.A.A. and A.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: alvarez@wjh.harvard.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0808981106/DCSupplemental.

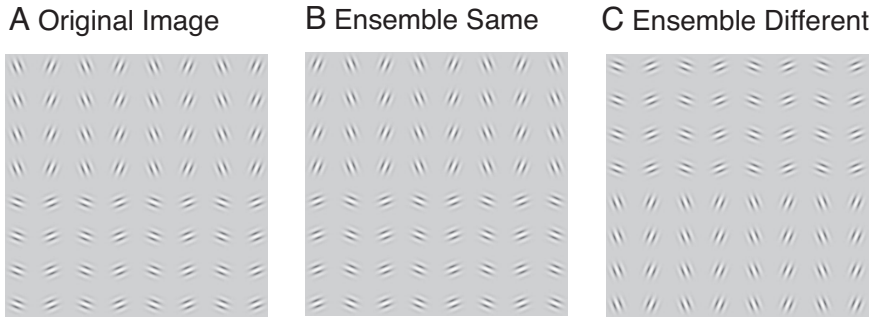

Fig. 1. Ensemble structure defined by the spatial pattern of orientation information in the top and bottom regions of the image. (*A*) The original image has a top vertical/bottom horizontal structure. (*B*) This image was generated by rotating each Gabor patch in the original image exactly 45° such that the ensemble structure remains top vertical/bottom horizontal. (*C*) This image was generated by rotating each Gabor patch in the original image exactly 45° such that the ensemble structure changes to top horizontal/bottom vertical.

withdrawn from the objects. For example, in an attentional tracking task (39), observers tracked objects moving continuously in a field of moving distractors. At a random moment during the trial, the distractors were deleted from the display. Observers could accurately report the position of a single missing target item but performed near chance when judging the position of a single missing distractor item. However, they could report the center of mass of the items (the centroid) well above chance, for both targets and distractors. Importantly, it is not the case that observers lacked knowledge about the distractors. Monte Carlo simulations suggested that the accuracy for judging the centroid position was predictable from the accuracy for judging the location of individual items. Thus, as one would predict based on the benefits of averaging, an accurate statistical summary can be computed even from very noisy local measurements.

This previous work explored the ability to extract a very compact statistical summary feature, the centroid. Such compressed representations play an important role in eye movement behavior. For instance, when saccading to a dot cluster, saccade-target landing positions appear to be based on a centroid representation (45, 46). However, there are other forms of statistical regularity in natural images, such as the layout of spatial frequency and orientation information in the visual field. These spatial statistics have proven useful for scene categorization (15, 36), and for guiding attention to task-relevant regions of the visual field (47). The purpose of the current study was to explore whether summary statistics that more closely resemble the spatial statistics of natural images can be represented under conditions of reduced attention. For instance, a coding scheme that collapses across local details to represent ecological spatial regularities (such as the differences between top and bottom part of a scene image) would benefit from averaging. As such, ensemble coding of spatial patterns would compensate for the costs of withdrawing attention by providing a compact, accurate representation of the pattern of information outside the current focus of attention, providing a rich scene representation despite limited attentional capacity.

The Span of Ensemble Statistics. Researchers have referred to a variety of statistical summary features by different terms, such as "global features" (36, 48), "holistic features" (49), or "sets" (30–32, 44). We refer to each of these types of features under the umbrella term "ensemble statistics." "Ensemble" refers to any representation that takes multiple image details and collapses across them or associates them across space, whether those details are contained within a specific spatial frequency band, and whether those details are attached to segmented objects, parts, or locations in space. Thus, an ensemble can include singular summary features, such as the number of objects (30), the mean size of objects (31, 32), the centroid of a collection of objects (33), and higher-level summary features, such as the average facial emotion or gender in a crowd (34); or spatial summary features, such as a particular combination of local orientation and spatial frequency information (14, 36). In each case the ensemble feature is a higher-level description that collapses across or otherwise combines local features.

The Current Study. The primary question addressed in this study is whether spatial ensemble statistics are efficient codes that can be represented under conditions of reduced attention. To address this question, the experiments required a stimulus for which we could dissociate higher-order image structure from local image structure. This is difficult to achieve in natural images, but it is possible with spatial patterns of oriented objects such as Gabor patches, whose statistical properties are manipulated to match certain regularities found in natural images. Most of the spatial and spectral variance encountered by humans in natural and carpentered environments is represented by a pattern consisting of different orientations in the top and bottom part of the scene (15, 50, 51). Thus, we arranged Gabor patches to form 2 possible patterns defined by the average orientation in the top and bottom regions of the display: top vertical/bottom horizontal, or top horizontal/bottom vertical (Fig. 1).

Most importantly, these images enable us to dissociate changes in local image structure from changes to the ensemble structure. Fig. 1 *B* and *C* shows 2 visual patterns that are equally different from the original (Fig. 1*A*) in their local structure, as each individual Gabor patch has rotated by 45° relative to the original image. Critically, after these changes, 1 pattern conserves the same ensemble structure as the original (ensemble same, Fig. 1*B*, vertical at the top and horizontal at the bottom), whereas the other pattern results in an ensemble with a different structure (ensemble different, Fig. 1*C*, horizontal at the top and vertical at the bottom).

To probe the representation of these spatial ensembles with reduced attention, we combined a variant of the change detection task (52, 53) with an attentional tracking task (54), creating conditions similar to natural viewing where observers focused on particular objects in the scene, withdrawing some attentional resources from background information. Tracking tasks have been shown to be very attentionally demanding, keeping the focus of attention on the targets and away from other information in the display (33, 55–57).

In our experiments, observers were presented with 2 displays of Gabors (as in Fig. 1), one at a time, and asked to detect if any local Gabor patch changed orientation between the 2 displays. This task was trivially easy when it was the only task because observers would simply focus on 1 Gabor patch to report a change. But it became challenging when observers had to keep track of several white circles that moved haphazardly on top of the Gabor patterns. Fig. 2 illustrates how the 2 tasks were combined (see Fig. 2 legend and

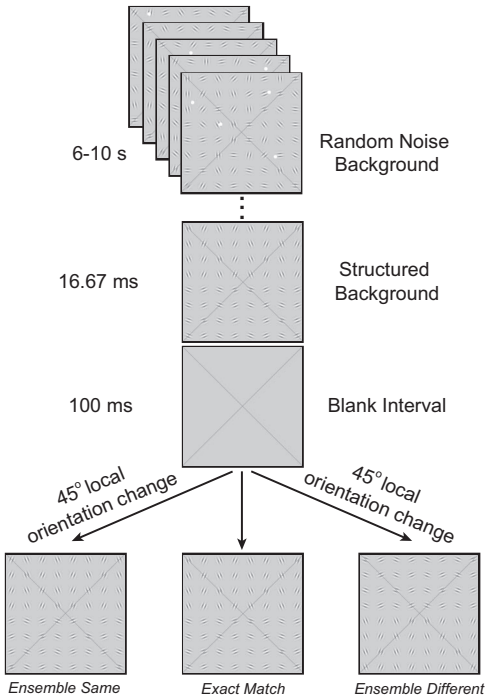

Fig. 2. Stimuli for experiment 1. For the first 6–10 s, the background consisted of Gabors that rotated and changed direction randomly. During this period the background appeared to be random and unstructured. Then the Gabors smoothly rotated into a coherent ensemble structure (e.g., vertical top, horizontal bottom, as shown), remained aligned for 16.67 ms, then disappeared for 100 ms. When the Gabors reappeared, each Gabor had exactly the same orientation (*Center*), or each Gabor changed orientation by 45° such that the ensemble structure remained the same (*Left* remains vertical top, horizontal bottom), or such that the ensemble structure changed (*Right* has changed to horizontal top and vertical bottom). In the dual task condition, the foreground also consisted of 4 white circles that moved around the display until the blank interval. In the dual task, observers focused their attention on counting the number of times the white circles touched or crossed the red lines, and then as a secondary task observers had to judge whether there was a change to the background when it reappeared.

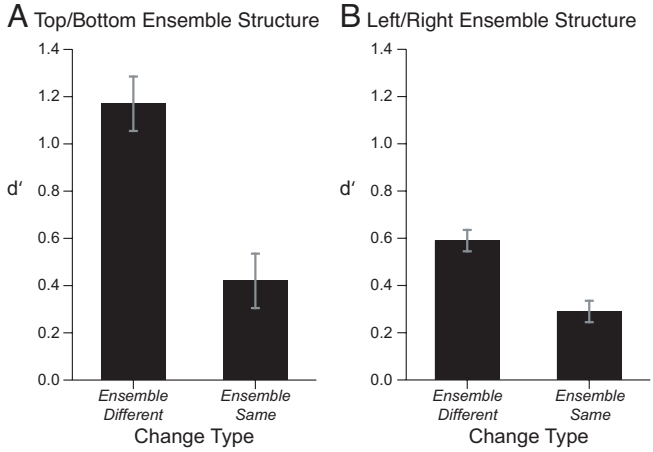

Fig. 3. Performance on the background change detection task in Experiments 1 and 2. (*A*) Results of experiment 1. Sensitivity to change (d') was greater when the ensemble structure was altered (ensemble-different) than when the ensemble structure remained the same (ensemble-same), even though the magnitude of local change was 45° for each Gabor in both conditions. (*B*) Results of experiment 2. Sensitivity to change (d') was again greater when the ensemble structure was altered (ensemble-different) than when the ensemble structure remained the same (ensemble-same). In both *A* and *B*, error bars represent within-subject standard error of the mean calculated with Cousineau's (73) modification of Loftus and Masson's method (74).

Methods for details). Critically, the background display appeared in a perfectly aligned ensemble structure at a random time during the tracking task, and only for a single video frame (≈17 ms). This manipulation was intended to discourage observers from paying attention to the background change detection task—because at any given moment the background was unlikely to contain useful information—and to encourage observers to focus mainly on the foreground tracking task.

By combining the attentional tracking and change detection tasks, we were able to probe the visual system's sensitivity to 2 types of change that were perfectly matched for local differences but varied in the degree to which they altered the ensemble structure of the image. To anticipate the results under conditions of reduced attention that increase noise in local feature representations, background changes are more noticeable when the ensemble structure of the image is altered (experiments 1 and 2). Experiment 3 rules out an alternative, local explanation for these results. Combined, these findings indicate that spatial ensemble statistics can be robustly represented with reduced attention. Such ensemble codes are compact, compressed representations that lack local precision but benefit from averaging across local features and therefore accurately represent the pattern of information in a scene.

Results

Experiment 1: Ensemble Feature Changes Are Detected with Reduced Attention. We used the combined change-detection and attentional tracking task (see Fig. 2 and *Methods*) to determine whether the ensemble structure of a background display was represented when attention was focused on a tracking task. The ensemble structure of the Gabor patches was set so that all items in the top half of the display were vertical ±22.5° and all items in the bottom half of the display were horizontal ±22.5° (top vertical/bottom horizontal displays), or vice versa (top horizontal/bottom vertical displays).

Overall, participants accurately performed the primary tracking task, typically missing 1 or 2 line touches (85% accuracy). The remaining analyses focus on the change-detection task when no more than 2 touches were missed, with raw accuracy scores reported in Table S1. We also computed sensitivity for detecting a change ($d' = Z[\text{HIT}] - Z[\text{FA}]$), where HIT is the hit rate (correctly reporting a change) and FA is the false alarm rate (incorrectly reporting a change). During the practice trials, change detection accuracy was high for both the ensemble-different condition (mean $M = 2.43$, SEM = ±0.32) and the ensemble-same condition ($M = 2.24$, SEM = ±0.55) and did not differ between the 2 conditions ($t < 1$).

In contrast, there was a large, robust difference in d' between the 2 change types in the dual task condition (see Fig. 3*A*), with a significant advantage for detecting changes in the ensemble-different condition relative to the ensemble-same condition [$t(7) = 3.2$, $P = 0.015$, $r^2 = 0.60$]. The local orientation change at each location in the display was 45° in both the ensemble-different and the ensemble-same condition, and yet there was a large reliable difference in detection of these 2 change types. Thus, when attention is withdrawn from the background elements of the display, changes that alter the ensemble structure of orientation information are more noticeable than changes of equal magnitude that do not alter the ensemble structure. This suggests that the background is represented at a higher-order, abstract level that collapses across local details and describes the pattern of orientation information over large regions of the display. This global pattern is not contained within the low spatial frequencies of the display: it is only within the ensemble of relatively high-frequency orientation information that the ensemble pattern emerges. In experiment 2, we test whether the same result will be obtained for a different spatial ensemble structure: left/right.

Experiment 2: Representation of Left/Right Structure with Reduced Attention. Here we arranged local orientation information to form 2 possible patterns that were again defined by the average orientation in 2 regions of the display: left vertical/right horizontal, or left horizontal/right vertical. This left/right structure is an accidental structure, and it accounts for less variance in natural images than the top/bottom ensemble structure used in experiment 1 (15, 51). If the visual system efficiently encodes any ensemble structure, then we would expect the results to be identical to experiment 1. However, if the system is tuned to patterns that are very diagnostic of natural image variations, as an efficient coding theory would suggest, then observers should be less sensitive to changes in the left/right structure than they were to changes in the top/bottom structure, and the advantage for detecting changes that alter the ensemble structure should be reduced or eliminated.

All aspects of the stimuli and procedure were identical to those in experiment 1, except that the variation of the ensemble structure was between left/right. Participants' accuracy at the primary tracking task was high (94%). Remaining analyses focus on d' in the change-detection task. During the practice trials, change detection accuracy was high for both the ensemble-different condition (M = 3.15, SEM = $\pm$0.52) and the ensemble-same condition (M = 3.02, SEM = $\pm$0.48) and did not differ between the 2 conditions ($t < 1$).

In contrast, there was a robust difference between the 2 change types in the dual task condition (see Fig. 3*B*), with a significant advantage for detecting changes in the ensemble-different condition relative to the ensemble-same condition [$t(7) = 2.87, P = 0.024$, $r^2 = 0.54$]. Thus, again we find that when attention is withdrawn from the background, observers are more sensitive to changes that alter the ensemble structure of the background than changes that do not alter the ensemble structure. Although the advantage in d' for the ensemble-different condition in the current experiment (M = 0.30, SEM = $\pm$0.10) was less pronounced than in experiment 1 (M = 0.75, SEM = $\pm$0.23), the interaction between change type and experiment was marginal [$F(1, 14) = 3.09$, $P = 0.10$]. Thus, further work will be necessary to determine whether changes in the top/bottom pattern of orientation information are more noticeable than differences in the left/right pattern of orientation information. This would be expected if the background is represented at an abstract level of representation along ensemble feature dimensions that reflect the statistics of the natural world. Consistent with this interpretation, top/bottom structure has been shown to be more prevalent in natural images than left/right structure. However, it is also possible that the asymmetry is related to other factors, such as differences in the strength of grouping across the left and right visual field (58), or differences between the top and bottom visual field in figure ground segmentation (59) and attentional acuity (60). Most importantly, experiment 2 replicates the finding that changes in ensemble structure are more noticeable than locally matched changes that do not alter the ensemble structure of the background. In experiment 3, we rule out an alternative explanation for the results of experiments 1 and 2 that appeals to a purely local difference between conditions.

Experiment 3: No Effect of the Categorical Nature of Local Feature Changes. We have assumed that the only difference between the ensemble-different and ensemble-same conditions is whether the ensemble structure of the display has changed. However, there is a possible local factor that must be addressed. Previous work has shown that orientation information might be coded categorically, say as steep, shallow, left-tilted, or right-tilted (61), suggesting that categorical changes in orientation might be more noticeable. For example, it might be easier to detect an orientation change when an item changes clockwise by 45° from near vertical (+22.5°) to near horizontal (+67.5°) than when the same size change in the counterclockwise direction takes an item from near vertical (+22.5°) to still near vertical (−22.5°). If this is the case, then this entirely local explanation could account for the results of experiments 1 and 2. The ensemble-different condition had categorical orientation changes (items changed from near vertical to near horizontal, or vice versa), whereas the ensemble-same condition did not (after changing near vertical items remained near vertical, and near horizontal items remained near horizontal).

To determine whether categorical orientation changes are more easily detected with reduced attention, observers performed the same dual task as in the previous experiments, but there was no coherent ensemble structure in the background. Instead, we generated random displays of Gabors, with each individual Gabor appearing either near vertical or near horizontal. There were 2 types of changes: category-different, in which each item changed orientation by 45° such that categorical orientation changes occurred; and category-same, in which each item changed orientation by 45° such that categorical orientation did not change. If categorical orientation changes are more easily detected with reduced attention, then we should see better performance in the category-different condition. If there were no such advantage, then it would confirm that the results of experiments 1 and 2 were not due to local differences. All other aspects of the stimuli and procedure were identical to experiment 1.

Overall participants accurately performed the primary tracking task with 91% accuracy. During the practice trials, change detection sensitivity (d') was high for both the category-different condition (M = 3.08, SEM = $\pm$0.32) and the category-same condition (M = 2.79, SEM = $\pm$0.14) and did not differ between the 2 conditions [$t(7) = 1.02$, $P = 0.342$, $r^2 = 0.129$].

Unlike experiments 1 and 2, change detection sensitivity (d') was not significantly different ($t < 1$) for the category-different condition (M = 0.54, SEM = $\pm$0.16) and the category-same condition (M = 0.51, SEM = $\pm$0.17). These results suggest that with reduced attention, there is no advantage for detecting changes that alter the categorical orientation of the Gabor patches: A change from vertical to near horizontal is no more noticeable than a change from near vertical to near vertical. Thus, the results of experiments 1 and 2 cannot be attributed to local differences between conditions.

Discussion

It is widely acknowledged that reducing the amount of attention paid to objects or regions decreases the precision with which their features can be represented (37–39). We have previously shown that the visual system can compensate for this decreased precision by pooling local features to represent the entire ensemble with fairly high accuracy (33), providing a relatively accurate representation of ensemble statistics even with reduced attention. This previous work explored a relatively concise ensemble feature: the center of mass of a collection of objects (the centroid). Here we tested whether ensembles that characterize the spatial distribution of spatial frequency and orientation information could be represented with reduced attention by using Gabor patterns with statistics that were constrained to resemble a distribution of orientation information that is commonly found in natural images. High noise in the representation of local image details could potentially be overcome by collapsing across those local measurements to represent the pattern of information at a more abstract level (e.g., the average orientation is "vertical" on top, and "horizontal" on the bottom of this scene). Such an ensemble code would be more compact and would benefit in accuracy from pooling across multiple local measurements. The current results indicate that such spatial patterns of orientation information can be represented with reduced attention, supporting the idea that spatial ensemble statistics can serve as a compact representation that enables the visual system to overcome its severe capacity limitations.

Relation to Perceptual Grouping With/Without Attention. The ability to compute and represent spatial ensemble statistics under conditions of reduced attention is conceptually related to whether perceptual grouping can occur with or without attention. Early

research using the inattentional blindness paradigm suggested that perceptual grouping and texture segregation of background stimuli does not occur without attention (62). However, this paradigm requires observers to recall their past experiences, and therefore it is unclear whether observers actually did not perceive the background, or simply could not recall what it looked like (63). To overcome this problem, subsequent research used an online measure of whether the background elements formed a perceptual group (64). The results showed that online task performance was biased by a perceptual illusion induced by the grouping of background elements, even though subsequently observers could not accurately recall the appearance of the background.

Although texture segregation and perceptual grouping are closely related to ensemble processing, our claim is *not* that ensemble statistics can be processed without attention or awareness (although they may be). There is strong evidence that there is no conscious perception of information without attention (65–68), whereas our hypothesis is about conscious, attentive vision. Specifically, our claim has 2 components: (*i*) reducing the amount of attention paid to an object or region reduces the quality and precision with which local visual features are represented, and (*ii*) ensemble coding can compensate for this loss of local precision by collapsing across local details to form an accurate representation of the ensemble. On this view, if attention was completely withdrawn, leaving no representation of local details, then it would be impossible to represent and consciously perceive the ensemble. Thus, our claims about ensemble statistics are about explicit, conscious perception under conditions of reduced attention—not in the absence of attention.

Role of Attention in the Perception of Ensemble Statistics. Previous research has explored the possibility that ensemble statistical processing is more efficient under conditions of distributed attention (when attention is spread across the entire display), than under conditions of focal attention (when attention is focused narrowly) (69). On a first pass, the attention task used in the current paradigm might seem to require distributed attention, given that observers are attending to multiple objects that are spatially distributed across the display. Counter to this intuition, previous work suggests that attention is actually focally allocated to targets in such attentive tracking tasks, and does not spread over the space between targets (55, 56). Attention appears to select and track targets as if there were multiple, independent foci of attention (70). On this view, the current results would suggest that, even when our attention is focally allocated to a subset of items, spatial ensemble statistics could be computed outside the focus of attention. However, without independent evidence to support the assumption that targets are tracked by means of focal attention, it remains possible that the background received diffuse attention in these displays, and that such diffuse attention is necessary for computing spatial ensemble statistics. Future experiments can address this question by requiring observers to track a single target versus multiple targets (matching these conditions for difficulty by adjusting speed) (37). If spatial ensemble statistics are computed more accurately with diffuse attention, then the ability to detect changes to the ensemble should be greater when tracking multiple targets with diffuse attention than when tracking a single target with focal attention.

Conclusion

Understanding of low-level sensory coding has been greatly advanced by exploring the relationship between the statistical regularities present in the natural environment (particularly spatial regularities) and perceptual coding mechanisms. Although relatively little research on high-level vision has explored efficient coding per se, research on high-level vision has explored the ability to perceive compact ensemble statistics of a visual scene (e.g., the average size of objects). Although such statistical processing can be thought of as a form of efficient coding, a direct link to efficient coding in low-level vision has been missing, partly because the statistical properties investigated have been so different. Here we provided evidence that the visual system represents spatial ensemble statistics—patterns of spatial frequency and orientation information—and that these representations are robust to the withdrawal of attention. Expanding the notion of ensemble statistics to such spatial patterns provides an important bridge between studies of efficient coding in high-level vision (ensemble coding, set perception, statistical perception, holistic processing), and efficient coding in low-level sensory coding. This opens the door to exploring how regularities in natural images affect not only low-level sensory coding but also high-level vision. An important avenue for future research is to explore the relationship between natural image statistics and the efficiency of spatial ensemble coding. One intriguing possibility is that the efficiency with which a particular spatial ensemble is represented is proportional to the likelihood of that spatial ensemble occurring in natural images.

Methods

Participants. Separate groups of 8 observers participated in experiments 1 and 2. A partially overlapping group of 8 observers participated in experiment 3 (6 new, 1 who participated in experiment 1, 1 who had participated in experiment 2). All participants were 18 to 35 years old, gave informed consent, and were paid $10/h.

Apparatus. Experiments were run using the Psychophysics Toolbox extensions (71, 72) on a 35° by 28° cathode ray tube display, viewed from ≈57 cm.

Stimuli. Fig. 2 shows the stimuli for a sample trial. Four white circles (radius = 0.35°) moved at a constant rate of 4°/s, within a central region of the screen marked by a black, square outline (24.5° × 24.5°, line thickness = 0.1°). Two diagonal red lines connected the corners of the square (line thickness = 0.1°), and the background was gray. The motion direction of the circles was constrained such that items appeared to avoid one another while moving randomly about the display.

The stimuli for the background change detection task consisted of an 8 × 8 grid of 100% contrast Gabor patches (2 cycles per degree), each subtending ≈2.5° by 2.5°. Displays were linearized with gamma correction (gamma = 2.25 for all color guns) to prevent higher-order luminance artifacts. To mask the ensemble structure of the background items, the Gabors were initially randomly oriented. Then each Gabor patch began to spin at 60°/s in the clockwise or counterclockwise direction, changing direction with a 1/30 chance on each video frame (≈2 times per second). The initial rotation direction and direction changes were chosen randomly and independently for each Gabor. During this phase of the trial, the background appeared to be noisy and unstructured. Then, at a randomly determined time between 6 and 10 s, the Gabors smoothly aligned into a coherent ensemble structure. In experiment 1, the patches would align so that the top of the screen consisted of nearly vertical items (±22.5° from vertical), and the bottom consisted of nearly horizontal items (±22.5° from horizontal), or the opposite pattern. In experiment 2, the patches would align so that the left of the screen consisted of nearly vertical items (±22.5° from vertical), and the right consisted of nearly horizontal items (±22.5° from horizontal), or the opposite pattern. In experiment 3, the patches would align so that the half of the items appeared nearly vertical (±22.5° from vertical), and the other half appeared nearly horizontal (±22.5° from horizontal), but the items were randomly positioned so that there was no ensemble structure.

The displays were perfectly aligned in this coherent ensemble structure for only 1 video frame (16.67 ms), then disappeared for 100 ms , and reappeared in 1 of 3 conditions: exact-match (Fig. 2 *Bottom Center*), ensemble-same (Fig. 2 *Bottom Left*), or ensemble-different (Fig. 2 *Bottom Right*), as described above. Critically, in both the ensemble-same and ensemble-different conditions, the Gabors changed orientation by exactly the same amount (45°), so these conditions were perfectly matched in terms of the magnitude of local feature change.

Procedure. ***Single-task practice phase.*** Over the course of 6–10 s, the Gabor patches rotated and occasionally changed direction, such that the background appeared to be random and unstructured. Then the Gabors rotated into a coherent ensemble structure (e.g., vertical top, horizontal bottom, as shown in Fig. 2), remained aligned for 16.67 ms, disappeared for 100 ms, and then reappeared. Observers were instructed to determine whether any of the Gabors in the display changed orientation when they reappeared. No explicit instruction was given regarding the ensemble structure in the display. Observers were informed that when 1 Gabor changed, all of them changed, and so if they noticed anything

change, they should respond ''change,'' and otherwise they should respond ''no change.''

Dual-task phase. Four white circles appeared in the foreground of the display until the blank interval. At the beginning of each trial, the circles flashed off and on twice a second for 2 s to remind observers to focus their attention on them. Then each item moved about the display and the primary task was to attentively track the targets and to keep count of the number of times a target item touched or crossed one of the red lines that connected the corners of the display (one running count collapsed across all targets, not a separate count for each individual target). At the end of the trial, observers indicated whether they noticed any background items change, and then typed in the number of times the tracking targets touched or crossed the red lines. Although they entered their counting response second, they were instructed to focus primarily on the tracking task.

Participants performed 90 practice trials and 90 dual-task trials. The 3 conditions (exact-match, ensemble-same, ensemble-different) were equally likely and randomly intermixed within a block.

ACKNOWLEDGMENTS. We thank Timothy F. Brady, Talia Konkle, Ruth Rosenholtz, and Antonio Torralba for helpful conversation. G.A.A. was supported by National Institutes of Health/National Eye Institute fellowship F32EY016982. A.O. was supported by National Science Foundation CAREER Award 0546262 and National Science Foundation Grant 0705677.

1. Kersten D (1987) Predictability and redundancy of natural images. *J Opt Soc Am A* 4:2395–2400.
2. Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A* 4:2379–2394.
3. Field DJ (1989) What the statistics of natural images tell us about visual coding. *SPIE: Human Vision, Visual Processing, Digital Display* 1077:269–276.
4. Chandler DM, Field DJ (2007) Estimates of the information content and dimensionality of natural scenes from proximity distributions. *J Opt Soc Am A* 24:922–941.
5. Brady N, Field DJ (2000) Local contrast in natural images: Normalisation and coding efficiency. *Perception* 29:1041–1055.
6. Frazor RA, Geisler WS (2006) Local luminance and contrast in natural images. *Vision Res* 46:1585–1598.
7. Webster MA, Mollon JD (1997) Adaptation and the color statistics of natural images. *Vision Res* 37:3283–3298.
8. Hyvärinen A, Hoyer PO (2000) Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput* 12:1705–1720.
9. Judd DB, MacAdam DL, Wyszecki GW (1964) Spectral distribution of typical daylight as a function of correlated color temperature. *J Opt Soc Am A* 54:1031–1040.
10. Long F, Yang Z, Purves D (2006) Spectral statistics in natural scenes predict hue, saturation, and brightness. *Proc Natl Acad Sci USA* 103:6013–6018.
11. Maloney LT (1986) Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J Opt Soc Am A* 3:1673–1683.
12. Maloney LT, Wandell BA (1986) Color constancy: A method for recovering surface spectral reflectance. *J Opt Soc Am A* 3:29–33.
13. Burton GJ, Moorehead IR (1987) Color and spatial structure in natural scenes. *Appl Opt* 26:157–170.
14. Geisler WS, Perry JS, Super BJ, Gallogly DP (2001) Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res* 41:711–724.
15. Torralba A, Oliva A (2003) Statistics of natural image categories. *Network* 14:391–412.
16. Huffman DA (1952) A method for construction of minimum redundancy codes. *Proc IRE* 40:1098–1101.
17. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423, 523–656.
18. Shannon CE, Weaver W (1949) *The Mathematical Theory of Communication* (Univ Illinois Press, Urbana, IL).
19. Barlow HB (2001) The exploitation of regularities in the environment by the brain. *Behav Brain Sci* 24:602–607, and discussion (2001) 24:652–671.
20. Attneave F (1954) Some informational aspects of visual perception. *Psychol Rev* 61:183–193.
21. Barlow HB (1961) The coding of sensory messages. *Current Problems in Animal Behaviour*, eds Thorpe WH, Zangwill OL (Cambridge Univ Press, Cambridge, UK), pp 331–360.
22. Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. *Annu Rev Neurosci* 24:1193–1216.
23. Atick JJ, Redlich AN (1992) What does the retina know about natural scenes? *Neural Comput* 4:196–210.
24. Atick JJ (1992) Could information theory provide an ecological theory of sensory processing? *Network Comput Neural Syst* 3:213–251.
25. Atick JJ, Redlich AN (1990) Mathematical-model of the simple cells of the visual cortex. *Biol Cybern* 63:99–109.
26. Barlow HB, Foldiak P (1989) Adaptation and decorrelation in the cortex. *The Computing Neuron*, eds Durbin R, Miall C, Mitchison G (Addison–Wesley, Reading, MA), pp 54–72.
27. Daugman JG (1988) Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Transact Acoustics, Speech Signal Process* 36:1169–1179.
28. Olshausen BA, Field DJ (1996) Natural image statistics and efficient coding. *Network* 7:333–339.
29. Lewicki MS (2002) Efficient coding of natural sounds. *Nat Neurosci* 5:356–363.
30. Halberda J, Sires SF, Feigenson L (2006) Multiple spatially overlapping sets can be enumerated in parallel. *Psychol Sci* 17:572–576.
31. Ariely D (2001) Seeing sets: Representation by statistical properties. *Psychol Sci* 12:157–162.
32. Chong SC, Treisman A (2005) Statistical processing: Computing the average size in perceptual groups. *Vision Res* 45:891–900.
33. Alvarez GA, Oliva A (2008) The representation of simple ensemble visual features outside the focus of attention. *Psychol Sci* 19:392–398.
34. Haberman J, Whitney D (2007) Rapid extraction of mean emotion and gender from sets of faces. *Curr Biol* 17:R751–R753.
35. Torralba A, Oliva A (2002) Depth estimation from image structure. *IEEE Pattern Anal Mach Intell* 24:1226–1238.
36. Oliva A, Torralba A (2001) Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int J Comput Vision* 42:145–175.
37. Alvarez GA, Franconeri SL (2007) How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *J Vision* 7:1–10.
38. Franconeri SL, Alvarez GA, Enns JT (2007) How many locations can be selected at once? *J Exp Psychol Hum Percept Perform* 33:1003–1012.
39. Palmer J (1990) Attentional limits on the perception and memory of visual information. *J Exp Psychol Hum Percept Perform* 16:332–350.
40. Titchener EB (1908) *Lectures on the Elementary Psychology of Feeling and Attention* (Macmillan, New York).
41. Carrasco M, Ling S, Read S (2004) Attention alters appearance. *Nat Neurosci* 7:308–313.
42. Carrasco M, Williams PE, Yeshurun Y (2002) Covert attention increases spatial resolution with or without masks: Support for signal enhancement. *J Vision* 2:467–479.
43. Yeshurun Y, Carrasco M (1998) Attention improves or impairs visual performance by enhancing spatial resolution. *Nature* 396:72–75.
44. Chong SC, Treisman A (2003) Representation of statistical properties. *Vision Res* 43:393–404.
45. Melcher D, Kowler E (1999) Shapes, surfaces and saccades. *Vision Res* 39:2929–2946.
46. McGowan JW, Kowler E, Sharma A, Chubb C (1998) Saccadic localization of random dot targets. *Vision Res* 38:895–909.
47. Torralba A, Oliva A, Castelhano MS, Henderson JM (2006) Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychol Rev* 113:766–786.
48. Navon D (1977) Forest before trees: The precedence of global features in visual perception. *Cognit Psychol* 9:353–383.
49. Kimchi R (1992) Primacy of wholistic processing and global/local paradigm: A critical review. *Psychol Bull* 112:24–38.
50. Field DJ (1994) What is the goal of sensory coding? *Neural Comput* 6:559–601.
51. Oliva A, Torralba A (2006) Building the gist of a scene: The role of global image features in recognition. *Prog Brain Res* 155:23–36.
52. Pashler H (1988) Familiarity and visual change detection. *Percept Psychophys* 44:369–378.
53. Phillips WA (1974) On the distinction between sensory storage and short-term visual memory. *Percept Psychophys* 16:283–290.
54. Pylyshyn ZW, Storm RW (1988) Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spat Vision* 3:179–197.
55. Intriligator J, Cavanagh P (2001) The spatial resolution of visual attention. *Cognit Psychol* 43:171–216.
56. Sears CR, Pylyshyn ZW (2000) Multiple object tracking and attentional processing. *Can J Exp Psychol* 54:1–14.
57. Most SB, et al. (2001) How not to be seen: The contribution of similarity and selective ignoring to sustained inattentional blindness. *Psychol Sci* 12:9–17.
58. Pillow J, Rubin N (2002) Perceptual completion across the vertical meridian and the role of early visual cortex. *Neuron* 33:805–813.
59. Rubin N, Nakayama K, Shapley R (1996) Enhanced perception of illusory contours in the lower versus upper visual hemifields. *Science* 271:651–653.
60. He S, Cavanagh P, Intriligator J (1996) Attentional resolution and the locus of visual awareness. *Nature* 383:334–337.
61. Wolfe JM, Friedman-Hill SR, Stewart MI, O'Connell KM (1992) The role of categorization in visual search for orientation. *J Exp Psychol Hum Percept Perform* 18:34–49.
62. Mack A, Tang B, Tuma R, Kahn S, Rock I (1992) Perceptual organization and attention. *Cognit Psychol* 24:475–501.
63. Wolfe JM (2000) Inattentional amnesia. *Fleeting Memories*, ed Coltheart V (MIT Press, Cambridge, MA), pp 71–94.
64. Moore CM, Egeth H (1997) Perception without attention: Evidence of grouping under conditions of inattention. *J Exp Psychol Hum Percept Perform* 23:339–352.
65. Mack A, Rock I (1998) *Inattentional Blindness* (MIT Press, Cambridge, MA).
66. Most SB, Scholl BJ, Clifford ER, Simons DJ (2005) What you see is what you set: Sustained inattentional blindness and the capture of awareness. *Psychol Rev* 112:217–242.
67. Neisser U, Becklen R (1975) Selective looking: Attending to visually specified events. *Cognit Psychol* 7:480–494.
68. Yantis S (2003) To see is to attend. *Science* 299:54–56.
69. Chong SC, Treisman A (2005) Attentional spread in the statistical processing of visual displays. *Percept Psychophys* 67:1–13.
70. Cavanagh P, Alvarez GA (2005) Tracking multiple targets with multifocal attention. *Trends Cognit Sci* 9:349–354.
71. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vision* 10:433–436.
72. Pelli DG (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat Vision* 10:437–442.
73. Cousineau D (2005) Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology* 1:42–45.
74. Loftus GR, Masson ME (1994) Using confidence intervals in within-subject designs. *Psychon Bull Rev* 1:476–490.