Journal of Experimental Psychology: General
2015, Vol. 144, No. 2, 432–446

0096-3445/15/$12.00 http://dx.doi.org/10.1037/xge0000053

Individual Differences in Ensemble Perception Reveal Multiple, Independent Levels of Ensemble Representation

Jason Haberman
Harvard University and Rhodes College

Timothy F. Brady and George A. Alvarez
Harvard University

Ensemble perception, including the ability to "see the average" from a group of items, operates in numerous feature domains (size, orientation, speed, facial expression, etc.). Although the ubiquity of ensemble representations is well established, the large-scale cognitive architecture of this process remains poorly defined. We address this using an individual differences approach. In a series of experiments, observers saw groups of objects and reported either a single item from the group or the average of the entire group. High-level ensemble representations (e.g., average facial expression) showed complete independence from low-level ensemble representations (e.g., average orientation). In contrast, low-level ensemble representations (e.g., orientation and color) were correlated with each other, but not with high-level ensemble representations (e.g., facial expression and person identity). These results suggest that there is not a single domain-general ensemble mechanism, and that the relationship among various ensemble representations depends on how proximal they are in representational space.

Keywords: ensembles, summary statistics, individual differences, ensemble mechanisms

The natural world is rife with visual redundancy: For example, any given blade of grass looks like many other blades of grass. When seen as a group, overlapping features give rise to an unambiguously unified percept (e.g., a lawn). The visual system is adept at exploiting such featural redundancy, creating compressed codes in the form of summary statistics (Alvarez, 2011). Seeing the average or statistical summary of a group, often referred to as *ensemble perception*, is a robust phenomenon that operates across a host of visual domains (Haberman & Whitney, 2012), including orientation (Dakin & Watt, 1997; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), size (Ariely, 2001; Chong & Treisman, 2003), position (Alvarez & Oliva, 2008), motion (Sweeney, Haroz, & Whitney, 2012; Watamaniuk, 1993), speed (Watamaniuk & Duchon, 1992), number (Burr & Ross, 2008; Halberda, Sires, & Feigenson, 2006), and faces varying in emotion (Fischer & Whitney, 2011; Haberman & Whitney, 2007) and identity (de Fockert & Wolfenstein, 2009; Neumann, Schweinberger, & Burton, 2013). Ensembles are represented across space and time (Albrecht & Scholl, 2010; Haberman, Harp, & Whitney, 2009), are immune to outliers (Haberman & Whitney, 2010), come in several statistical forms (e.g., average, variance, and range; Dakin, 1999; Morgan, Chubb, & Solomon, 2008; Solomon, 2010) and can be computed with minimal attentional effort (Alvarez & Oliva, 2008; Haberman & Whitney, 2011).

The utility of an efficient ensemble perception system that is separate from the object-based perceptual system has broad appeal (Alvarez, 2011; Ariely, 2001; Haberman & Whitney, 2012). The ease with which the visual system represents visual scene statistics, such as orientation and color, may help to reconcile the contradictory experience of a sense of visual completeness and the well-established limitations of the representation of individual items (e.g., change blindness, attentional blink; Raymond, Shapiro, & Arnell, 1992; Rensink, 2004; Simons & Ambinder, 2005). Having access to such global scene statistics may also be instrumental in identifying relevant features, such as a visual pop-out (i.e., items that differ substantially from the average; Duncan & Humphreys, 1989), and thus be relevant for guiding attention. Other, higher level ensembles, such as the average expression of a crowd of faces (Haberman & Whitney, 2007), may be critical for identifying potential threats (e.g., the intention of the mob), can be useful for assessing whether students are confused during a class lecture, and is related to people's level of social anxiety (Yang, Yoon, Chong, & Oh, 2013).

The efficiency of ensemble perception, along with its breadth and flexibility, has led researchers to propose the existence of dedicated and specialized ensemble mechanisms. Evidence for the existence of specialized mechanisms comes from the absence of a set size effect (i.e., ensemble representation precision is fairly constant regardless of the number of items in the set; Attarha, Moore, & Vecera, 2014; Chong & Treisman, 2003, 2005; Haberman & Whitney, 2009), the speed at which ensembles may be derived (as low as 50 ms), and the availability of ensemble information even when individual item information is unavailable (i.e., as undercrowded conditions or change blindness paradigms; Fischer & Whitney, 2011; Haberman & Whitney, 2011; Parkes et

Jason Haberman, Department of Psychology, Harvard University, and Department of Psychology, Rhodes College; Timothy F. Brady, and George A. Alvarez, Department of Psychology, Harvard University.

This work was supported by a National Science Foundation CAREER Award to George A. Alvarez [BCS-0953730]. Thank you to Julie Belkova for assistance with data collection.

Correspondence concerning this article should be addressed to Jason Haberman Rhodes College, 2000 North Parkway, Memphis, TN 38107. E-mail: habermanj@rhodes.edu

al., 2001). Although there is some suggestion that ensemble representation of size could potentially be explained by sampling just one or two items from an entire array (i.e., not a specialized ensemble mechanism; Myczek & Simons, 2008), there is substantial evidence showing that many ensemble types are processed in a manner akin to textures (Balas, Nakano, & Rosenholtz, 2009), even when individual item information cannot be resolved (Fischer & Whitney, 2011; Haberman & Whitney, 2011) and therefore cannot be individually sampled. Further evidence of the specialized nature of ensemble perception comes from neuropsychological work, which has revealed a preserved ability to represent the average expression in patients with prosopagnosia (Leib et al., 2012).

The seeming ubiquity of ensemble perception suggests that it is a fundamental visual process; however, little work has focused on exploring the functional organization of ensemble processing. Critically, it remains unknown how ensemble representations relate to one another within the broader cognitive framework. Is there a single, domain-general mechanism supporting all ensemble representation subtypes, or are there multiple domain-specific mechanisms at work?

In the domain-general view, an individual who can precisely represent one feature, say, the average orientation of a set of rotated gabors, should also precisely represent other features, such as the average color and average facial expression. That is, there should be a performance relationship among disparate ensemble tasks. This framework has some appeal, as several investigators view ensemble perception as a form of texture processing (e.g., Balas et al., 2009; Parkes et al., 2001). This could lead to consistent ensemble performance across all domains, as observers' performance on ensemble tasks might depend largely on their general texture processing ability. There is neuroimaging evidence to support this view, showing ensemble-specific neural adaptation for different stimulus types in regions associated with texture processing (Cant & Xu, 2012). Rather than being based on texture processing, a domain-general mechanism could also be a dedicated and centralized "ensemble processor," agnostic to visual domain,[1] receiving and averaging individual item information from multiple sources (e.g., orientation, spatial frequency, global motion). In fact, it is possible that a single shared mechanism underlies many forms of statistical processing, including both ensemble processing and statistical learning (Zhao, Ngo, McKendrick, & Turk-Browne, 2011).

The alternative, domain-specific framework suggests that there are multiple ensemble processors specific to each type of visual domain. In other words, there are separate "units" for extracting average orientation, average person identity, and so forth. At the limit, there could be an independent "ensemble mechanism" for every separable kind of feature information, if the ability to perform ensemble processing (e.g., extract an average from a set) is a basic characteristic of the coding scheme employed by the visual system. For example, ensemble processing could arise from pooling mechanisms that operate locally, and thus only over specific sets of features (e.g., with some units pooling over low-level orientation-tuned cells, and other units pooling over higher level cells tuned to facial features, and so on). These pooling regions would act as separate ensemble mechanisms for every kind of feature.

To address whether ensemble processing is supported by a domain-general or domain-specific mechanism(s), we employed an individual differences approach. Individual differences paradigms are particularly useful for addressing questions of cognitive architecture because they capitalize on the inherent variability present in a population sample. Although such studies are traditionally used to identify clinically relevant subsamples (e.g., autism spectrum disorders), they are also well suited to explore the functional organization of cognitive processing (Huang, Mo, & Li, 2012; Underwood, 1975; Wilmer, 2008). By examining how performance on different tasks correlate (e.g., Is skill at ensemble orientation judgments correlated with skill at ensemble face judgments?), we can infer whether such processes are likely supported by the same underlying mechanism or independently operating mechanisms.

In Experiment 1, we explore the relationship between averaging person identity and averaging gabor orientation, because these stimuli are bookends in representational space—faces are high-level, meaningful stimuli, whose appearance depends on configural processing of multiple features (Maurer, Grand, & Mondloch, 2002; McKone, 2004; Tanaka & Farah, 1993); gabors are a low-level stimulus, optimized for mimicking the response properties of V1 simple cells (Daugman, 1980; Marčelja, 1980). Ample evidence has also demonstrated ensemble face processing as distinctly high level. For example, average emotion performance is disrupted when sets are inverted (Haberman & Whitney, 2009), and sophisticated social information such as the average direction of eye gaze is readily available (Sweeny & Whitney, 2014). Should these high-level versus low-level tasks be highly correlated, it would be the strongest test of the claim that ensembles are supported by a single, domain-general mechanism. If, however, they were not correlated, it would be an indication that different ensemble representations are at least partly independent.

In addition to exploring the relationship among different ensemble domains, Experiment 1 was designed to test the relationship between individual item representation and ensemble representation. Characterizing this relationship addresses whether individual item representations and ensemble representations share a common source of noise (Alvarez & Oliva, 2008), which need not be the case if they are computed over different features. For example, it is also possible that average size is computed indirectly using global spatial frequency information and an estimate of item number (i.e., bypassing any measures of individual item size; Šetić, Švegar, & Domijan, 2007), in which case the precision of individual size and average size representations could be completely independent. Alternatively, it is possible that average features are computed directly from estimates of individual feature values (Alvarez & Oliva, 2008), in which case the precision of item representations and ensemble representations would be highly correlated. By testing representations at both the individual and ensemble levels, Experiment 1 allowed us to examine this issue using an individual differences approach.

[1] By "domain," we mean stimulus category. We use "within domain" to refer to processes related to a single category type (e.g., faces), whereas we use "across domain" to refer to processes related to multiple category types (e.g., faces and oriented gabors).

In Experiments 2 through 8, we explore a wide range of feature domains, from low-level features (orientation and color) to high-level features (person identity, emotional expression). Tasks that are uncorrelated point to independently operating mechanisms, whereas tasks that are correlated are at least consistent with the idea that a common mechanism supports those tasks. To preview the results, we find that ensemble perception tends to be correlated among high-level features (person identity, emotional expression) and among low-level features (e.g., orientation and color), but not between high-level and low-level features (e.g., identity and orientation). Together these results suggest that ensemble perception is not supported by a single, monolithic statistical processing mechanism, and that the architecture of ensemble representation shows a division between low-level and high-level ensemble processing.

Experiment 1

In the first experiment, the items were either faces that varied continuously in identity, or gabors that varied continuously in orientation (see Figure 1). Participants adjusted a test item to match the average of all the items in the set, or to match a single individual from the set. The objective of these experiments was to (a) determine whether separate ensemble mechanisms exist for different stimulus domains, and (b) determine how ensemble representations relate to individual item representations. Because one of our goals was to compare ensemble representations with individual item representations, it was important to monitor for eye movements. For example, if observers fixated the cued individual item, the representation of the item in the individual task would be viewed in the fovea, whereas most (or all) items in the ensemble task would be viewed peripherally. This discrepancy could artificially decrease any correlation between performance on the individual and ensemble tasks, particularly if different observers adopted different eye-movement strategies. Thus, to enable a direct comparison between individual and ensemble representations, observers were required to maintain fixation for both tasks and eye position was tracked using an eye-tracker. Thus, Experiment 1 was run in the lab with eye-tracking. Subsequent experiments (2 through 8) focused on comparisons between ensemble tasks only, in which eye movements were not an issue, and were run online, where a much larger sample could be recruited.

Method

Participants. Fifty-five affiliates of Harvard University, ages 18 to 35 years, participated in this study. Informed consent was obtained for all participants, who were compensated for their time and had normal or corrected-to-normal vision. This research was approved by Harvard University's institutional review board.

Stimuli and design. Stimuli were oriented gabors and identity face morphs. Gabors were generated using Psychophysics Toolbox (Brainard, 1997) in MATLAB (94% max Michelson contrast, 1.15 c/deg). Faces were 360 linearly interpolated identity morphs, taken from the Harvard Face Database, of three distinct male faces (A-B-C-A; see Figure 1), generated using MorphAge software (version 4.1.3, Creaceed). Face morphs were nominally separated from one another in identity units, which corresponded to steps in the morph space. Prior to morphing, face images were luminance normalized. Both orientation and identity were circular stimulus spaces, with orientation spanning 180° (e.g., leftward and rightward gabors were identical) and face identity spanning 360°. All stimuli in this and future experiments were presented in grayscale.

In a block design, observers viewed either four gabors varying in orientation (always ±5° and ±15° around the mean orientation) or four faces varying in identity (always ±10 and ±30 identity units around the mean identity). These parameters were selected based on pilot studies, with each item in a given set exceeding at least one just noticeable difference. Differences in discriminability among the stimulus sets was not explicitly controlled for, although some of these differences are attributable to the fact that orientation space ranges from 0 to 180°, whereas the face space ranges from 0 to 360°. Each individual item subtended 4.3° × 4.3° of visual angle. Each set contained four items, with one item appearing in each quadrant 3° from fixation. The entire set subtended 8.6° × 8.6°.

Eye-tracking was employed to ensure observers maintained fixation throughout the trial using an Eyelink 1000 (SR Research Ltd.). Trial completion was fixation contingent; when observers broke fixation, the trial was terminated and observers were required to repeat the trial.

Procedure. Observers' task was to report the identity of an individual object, or the average identity of a group of objects. The objects were either faces drawn from the face wheel or oriented gabor patches (see Figure 1). Observers completed four tasks in

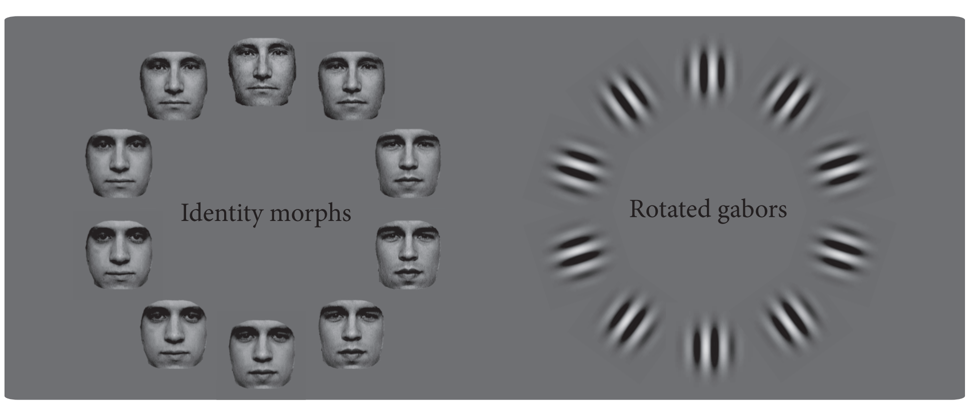

Figure 1. Stimuli consisted of morphed faces or rotated gabors. The faces were morphed continuously between three different individuals to create a circular identity morph space.

separate blocks: judging the orientation of an individual gabor, judging the identity of an individual face, judging the average orientation of a set of four gabors, and judging the average identity of a set of four faces. In the individual tasks, observers were cued to a single item and asked to judge its identity or orientation (see Figure 2). In the ensemble tasks, observers were cued to all four items, and asked to judge the average identity for faces, or the average orientation for gabors. After 1 s, the set was replaced with a single test item at the center of the screen. Observers used the mouse to adjust the test stimulus to match either the individual or the average, depending on the cue. Moving the mouse continuously in one direction altered the appearance of the stimulus, either rotating the gabor or changing the identity of the face morph (although the mouse cursor was invisible). The advantage of the continuous report paradigm employed here is that it provides a direct measure of how accurately the individual or ensemble was perceived.

The order of blocks was counterbalanced across observers using a balanced Latin square, which controls for first-order carryover effects. Because this was an individual differences design, we wanted to minimize performance differences caused by trial order effects, and therefore the trial order within blocks was fixed to a single random order used for each participant.

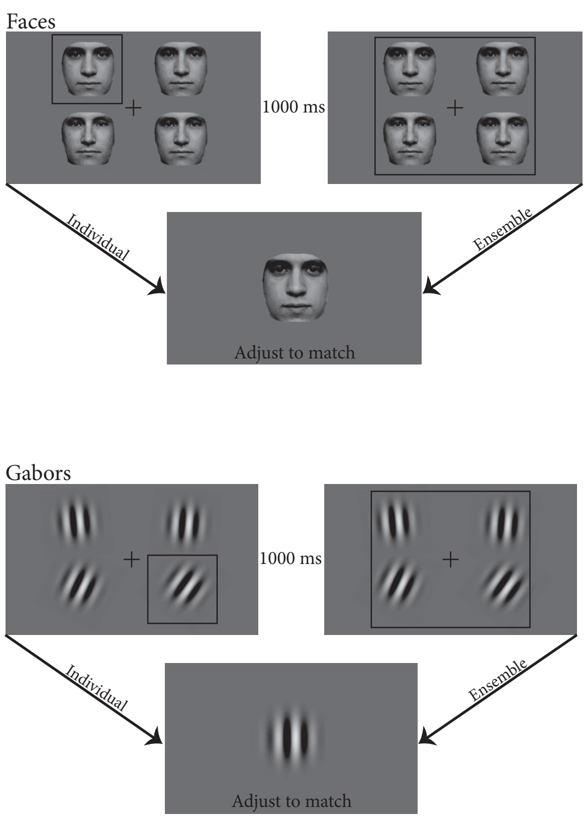

Figure 2. Experiment 1 tasks. Observers were presented with four faces (top) or four gabors (bottom) while they maintained fixation at the center of the screen. In the individual judgment condition (left), observers had to adjust the test face or the test gabor to match the cued individual and ignore the other items. In the ensemble judgment condition (right), all four items were cued, and observers had to adjust the test item to match the perceptual average of all four of these items.

To enable us to compute reliability estimates using Cronbach's alpha, observers saw a total of 20 unique face sets and 20 unique gabor sets 8 times each (4 times in the individual condition, 4 times in the ensemble condition). The sets were equally spaced across the span of the stimulus set.

Before the experiment began, observers performed 32 practice trials (eight in each condition) with feedback to ensure they understood task instructions. Observers then performed 80 test trials in each of the four conditions, for a total of 320 test trials.

Data analysis. Our primary interest is the degree to which individual observers' performance in different tasks correlate with one another. However, the correlation observed between two variables is limited by the reliability with which those variables are measured. In general,

$$r_{\max x,y} = \sqrt{(\alpha_x \times \alpha_y)} \tag{1}$$

where α is the measure of reliability. The maximum observable correlation between x and y is equal to the geometric mean of the reliabilities with which they are measured (i.e., the square root of their product). Thus, reliability places a bound on the maximum observable correlation (Nunnally, 1970), and therefore must be taken into account when comparing correlations across tasks.

We computed reliability for each task using Cronbach's alpha (Cortina, 1993; Cronbach, 1951). Cronbach's alpha is a measure of internal consistency, increasing as the intercorrelations among items (i.e., the 20 unique displays) increases.

We report the observed correlation value for each pair of tasks as well as each tasks' reliability. One could also report the adjusted correlation, which estimates the correlation between tasks x and y by taking their reliability into account (defined as the ratio between the observed correlation and the maximum observable correlation determined by the reliability). However, such adjustments rely on assumptions that are difficult to validate (e.g., accurate estimates of reliability), and also tend to obscure the raw, observed results. We find that our results with the adjusted correlations are qualitatively identical to those with the unadjusted correlations. Thus, in the interest of simplification and transparency, we report the observed correlation values only.

Results and Discussion

For each observer and condition, we calculated the mean absolute error, which can be used as an index of how precisely information was represented. Smaller absolute error suggests a more precise representation. Observers whose performance was 2.5 standard deviations worse than average performance on any task were excluded from analysis, resulting in $N = 47$. Five of the eight excluded had error distributions that did not differ from a uniform distribution, as determined by a modified Rayleigh test (Durand & Greenwood, 1958; Fisher, 1995)—an indication that they were randomly guessing.

Performance was highly reliable for all tasks, as measured by Cronbach's alpha (Cortina, 1993; Cronbach, 1951): individual orientation, $\alpha = .80$; ensemble orientation, $\alpha = .86$; individual face, $\alpha = .76$; ensemble face, $\alpha = .84$. These high Cronbach's alpha values suggest that our displays were internally consistent and that our measures were reliable. These reliabilities place limits on the maximum observable correlations, such that we should not expect to see correlation values above approximately 0.80 (range = 0.78 to 0.85), even if our tasks were tapping an identical underlying mechanism.

Our primary analysis focuses on the correlation in performance across participants for each pair of tasks (see Figure 3). As shown in the bottom panel of Figure 3, there was a strong, significant correlation *within feature domains*, both for orientation (individual orientation task vs. ensemble orientation task, $r = .43$, $p < .01$; 95% confidence interval [CI] [0.17, 0.63]) and for faces (individual face task vs. ensemble face task, $r = .76$, $p < .01$; 95% CI [0.61, 0.86]). In contrast, the top panel of Figure 3 shows that the correlations *across feature domains* (orientation task to face task) were weak both for individual judgments (individual orientation vs. individual face, $r = .04$, $p = .81$; 95% CI [−0.25, 0.31]), and for ensemble judgments (ensemble orientation vs. ensemble face, $r = .05$, $p = .72$; 95% CI [−0.23, 0.34]). Thus, accurately seeing the average facial identity does not predict an observers' ability to accurately see the average orientation.

To verify that the correlations within a feature domain were stronger than the correlations across feature domains, we employed a statistical test that allows comparisons between correlated correlation coefficients (Meng, Rosenthal, & Rubin, 1992). To compare r_{xy} with r_{xz}, it is necessary to take into account the fact that both correlations share one variable. Thus, using this test, we can ask whether the ensemble orientation task is more strongly correlated with the individual orientation task or the ensemble face task. These statistical tests confirm that both within-domain relationships (bottom of Figure 3) were stronger than both across-domain relationships (top of Figure 3; all z scores ≥ 2.17, all p values < 0.05). The lack of relationship across domains suggests the presence of separate, independently operating, ensemble mechanisms for average identity and average orientation.

The remaining relationships for individual orientation versus ensemble face and individual face versus ensemble orientation were not significant ($r = -0.12$ and $r = .14$, respectively). The lack of correlation between these tasks mitigates any concerns that the strong within-feature correlations are explained by general factors, such as intelligence or overall effort. Such factors would have produced significant correlations even for these across-domain, across-task measures.

This experiment provides evidence that the performance on ensemble tasks can be predicted from performance on individual tasks. However, because the present study is correlational, the cause of this relationship is unknown: We cannot conclude that

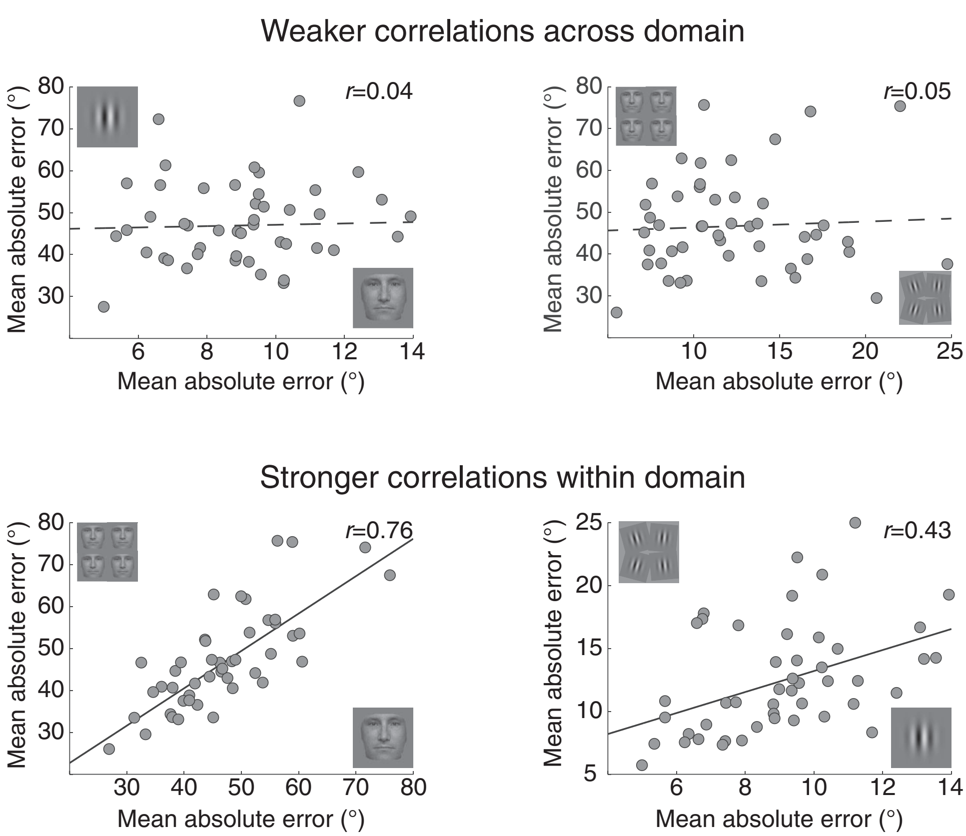

Figure 3. Results of the individual difference correlation analysis. Each point represents an individual observer's performance on the two tasks. Top panel: There was no relationship across feature domains; performance on the individual gabor and individual face tasks was not related across observers, nor was performance on the ensemble gabor orientation and ensemble face identity task. Bottom panel: There was, however, a strong relationship within a feature domain, such that observers' performance on ensemble perception of face identity was related to their performance on individual face identity, and their performance on ensemble gabor orientation was related to their performance on individual gabor orientation.

ensemble judgments are the same as individual item judgments, or that this correlation between individuals and ensembles is solely caused by the precision of individual representations affecting the precision of the ensemble representation. There is ample evidence suggesting that ensemble computations are driven by distinct cognitive processes (Ariely, 2001; Chong & Treisman, 2003), which can even operate under conditions in which individual items cannot be discriminated (i.e., crowding; Fischer & Whitney, 2011; Parkes et al., 2001). Even when the items are discriminable, psychophysical evidence suggests that averaging is an imperfect and inefficient operation (Solomon, 2010; Solomon, Morgan, & Chubb, 2011), above and beyond inefficiency in individual item representation. Indeed, because the present work is correlational in nature, it is impossible to know the directionality of information transfer, and it is likely the case that the ensemble and individual representations are mutually interactive (e.g., (Brady & Alvarez, 2011; de Fockert & Marchant, 2008; Hochstein & Ahissar, 2002), for example, if observers' reports of a particular face were affected by the average of the set, this could induce a correlation between individuals and ensembles. Nevertheless, the correlation between individuals and ensembles in our task does suggest that there are shared limits on individual item processing and ensemble processing.

These data are also consistent with a domain-specific architecture for ensemble representations, at least for stimuli that divide along high- and low-level visual features. The ability to see the average orientation of a set of gabors does not predict the ability to see the average identity of a crowd of faces. Given that there is a host of domains over which ensembles may be extracted, however, the domain-specificity hypothesis requires additional experiments using varied stimuli. In the subsequent experiments, we test a range of visual domains to more fully characterize an ensemble framework.

Experiments 2 Through 8: Multiple Ensemble Domains

To example the relationships among a wide range of ensemble feature domains, we deployed a series of experiments online using Amazon's Mechanical Turk (MTurk). Using MTurk allowed us to recruit a large number of participants, and MTurk users form a representative subset of adults in the United States (Berinsky, Huber, & Lenz, 2012; Buhrmester, Kwang, & Gosling, 2011). Data from MTurk are known to closely match data from the lab on related tasks (e.g., Brady & Alvarez, 2011; Brady & Tenenbaum, 2013).

For each experiment, participants performed two ensemble judgment tasks (e.g., average orientation and average color), and the correlation in performance on those tasks across participants was determined. Of principal interest was whether performance would be uncorrelated for all pairs of tasks, suggesting that ensemble processing is entirely domain-specific, or whether certain pairs of tasks would be highly correlated. Identifying which tasks, if any, tend to be correlated will provide key insight into the cognitive architecture underlying ensemble processing.

Overview of Experiments

The basic task for each experiment was similar to the ensemble task described in Experiment 1: Observers adjusted a test stimulus to match the perceptual average of a group of objects. The individual item adjustment task was not included in these experiments because the relationship between the individual and the ensemble was well characterized in Experiment 1, and because we could not control for eye movements in the online testing environment; without this control, there could be differences in the way in which an individual item from the set was perceived (foveated) in the individual task compared with the ensemble task. Instead, the focus of these experiments was entirely on the relationships among various ensemble representations. These experiments were designed to span the visual hierarchy, offering the most complete picture of the cognitive architecture of ensemble representations to date.

Experiment 2: Identity of individuals versus orientation of gabors. This experiment was a direct replication of the ensemble portion of Experiment 1.

Experiment 3: Emotional expression of faces versus color of dots. This experiment examined another example of two ensemble features (distinct from those used in Experiments 1 and 2) that might be considered bookends in representational space: color and emotional expression. Color is a basic low-level property that is processed early and in parallel across the visual field (e.g., Wolfe, 1994), and has recently been identified as a basic ensemble feature (Maule, Witzel, & Franklin, 2014). By contrast, processing faces is a high-level process that involves holistic, configural processes (e.g., McKone, 2004; Tanaka & Farah, 1993). Whereas Experiment 1 examined processing of the identity of a face, the processing of the emotional expression of a face is generally thought to rely on different systems than those supporting the processing of face identity (e.g., Bruce & Young, 1986), and thus provides another, distinct example of a high-level visual property. In particular, there are different pathways for processing changeable face properties, like expression and eye gaze, and invariant facial properties, like identity (Haxby & Gobbini, 2011), and the processing of emotional expression is known to involve different neural mechanisms than identity perception (e.g., amygdala, insula, striatum; Haxby & Gobbini, 2011). In addition, individuals with prosopagnosia are often impaired at identity perception but not emotion perception (Tranel, Damasio, & Damasio, 1988; Young, Newcombe, de Haan, Small, & Hay, 1993). Thus, Experiment 3 provided a different test of high-level versus low-level ensembles properties.

Experiment 4: Orientation of triangles versus color of triangles. Orientation and color perception are both available as a result of low-level visual processing, with cells selective for both color and orientation present at least as early as primary visual cortex (e.g., Hubel & Livingstone, 1990; Hubel & Wiesel, 1962). In this experiment, we correlated performance on ensemble versions of these two tasks. Triangles were used for both tasks, such that in one block, observers were instructed to judge the average orientation of the triangles, and in another block (identical in terms of stimuli), they were instructed to judge their average color. This design, combined with the designs of Experiments 5 and 6, allows us to test whether the relationships among ensemble representations are dependent upon the visual feature itself (e.g., orientation), the object within which that feature is embedded (e.g., triangles), or some combination of both feature and object properties. Given that both of these features are low level, and are even coded in some of the same cells in primary visual cortex (e.g., Conway, 2001), this experiment provides a particularly strong test of domain specificity.

Experiment 5: Orientation of gabors versus color of dots. This experiment provides another test of the extent to which two kinds of low-level ensemble processing might be related. Here we examine the relationship between orientation and color with two different types of objects providing the basis for the features (gabors vs. dots) rather than both being present on the same object (as with the triangles in Experiment 4).

Experiment 6: Orientation of gabors versus orientation of triangles. In this experiment, observers performed ensemble judgment tasks on identical visual features but with different objects as the carriers of those features (orientation of gabors or triangles). This design, combined with Experiment 4, helps to tease apart whether the visual feature or the object itself determines the relationships among ensemble representations. There is some reason to believe the orientation of a gabor is stored and processed differently than the orientation of a triangle. In particular, the orientation of a gabor is a surface feature (determined by a texture within a circular boundary), whereas the orientation of a triangle is a boundary feature. Boundary representations are likely higher level, as they seem to provide the access point to visual working memory representations (e.g., Alvarez & Cavanagh, 2008), and show more spatial heterogeneity across the visual field than nonboundary judgments (e.g., Afraz, Pashkam, & Cavanagh, 2010), which is a signature of how high-level feature processing is within the visual system.

Experiment 7: Identity of individuals versus emotional expression of faces. In the same way that color and orientation are considered low level, this experiment examined two high-level ensemble feature domains (identity and emotional expression). These feature domains are closely related, as they are both properties of faces, but are supported by ostensibly separate cognitive and neural systems (Bruce & Young, 1986; Haxby & Gobbini, 2011). Finding a lack of correlation between these dimensions would provide strong support for domain-specific hypothesis.

Experiment 8a: Orientation of high spatial frequency gabors versus orientation of low spatial frequency gabors. In this experiment, we compared two highly related ensemble tasks: average orientation of high- and low-spatial-frequency gabors. Given these stimuli are nearly identical, we should observe a high correlation. Thus, this experiment was used to establish a theoretical ceiling on an observable relationship in an online testing environment. Unlike the r_{max}, which is estimated from the reliabilities of each task, this task offers an empirical estimate of the upper bound of r when observers are performing nearly identical tasks and are doing so across two different blocks separated in time.

Experiment 8b: Orientation of gabors versus letter span. In this experiment, we compared an ensemble task with a verbal working memory task that depended only minimally on visual processing and had no ensemble component. These tasks were used to establish a theoretical floor on an observable relationship in an online testing environment. Any relationship observed between these two tasks should reflect only correlations in general factors, such as motivation, general skill at computer usage, working memory capacity, and so forth. By examining this relationship, we can take these general factors into account.

Our letter span task used an adaptive procedure to determine the letter span for each participant, an index of working memory ability that should not depend on ensemble processing or even visual processing (e.g., Logie, Zucco, & Baddeley, 1990).

Method

Participants. For each experiment, 100 participants from the United States were included. We chose 100 participants because that number provides 92% power to detect a correlation of 0.3, a value we considered to be importantly different from the null hypothesis. Participants were excluded from final analysis if their adjustment performance did not differ from a uniform distribution (as described in Experiment 1). Therefore, different numbers of participants were excluded from each experiment and replaced with new participants (ranging from 0 to 35 depending on the task) in order to achieve the desired $N = 100$. Informed consent was obtained for all online volunteers, who were compensated $1.50 for approximately 10 min of their time. We did not prevent participants from participating in more than one ensemble experiment, although fewer than 10 observers overlapped between any given pair of experiments. This research was approved by Harvard University's Institutional Review Board.

Stimuli. Stimuli were gabors (medium-, high-, and low-spatial-frequency versions), faces varying in identity, faces varying in expression, colored isosceles triangles, isosceles oriented triangles (grayscale), and colored dots. Gabors were created in Matlab's Pscyhtoolbox (Brainard, 1997; Pelli, 1997). Stimuli were all 250×250 pixels in size, but actual retinal image size depended on participant viewing distance. Gabors and faces were identical to those described in Experiment 1 (except the low-frequency gabors had a spatial frequency of approximately 0.40 cycles/degree, and the high-frequency gabors had a spatial frequency of approximately 1.8 cycles/degree; these values may have differed slightly depending on a given observer's screen resolution and distance from the screen; however, what matters for our purposes is the relative difference in spatial frequency between the two conditions). The procedure for creating faces varying in expression was the same as for identity, except the source faces were of a single individual displaying either a happy, neutral, or sad expression (Ekman & Friesen, 2003). Each facial expression was separated by 120 linear morph steps, so the face wheel comprised 360 unique "expressive units." All stimulus sets adhered to a circular space.

Variance of the sets was as described in Experiment 1. For orientation judgments (gabors or triangles) and color judgments, the set items were $\pm 5°$ and $\pm 15°$ from the mean. For face judgments (identity or expression), the set items were $\pm 10°$ and $\pm 30°$ from the mean.

Procedure. The task for all experiments, except letter span, was to report the perceptual average of a group of objects. Each participant saw two kinds of objects in a given experiment, as described above. The objects were faces drawn from the "face wheel" (identity or expression), oriented gabor patches, oriented colored triangles, or colored dots drawn from a color wheel. On a given trial, observers were shown four items and asked to judge the perceptual average. The items appeared for 1 s and then disappeared. After 1 s, a single test item at the center of the screen was shown. Observers used the mouse to adjust the object (e.g., the orientation of the gabor or the identity of the face) to match the average of the initial set. Moving the mouse continuously in a circle would alter the appearance of the stimulus (e.g., rotating the gabor or changing the identity of the face morph).

The letter span task differed from the other tasks. In the letter span task, observers initially saw a sequence of three consonants presented sequentially for 500 ms, with a 500-ms interstimulus interval, and then typed those letters after a brief delay (1,000 ms). If they remembered all the letters correctly two trials in a row, they were shown one additional letter on the next trial. If they inserted an extra letter or missed a letter that was presented two trials in a row, they were shown one less letter on the next trial. The mean number of letters they were shown throughout the task then served as our estimate of their letter span.

Observers completed all trials of one kind (e.g., orientation of gabors) before moving on to all trials of the other kind (e.g., identity of faces). The order of each pair of tasks was counterbalanced across individuals. Before an experiment began, observers performed eight practice trials (four in each condition) with feedback to ensure they understood task instructions. Observers then performed 60 test trials in each of the two conditions, for a total of 120 test trials.

Reliability for these tasks was assessed using Cronbach's alpha, as described in Experiment 1. Participants saw three instances of each of the 20 unique ensemble sets (as opposed to four instances, as was the case in Experiment 1). As before, we report the correlation between observers' performance on a given pair of ensemble tasks.

Results

Analyses were carried out as described in Experiment 1. Each participant's average error (i.e., how far off participants were from the true set mean on average) was calculated. The smaller the average error was, the more precise was the ensemble representation. The relationships among each pair of ensemble tasks are depicted in summary form in Figure 4 and Table 1, individually in Figures 5 and 6, and schematically in Figure 7.

Our primary analysis focuses on the correlation in performance across participants for each pair of tasks. The pattern that emerges (see Figure 4) is a near absence of a relationship between high- and low-level ensembles (e.g., average identity and average orientation; average expression and average color), juxtaposed with strong correlations between low-level domains (e.g., color and orientation) and high-level domains (face identity and emotion).

Performance was reliable for all tasks, as measured by Cronbach's alpha (Cortina, 1993; Cronbach, 1951), with all but two values ranging between 0.77 and 0.88 (the exception was for averaging emotional faces and averaging color in Experiment 3, with α of 0.63 and 0.68, respectively, although the very same stimuli in Experiments 5 and 7 elicited higher reliabilities, α of 0.84 for emotional faces and α of 0.89 for color). Although reliability places an upper bound on the observable correlation between two tasks, we also employed a more conservative estimate of this upper bound by empirically deriving the correlation between two highly related tasks (Experiment 8a): Average orientation of low-and high-spatial-frequency gabor sets. This experiment allows us to estimate how large a correlation we could expect to see between two tasks, given that they are limited in reliability and they also occur in different blocks of the experiment several minutes apart. The processing of the average orientation of high-

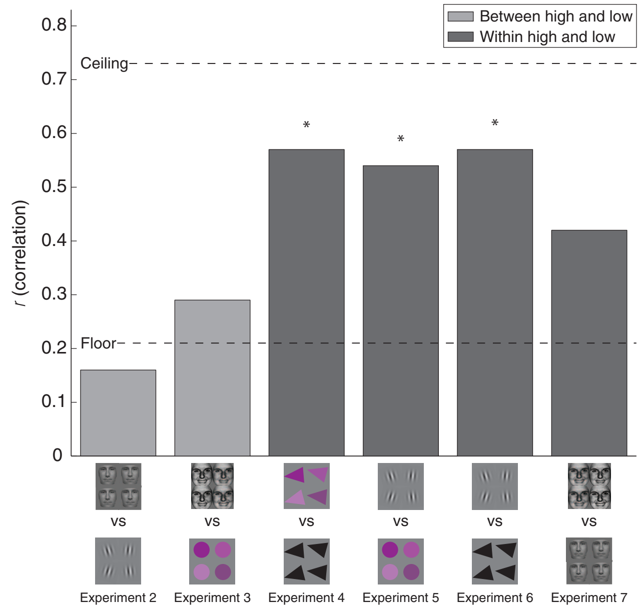

Figure 4. Correlation between observers' performance in each pair of tasks (Experiments 2 to 7). Correlations between high- and low-level features are in light gray (between levels), and correlations between either two low-level or two high-level features are in dark gray (within levels). The dashed line representing the expected ceiling on correlations is based on the result of Experiment 8a, and the line representing the expected floor on correlations is based on the result of Experiment 8b. The between-level correlations were low, and neither was greater than floor. The within-level correlations were higher in general and were significantly greater than floor (with the exception of face-identity vs. face-emotion, in which the comparison with the floor was marginal, $p = .10$; $^{*}\ p < .01$). See the online article for the color version of this figure.

and low-spatial-frequency gabors showed a high correlation ($r = .73$, $p < .0001$; 95% CI [0.62, 0.81]; Figure 5), establishing a reasonable ceiling for how correlated we could expect any two tasks to be. Note that, as expected, this correlation is below the ceiling we would calculate if we only took into account reliability ($r_{\max} = 0.88$).

Whereas Experiment 8a established an empirical ceiling for the correlation between two different tasks, Experiment 8b empirically derived the expected floor. This experiment was designed to establish a floor by estimating the correlation between an ensemble task and a task that does not require ensemble processing or even significant visual processing (letter span). The correlation between average orientation processing and letter span was 0.21 ($p = .04$; 95% CI [0.02, 0.39]). Thus, we can conclude that the expected correlation for two unrelated ensemble processes should be approximately 0.21; that is, if two ensemble tasks do not share any more processing than an ensemble task and a verbal memory task, they should show a correlation of approximately 0.21.

With these bounds in mind, we can examine and contextualize the correlations between the different ensemble tasks (Experiments 2 through 7). We find that the correlation between average identity and average orientation (Experiment 2), stimuli putatively processed by high- and low-level visual mechanisms, respectively, is small ($r = .16$, $p = .11$; 95% CI [−0.04, 0.34]). This value was

Table 1
Descriptive Statistics Across All Experiments

Format	Experiment (Task 1 vs. Task 2)	*r*	Task 1 Mean	Task 1 *SD*	Task 2 Mean	Task 2 *SD*
In-lab expts	Ens. identity vs. Ens. oriented gabor	0.05	46.7	11.7	12.4	4.4
	Ind. identity vs. Ind. oriented gabor	0.04	47.0	10.0	9.1	2.2
	Ind. identity vs. Ens. identity	0.76[a]	47.0	10.0	46.7	11.7
	Ind. oriented gabor vs. Ens. oriented gabor	0.43[a]	9.1	2.2	12.4	4.4
Online expts (all ensemble)	Identity vs. Oriented gabor	0.16	47.6	12.7	15.5	6.5
	Expression vs. Color circle	0.29	34.8	7.1	13.7	4.1
	Oriented triangle vs. Color triangle	0.57[a]	10.5	4.4	14.4	6.0
	Oriented gabor vs. Color circle	0.54[a]	14.3	6.9	14.3	4.7
	Oriented gabor vs. Oriented triangle	0.57[a]	12.7	4.8	9.6	3.2
	Identity vs. Expression	0.42	48.7	13.0	35.6	7.5
	HF oriented gabor vs. LF oriented gabor	0.73[a]	14.1	6.7	12.6	6.4
	Oriented gabor vs. Letter span	0.21	13.2	6.1	6.1	1.0

Note. All units are in 360° space, except orientation units, which are in 180° space, and letter span units, which is the number of letters recalled. expts = experiments; Ens. = ensemble; Ind. = individual; HF = high spatial frequency; LF = low spatial frequency.
[a] Significantly above zero or floor for in-lab and online experiments, respectively.

significantly lower than our expected ceiling on correlations, as determined by comparing Fischer *z*-transformed *r* values ($p < .001$), but was not significantly different than our expected floor ($p = .72$). This replicates the results of Experiment 1, showing the disconnect between ensemble processing of low-level and high-level properties.

In Experiment 3, examining two additional low- and high-level domains (average color vs. average expression), we once again found that the correlation was small (Experiment 3; $r = .29$, $p = .003$; 95% CI [0.10, 0.46]). This value was significantly below our expected ceiling on correlation ($p < .001$), but not significantly greater than our expected floor ($p = .59$). Taken together, Experiments 2 and 3 point to independent mechanisms supporting low- and high-level ensemble representations. Gabor orientation and face identity are no more correlated than an ensemble task and a verbal memory task; similarly, ensemble perception of the average color and average face expression are also no more correlated than the two baseline tasks.

The relationship among ensemble tasks within their respective high- and low-level domains revealed a very different pattern of results (see Figure 4, dark bars). Low-level ensembles clustered together (color, orientation), as did those that may be regarded as high-level (i.e., the face expression task and the person identity task). In particular, Experiment 4 revealed a strong and significant relationship between average color of triangles and average orientation of triangles ($r = .57$, $p < .0001$; 95% CI [0.43, 0.69]), a value significantly above floor ($p = .001$), and marginally significantly different from ceiling ($p = .05$). Both ensemble features appeared on triangles, allowing us to use identical sets of objects for each ensemble judgment. However, this design leaves open the possibility that these two ensemble tasks were related, because the features appeared on the same object type and not because the ensemble

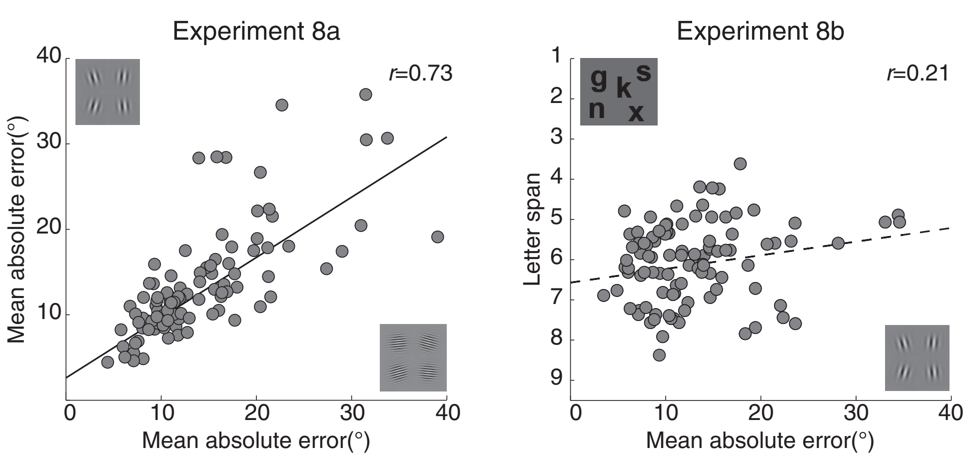

Figure 5. Results of Experiment 8. Each point represents a single observer's performance on the two tasks. The correlation between perception of low- and high-frequency gabors serves as our empirical ceiling (left), whereas the correlation between a verbal memory task and the gabor orientation task serves as our empirical floor (right).

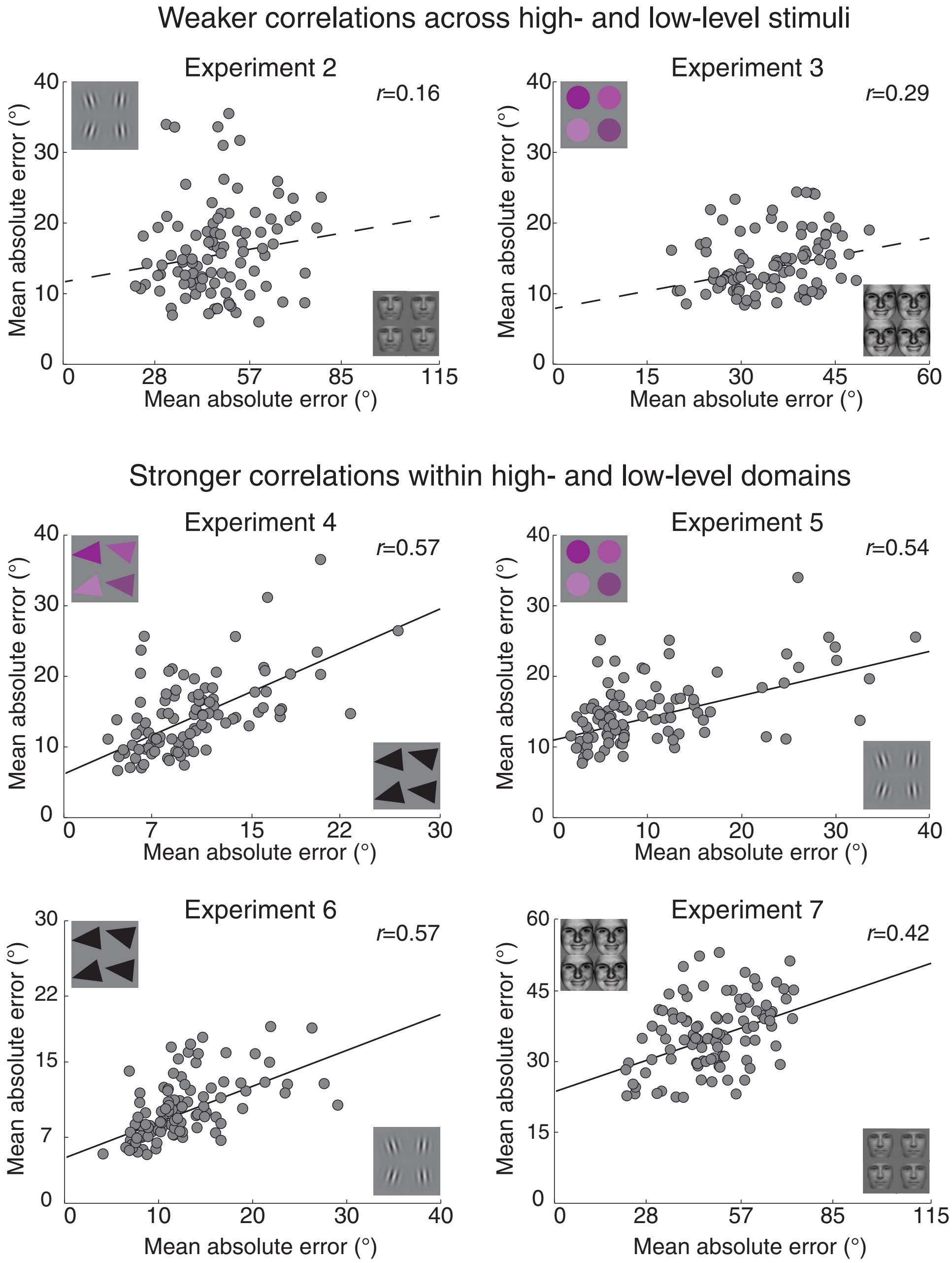

Figure 6. Scatterplots for Experiments 2 to 7. Each point represents a single observer's performance on the two tasks. The correlation values for these experiments are plotted separately in Figure 4. See the online article for the color version of this figure.

representations themselves are related. We addressed this in Experiment 5.

In Experiment 5, we once again tested ensemble representations of color and orientation, but in this design, color appeared on sets of dots and orientation appeared on sets of gabors. Replicating our findings in Experiment 4, we found a significant correlation between color and orientation ($r = .54$, $p < .0001$; 95% CI [0.38, 0.66]), a value significantly above floor ($p = .007$), and significantly below our ceiling value ($p = .02$). These data rule out the idea that the correlation in Experiment 4 was driven by driven by the fact that color and orientation appeared on the same object type (triangles). Rather, this correlation suggests some overlap among low-level ensemble representations in general. It is not the case, however, that seeing the average color is mechanistically identical to seeing the average orientation or that such data prove these two features are processed by an identical and fully shared ensemble mechanism. High correlations are consistent with the idea of a shared mechanism but do not necessitate complete overlap in processing; rather, another possible explanation for a correlation between the two features is a shared source of noise. Although

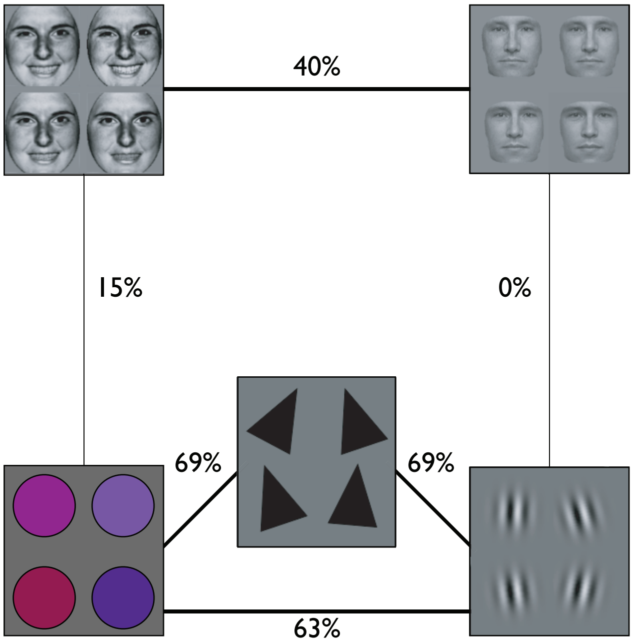

Figure 7. Schematized summary of Experiments 2 to 7. The correlations have been converted to a "percent overlap" score, which normalizes the correlation value to the empirically defined floor and ceiling, that is, $(r - r_{floor})/(r_{ceiling} - r) * 100$, so that values range from 0 to 100. There is reliable overlap within low-level features and high-level features, but little overlap between levels. See the online article for the color version of this figure.

there are clearly well-defined and distinct cognitive mechanisms supporting color and orientation perception (Hubel & Livingstone, 1990; Hubel & Wiesel, 1962; Wolfe, 1994), the current data suggest that ensemble tasks on these stimuli share more variance than, for example, orientation and person identity.

In Experiment 6, we had participants judge average orientation embedded within two different object types: gabors and triangles (see Figure 7). This design allowed us to explore whether stimulus type influences the ensemble representation of a single feature domain (orientation). Consistent with the previous experiments, our results revealed a strong relationship between the average orientation of triangles and the average orientation of gabors ($r = .57$, $p < .0001$; 95% CI [0.42, 0.69]), again significantly above floor ($p < .001$) and marginally significantly different than ceiling ($p = .05$). These data suggest the precision of the representation of the ensemble feature (i.e., average orientation) is consistent across observers regardless of the object on which it appears. This is true despite the fact that boundary features (like triangle orientation) seem to be treated differently by the visual system than surface features (like gabor orientation; e.g., Afraz et al., 2010; Alvarez & Cavanagh, 2008).

High-level ensembles (i.e., the face expression task and the person identity task) also clustered together. In particular, average identity was significantly correlated with average expression (Experiment 7, $r = .42$; $p < .0001$; 95% CI [0.24, 0.57]); significantly lower than ceiling, $p < .001$). Although this correlation was only marginally greater than our floor ($p = .10$), it was significantly greater than the correlation between average identity and the gabor task ($p = .05$). These correlations thus suggest a stronger relationship among high-level ensemble representations than between high-level and low-level ensembles. The greater correlation observed within low-level domains (mean $r = .56$) than within high-level domains ($r = .42$) was not significant (comparing with all within low-level experiments; all $ps > 0.10$). However, the observed correlations do raise the possibility that there might be a stronger relationship within low-level domains than within high-level domains, perhaps related to greater overlap in the regions responsible for primary processing of the low-level stimuli (e.g., primary visual cortex) compared with the two kinds of face information (e.g., identity information in anterior temporal cortex and emotion information in amygdala, insula, and striatum; Haxby & Gobbini, 2011).

The results are summarized in the bar graph in Figure 4. This visualization confirms the disparate nature of high- and low-level ensemble representations. The correlation between ensemble representations across disparate visual levels approach the theoretical floor estimated in Experiment 8a (lower dotted line in Figure 4). In contrast, the correlation between ensemble representations that are closer in representational space are above floor and approach the ceiling estimated in Experiment 8b (upper dotted line in Figure 4). Finally, we schematize these results in Figure 7, normalizing the correlations between floor and ceiling to highlight the relative strength of the relationships between tasks.

General Discussion

To begin to define the functional organization of ensemble perception, we employed an individual differences approach to explore the relationships among various ensemble features. This approach has proven useful for making distinctions between core cognitive processes (Huang et al., 2012; Underwood, 1975; Vogel & Awh, 2008; Wilmer, 2008), with tasks that are uncorrelated suggesting independently operating mechanisms. Our results revealed two examples of independence: between average identity and average orientation, and between average expression and average color, both of which highlight a lack of relationship between high- and low-level ensembles. Interestingly, ensembles tended to cluster within their respective high- and low-level domains, such that putative high-level stimuli (e.g., individual identity and facial expression), as well as low-level stimuli (e.g., color and orientation), showed high correlations. In addition, the results of Experiment 1 revealed a strong relationship between individual item representation and ensemble representation, suggesting that these two processes are mutually dependent.

The independence between high- and low-level ensemble representations offers the strongest evidence to date that ensemble perception is not a monolithic process. The correlation between average identity and average orientation did not significantly differ from our empirically derived correlational floor. The same was true between average expression and average color. This points to the existence of at least one major division in ensemble processing, with little to no relationship between high- and low-level ensemble computations.

The relationships among ensemble features were considerably stronger within their respective high- and low-level visual domains. For example, judging the average identity in a crowd of faces was related to judging the average expression from a different crowd of faces. Likewise, judging the average orientation of a set of gabors was related to judging the average color of a set of dots, and judging the average orientation of triangles was related to judging the average color of those same triangles. These experiments show that the relationship between two ensemble features persists regardless of how it is presented (i.e., on different objects or within the same object). Furthermore, the strong relationship between average orientation of triangles and average orientation of gabors shows that embedding the same ensemble feature (e.g., orientation) on objects belonging to discrete levels within the visual hierarchy (e.g., triangles, which use edge features to carry orientation, and gabors, which use surface features; Afraz et al., 2010; Alvarez & Cavanagh, 2008) does not disrupt the strong relationship between these ensemble judgments.

These results offer at least two possible characterizations of ensemble perception: One possibility is that there are two ensemble mechanisms, one for all types of high-level stimuli and one for all types of low-level stimuli. This is different than a domain-general model because there is more than one mechanism at work, but within the high- and low-level distinction, the mechanism may be agnostic to the visual stimulus. Another possibility is that ensemble representations are entirely domain-specific, and the individual differences approach is not sensitive enough to detect all of the distinctions between different ensemble mechanisms. For example, ensembles that reside close together in some representational space (e.g., color and orientation, both being low-level features) could share more perceptual noise than stimuli that are more representationally distinct (e.g., color and faces). This common source of perceptual noise may introduce correlations that do not reflect mechanistic overlap in the computation of ensemble averages. For example, if observers had a source of shared noise throughout their primary visual cortex, but this was not shared with high-level object cortex, this could induce a correlation between color and orientation judgments but not face judgments. However, although there may exist other, distinct mechanisms of ensemble representation that are difficult to isolate (e.g., separate for every feature; or separate for low-level, high-level, and maybe midlevel surface properties; Nakayama & Shimojo, 1992), the evidence for clearly independent low- and high-level ensemble representations make it unlikely that ensemble representations all arise from a single, monolithic, statistical averaging structure.

Our results additionally show that ensemble perception is linked to individual item perception within a feature domain, but not across feature domains. For example, accurate perception of individual faces predicted accurate perception of the average face, but not the average orientation. It is tempting to conclude that this relationship exists because ensemble representations are computed over sets of individual representations, and thus that ensembles inherit their noise directly from individuals. However, our correlational analysis does not allow us to determine the directionality of the dependence, and so we cannot conclude that individual object representations necessarily feed directly into ensemble representations. Other plausible interpretations include the notion that individual and ensemble representations are computed separately, but are both limited by a common source of perceptual noise, or that the ensemble level actually interacts with the individual level representations. Indeed, there is strong evidence of a hierarchical, bidirectional relationship between items and ensembles (Brady & Alvarez, 2011; de Fockert & Marchant, 2008). Integrated, hierarchical representations would allow for simultaneous activation of individual item information and global ensemble information, perhaps enabling observers to guide attention to critical aspects of the set (Alvarez, 2011), and perhaps to maintain stability in a dynamic perceptual world (Haberman & Whitney, 2012).

Our high-level stimuli only involve judgments about faces. Faces are often granted “special” status as a visual object because of the abundance of behavioral and neuroimaging evidence showing face-specific processing (e.g., Kanwisher, McDermott, & Chun, 1997), and the fact that face processing is particularly susceptible to inversion and configural effects (e.g., Farah, Wilson, Drain, & Tanaka, 1998; McKone, 2004; McKone, Martini, & Nakayama, 2001). Indeed, the “faces are special” proposal is the very reason faces are an ideal high-level stimulus in our experiments: Whether face processing is high level is not in question. However, a reasonable alternative interpretation of our data is that face processing requires certain face-specific computations that introduce noise above and beyond any noise in low-level feature representations, decorrelating the relationship between orientation and identity (indeed, this seems to be confirmed by the individual item data from Experiment 1). Under this view, the seeming independence between high- and low-level ensembles might actually arise at the individual item level, before ever reaching a putative domain-general ensemble processor.

There are several pieces of evidence that speak against this view. First, our data reveal minimal correlations between two separate low- to high-level ensemble tasks (orientation vs. identity averaging and color vs. expression averaging). For this to have resulted from specialized face processing and not separate ensemble systems, one must assume that identity and expression are supported by identical networks, despite the evidence showing some level of independence (e.g., Haxby & Gobbini, 2011; Tranel et al., 1988). Certainly, there is significant overlap between these processes, but they are nonetheless dissociated in both the neuropsychological and the neuroimaging literatures. Second, it is important to consider the ensemble computation itself, specifically, the idea that it is an imperfect, noisy process (Solomon, 2010; Solomon et al., 2011). Under a domain-general view of ensemble representation, any noise associated with the averaging process would be injected into all computations, regardless of stimulus domain. Therefore, the process of extracting the average orientation from a set of gabors would produce noise similar to that created while extracting the average identity from a set of faces, resulting in a significant and observable correlation. The correlations we observe, however, are either not different from zero or not different from an empirically derived correlational floor. In short, a domain-general ensemble mechanism would introduce correlations between low- and high-level ensemble tasks that we do not observe (see Figure 8 for a schematic summary).

Although our data are most consistent with a model composed of multiple, independent ensemble processors (at least two, perhaps more), it is nonetheless important to consider whether there exist other examples of high-level stimuli that might fit into our ensemble framework. An operational definition of high-level could be developed based on sensitivity to inversion and configural effects, invariance across the visual field (Afraz et al., 2010), or long-range spatial interference (Cohen, Rhee, & Alvarez, 2013). Using such an operational definition, future work can establish the level of representation supporting recognition of different types of objects, and can then make a priori predictions for which ensemble representations will be supported by shared mechanisms based on

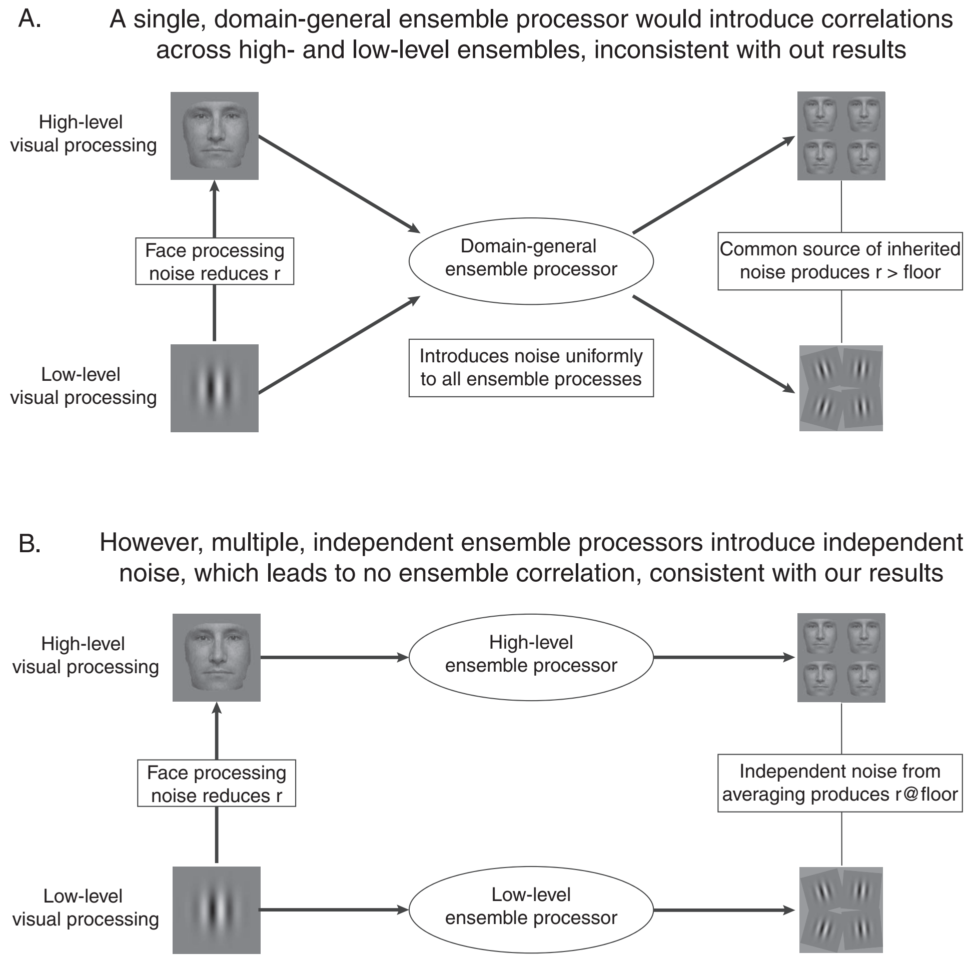

Figure 8. Competing process models for ensemble perception. (A) Even though initial input is uncorrelated (because of independent noise introduced by face processing), a domain-general ensemble mechanism will necessarily introduce observable correlations because of a shared source of noise related to the ensemble process. However, our ensemble correlations are at floor, which can only arise from the existence of (B) multiple ensemble mechanisms. The noise introduced by the ensemble process remains uncorrelated because there is more than one independently operating ensemble system.

this classification. Although this is an exciting direction for further investigation, the present work alone supports the conclusion that ensemble representations are not derived by a single, domain-general ensemble module.

Conclusions

Using an individual differences design, we found that ensemble processing operates independently across high-level and low-level feature domains. Although having a domain-general mechanism would perhaps be more economical, having multiple ensemble mechanisms makes some intuitive sense: An all-encompassing, domain-general ensemble mechanism would have to be remarkably flexible, and able to titrate information from a wide array of stimulus categories ranging from simple oriented gabors to complex faces. If ensemble representations were coarse and imprecise, such a mechanism might be plausible. However, ensemble representations are remarkably precise; thus, they might be "embedded computations," derived somewhat independently across the visual hierarchy. Thus, the present results suggest the possibility that ensemble representation is a canonical computation that operates separately across multiple feature domains represented by the visual system, underscoring the fundamental and ubiquitous nature of ensemble perception.

References

Afraz, A., Pashkam, M. V., & Cavanagh, P. (2010). Spatial heterogeneity in the perception of face and form attributes. *Current Biology*, *20*, 2112–2116. http://dx.doi.org/10.1016/j.cub.2010.11.017

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science*, *21*, 560–567. http://dx.doi.org/10.1177/0956797610363543

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*, 122–131. http://dx.doi.org/10.1016/j.tics.2011.01.003

Alvarez, G. A., & Cavanagh, P. (2008). Visual short-term memory operates more efficiently on boundary features than on surface features. *Perception & Psychophysics, 70,* 346–364. http://dx.doi.org/10.3758/PP.70.2.346

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19,* 392–398. http://dx.doi.org/10.1111/j.1467-9280.2008.02098.x

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12,* 157–162. http://dx.doi.org/10.1111/1467-9280.00327

Attarha, M., Moore, C. M., & Vecera, S. P. (2014). Summary statistics of size: Fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision, 9*(12), 13. http://dx.doi.org/10.1167/9.12.13

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis, 20,* 351–368. http://dx.doi.org/10.1093/pan/mpr057

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science, 22,* 384–392. http://dx.doi.org/10.1177/0956797610397956

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review, 120,* 85–109. http://dx.doi.org/10.1037/a0030779

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77,* 305–327. http://dx.doi.org/10.1111/j.2044-8295.1986.tb02199.x

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3–5. http://dx.doi.org/10.1177/1745691610393980

Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology, 18,* 425–428. http://dx.doi.org/10.1016/j.cub.2008.02.052

Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. *The Journal of Neuroscience, 32,* 7685–7700. http://dx.doi.org/10.1523/JNEUROSCI.3325-11.2012

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43,* 393–404. http://dx.doi.org/10.1016/S0042-6989(02)00596-5

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45,* 891–900. http://dx.doi.org/10.1016/j.visres.2004.10.004

Cohen, M., Rhee, J., & Alvarez, G. (2013). Hemifield effect for high-level, but not low-level, visual stimuli. *Journal of Vision, 13*(19), 643. http://dx.doi.org/10.1167/13.9.643

Conway, B. R. (2001). Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1). *The Journal of Neuroscience, 21,* 2768–2783.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98–104. http://dx.doi.org/10.1037/0021-9010.78.1.98

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334. http://dx.doi.org/10.1007/BF02310555

Dakin, S. C. (1999). Orientation variance as a quantifier of structure in texture. *Spatial Vision, 12,* 1–30. http://dx.doi.org/10.1163/156856899X00012

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37,* 3181–3192. http://dx.doi.org/10.1016/S0042-6989(97)00133-8

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research, 20,* 847–856. http://dx.doi.org/10.1016/0042-6989(80)90065-6

de Fockert, J. W., & Marchant, A. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics, 70,* 789–794.

de Fockert, J. W., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal or Experimental Psychology (2006), 62,* 1–7.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458. http://dx.doi.org/10.1037/0033-295X.96.3.433

Durand, D., & Greenwood, J. A. (1958). Modifications of the Rayleigh Test for uniformity in analysis of two-dimensional orientation data. *The Journal of Geology, 66,* 229–238. http://dx.doi.org/10.1086/626501

Ekman, P., & Friesen, W. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues.* Los Altos, CA: ISHK.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105,* 482–498. http://dx.doi.org/10.1037/0033-295X.105.3.482

Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology, 106,* 1389–1398. http://dx.doi.org/10.1152/jn.00904.2010

Fisher, N. I. (1995). *Statistical analysis of circular data.* New York, NY: Cambridge University Press.

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision, 9*(11), 1. http://dx.doi.org/10.1167/9.11.1

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*(17), R751–R753. http://dx.doi.org/10.1016/j.cub.2007.06.039

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 718–734. http://dx.doi.org/10.1037/a0013899

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics, 72,* 1825–1838. http://dx.doi.org/10.3758/APP.72.7.1825

Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review, 18,* 855–859. http://dx.doi.org/10.3758/s13423-011-0125-6

Haberman, J., & Whitney, D. (2012). Ensemble Perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe & L. Robertson (Eds.), *From perception to consciousness: Searching with Anne Treisman* (pp. 339–349). New York, NY: Oxford University Press.

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science, 17,* 572–576. http://dx.doi.org/10.1111/j.1467-9280.2006.01746.x

Haxby, J. V., & Gobbini, M. I. (2011). Distributed neural systems for face perception. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford handbook of face perception* (pp. 93–107). New York, NY: Oxford University Press.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36,* 791–804. http://dx.doi.org/10.1016/S0896-6273(02)01091-7

Huang, L., Mo, L., & Li, Y. (2012). Measuring the interrelations among multiple paradigms of visual attention: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 414–428. http://dx.doi.org/10.1037/a0026314

Hubel, D. H., & Livingstone, M. S. (1990). Color and contrast sensitivity in the lateral geniculate body and primary visual cortex of the macaque monkey. *The Journal of Neuroscience, 10,* 2223–2237.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of*

Physiology (London), 160, 106–154. http://dx.doi.org/10.1113/jphysiol.1962.sp006837

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17,* 4302–4311.

Leib, A. Y., Puri, A. M., Fischer, J., Bentin, S., Whitney, D., & Robertson, L. (2012). Crowd perception in prosopagnosia. *Neuropsychologia, 50,* 1698–1707. http://dx.doi.org/10.1016/j.neuropsychologia.2012.03.026

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica, 75,* 55–74. http://dx.doi.org/10.1016/0001-6918(90)90066-O

Marčelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America, 70,* 1297–1300. http://dx.doi.org/10.1364/JOSA.70.001297

Maule, J., Witzel, C., & Franklin, A. (2014). Getting the gist of multiple hues: Metric and categorical effects on ensemble perception of hue. *Journal of the Optical Society of America, A, Optics, Image Science, and Vision, 31,* A93–A102. http://dx.doi.org/10.1364/JOSAA.31.000A93

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6,* 255–260. http://dx.doi.org/10.1016/S1364-6613(02)01903-4

McKone, E. (2004). Isolating the special component of face recognition: Peripheral identification and a Mooney face. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 181–197. http://dx.doi.org/10.1037/0278-7393.30.1.181

McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 573–599. http://dx.doi.org/10.1037/0096-1523.27.3.573

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation-coefficients. *Psychological Bulletin, 111,* 172–175. http://dx.doi.org/10.1037/0033-2909.111.1.172

Morgan, M., Chubb, C., & Solomon, J. A. (2008). A "dipper" function for texture discrimination based on orientation variance. *Journal of Vision, 8*(11), 9. http://dx.doi.org/10.1167/8.11.9

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics, 70,* 772–788. http://dx.doi.org/10.3758/PP.70.5.772

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science, 257,* 1357–1363. http://dx.doi.org/10.1126/science.1529336

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition, 128,* 56–63. http://dx.doi.org/10.1016/j.cognition.2013.03.006

Nunnally, J. C. (1970). *Introduction to psychological measurement*. New York, NY: McGraw-Hill.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4,* 739–744. http://dx.doi.org/10.1038/89532

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. http://dx.doi.org/10.1163/156856897X00366

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18,* 849–860. http://dx.doi.org/10.1037/0096-1523.18.3.849

Rensink, R. A. (2004). Visual sensing without seeing. *Psychological Science, 15,* 27–32. http://dx.doi.org/10.1111/j.0963-7214.2004.01501005.x

Šetić, M., Švegar, D., & Domijan, D. (2007). Modelling the statistical processing of visual information. *Neurocomputing, 70,* 1808–1812. http://dx.doi.org/10.1016/j.neucom.2006.10.069

Simons, D. J., & Ambinder, M. S. (2005). Change blindness: Theory and consequences. *Current Directions in Psychological Science, 14,* 44–48. http://dx.doi.org/10.1111/j.0963-7214.2005.00332.x

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision, 10* (14), 19. http://dx.doi.org/10.1167/10.14.19

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision, 11*(12), 13. http://dx.doi.org/10.1167/11.12.13

Sweeney, T. D., Haroz, S., & Whitney, D. (2012). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 329–337. http://dx.doi.org/10.1037/a0028712

Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychological Science, 25,* 1903–1913. http://dx.doi.org/10.1177/0956797614544510

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 46,* 225–245. http://dx.doi.org/10.1080/14640749308401045

Tranel, D., Damasio, A. R., & Damasio, H. (1988). Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. *Neurology, 38,* 690–696. http://dx.doi.org/10.1212/WNL.38.5.690

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist, 30,* 128–134. http://dx.doi.org/10.1037/h0076759

Vogel, E. K., & Awh, E. (2008). How to exploit diversity for scientific gain using individual differences to constrain cognitive theory. *Current Directions in Psychological Science, 17,* 171–176. http://dx.doi.org/10.1111/j.1467-8721.2008.00569.x

Watamaniuk, S. N. (1993). Ideal observer for discrimination of the global direction of dynamic random-dot stimuli. *Journal of the Optical Society of America, A, Optics, Image & Science, 10,* 16–28. http://dx.doi.org/10.1364/JOSAA.10.000016

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research, 32,* 931–941. http://dx.doi.org/10.1016/0042-6989(92)90036-I

Wilmer, J. B. (2008). How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spatial Vision, 21,* 561–579. http://dx.doi.org/10.1163/156856808786451408

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238. http://dx.doi.org/10.3758/BF03200774

Yang, J.-W., Yoon, K. L., Chong, S. C., & Oh, K. J. (2013). Accurate but pathological: Social anxiety and ensemble coding of emotion. *Cognitive Therapy and Research, 37,* 572–578. http://dx.doi.org/10.1007/s10608-012-9500-5

Young, A. W., Newcombe, F., de Haan, E. H. F., Small, M., & Hay, D. C. (1993). Face perception after brain injury: Selective impairments affecting identity and expression. *Brain: A Journal of Neurology, 116,* 941–959. http://dx.doi.org/10.1093/brain/116.4.941

Zhao, J., Ngo, N., McKendrick, R., & Turk-Browne, N. B. (2011). Mutual interference between statistical summary perception and statistical learning. *Psychological Science, 22,* 1212–1219. http://dx.doi.org/10.1177/0956797611419304

Received May 14, 2014
Revision received December 29, 2014
Accepted January 2, 2015 ■