Mixed emotions: Sensitivity to facial variance in a crowd of faces

Jason Haberman

Department of Psychology, Rhodes College, Memphis, TN, USA

Pegan Lee

Department of Dermatopathology, University of California, San Francisco, CA, USA

David Whitney

Department of Psychology, University of California, Berkeley, CA, USA

The visual system automatically represents summary information from crowds of faces, such as the average expression. This is a useful heuristic insofar as it provides critical information about the state of the world, not simply information about the state of one individual. However, the average alone is not sufficient for making decisions about how to respond to a crowd. The variance or heterogeneity of the crowd—the mixture of emotions—conveys information about the *reliability* of the average, essential for determining whether the average can be trusted. Despite its importance, the representation of variance within a crowd of faces has yet to be examined. This is addressed here in three experiments. In the first experiment, observers viewed a sample set of faces that varied in emotion, and then adjusted a subsequent set to match the variance of the sample set. To isolate variance as the summary statistic of interest, the average emotion of both sets was random. Results suggested that observers had information regarding crowd variance. The second experiment verified that this was indeed a uniquely high-level phenomenon, as observers were unable to derive the variance of an inverted set of faces as precisely as an upright set of faces. The third experiment replicated and extended the first two experiments using method-of-constant-stimuli. Together, these results show that the visual system is sensitive to emergent information about the emotional heterogeneity, or ambivalence, in crowds of faces.

Introduction

Humans frequently encounter groups or even crowds of faces. In fact, encounters with groups of faces (seeing more than one face) may be our typical or default experience. Most research on face perception is on the individuals (Maurer, Le Grand, & Mondloch, 2002). However, crowds of faces convey critically important social information. For example, a single fearful face tells us something about the state of that individual; a *crowd* full of fearful faces tells us something about the state of the world (perhaps that we should run). Crowds also convey salient, behaviorally relevant cues. Indeed, when crowds direct their attention to an event, observers are much more likely to share the crowd attention (Milgram, Bickman, & Berkowitz, 1969). Information at the crowd level may strongly influence and guide our behavior. Because crowds convey such important social information, it makes sense that humans are exceedingly sensitive to the average expression in crowds of faces (Haberman & Whitney, 2007). In fact, humans can perceive the average emotional expression, gender, identity, gaze direction, and ethnicity (de Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007, 2009; Leib, Fischer, Liu, Whitney, & Robertson, 2013; Leib et al., 2012; Neumann, Schweinberger, & Burton, 2013; Rhodes, Neumann, Ewing, & Palermo, 2015; Sweeny & Whitney, 2014).

Several studies have found that observers represent the average expression or identity from a crowd, often as precisely as they can recognize the expression of any of the individuals (Haberman & Whitney, 2007, 2009). This crowd expression perception operates over space (e.g., a static crowd of faces) and time—observers are sensitive to crowd expression and identity when faces are presented in a temporal sequence, as happens when we shift our gaze across a crowd, or when a single face dynamically changes expression or speaks (Haberman, Harp, & Whitney, 2009; Post, Haberman, Iwaki, &

Whitney, 2012). Humans are therefore sensitive to the summary statistical information in groups of faces—groups of faces are represented quickly and efficiently as an ensemble (for a review, see Alvarez, 2011; Whitney, Haberman, & Sweeny, 2013). The advantages of this are clear: averaging many local estimates improves precision and accuracy in estimates of global properties or textures (for example, ensemble motion, texture, size; Albrecht & Scholl, 2010; Alvarez & Oliva, 2008; Ariely, 2001; Chong & Treisman, 2003, 2005; Morgan & Glennerster, 1991; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Watamaniuk & Duchon, 1992). Ensemble face representation may be a kind of facial texture coding, where the system can mitigate the effect of local (individual face) noise by averaging over the crowd (Leib et al., 2012). Further, as mentioned previously, crowds of faces convey significant information that is uniquely available at the level of the crowd. Processing the emotional expression of a whole *crowd* of people tells us something more significant, more accurate, and more reliable about the state of the world. Encoding ensemble faces (crowds) allows humans to visually represent social cues at the level of the group.

Although there have been some examinations of sensitivity to set variance (e.g., Marchant, Simons, & de Fockert, 2013; Solomon, 2010), to date, all of the work on ensemble face perception has focused on the sensitivity to summary statistical information about the first central moment—the average expression, gender, and identity (de Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007; Sweeny, Grabowecky, Paller, & Suzuki, 2009). However, if crowd characteristics are represented by specialized mechanisms, one might expect observers to be sensitive to other statistical information in crowds; for example, the second central moment—the variance—in the expression of the crowd. By encoding the variance of expressions in a crowd, the visual system could represent emergent information, such as mixed emotions or ambivalence—characteristics that are not necessarily conveyed by any individual face, but emerge naturally at the level of crowds. Here, we tested whether observers are sensitive to the variance in crowd emotional expression.

Experiment 1

Method

Participants

Three individuals (two females, mean age = 23.6 years) affiliated with the University of California, Davis participated. Informed consent was obtained for all volunteers, who were compensated for their time and had normal or corrected-to-normal vision. All research was approved by UC Davis' Institutional Review Board.

Stimuli

We created three sets of 50 faces by linearly interpolating (Morph 2.5, 1998, Gryphon Software, San Diego, CA) between two emotionally extreme faces of the same person, taken from the Ekman gallery (Ekman & Friesen, 1976). To create the range of morphs, multiple facial features (e.g., corners of the mouth, bridge of the nose, center of the eye, etc.) were matched between the emotionally extreme faces. The morphing software linearly interpolated between the start and endpoints specified, creating 50 image files. The stimulus sets ranged from happy to sad, sad to angry, and angry to happy. The amalgam of 150 faces formed the stimulus set, a virtual "circle" of emotions that was functionally infinite (Figure 1A). Morphed faces were nominally separated from one another by emotional units (e.g., face two was one emotional unit sadder than face one). The label "emotional unit" is arbitrary, and we do not mean to imply that every emotional unit corresponds to a categorically distinct emotion. Although emotion representation is thought to unfold nonlinearly in "emotion space" (Russell, 1980), this would not generate sensitivity to facial variance in crowds, per se, especially at a local scale between nearby facial expressions, where linearity can be assumed. Face images were gray-scaled (98% maximum Michelson contrast) and occupied 3.04 × 4.34 degrees of visual angle.

Sets of either four or 16 faces were presented on a 4 × 4 grid (Figure 1B) such that the central four spots always contained faces. Having two set size conditions allowed us to replicate results and confirm a common prediction of ensemble processing—that performance is relatively unaffected by the number of items in a set (e.g., Ariely, 2001). As such, if variance discrimination functions as other forms of ensemble processing, we expect similar performance in both set size 4 and set size 16. The 16 face sets subtended a total of 12.16 × 13.02 degrees of visual angle. The background relative to the average face had a maximum Michelson contrast of 29%. Each set was characterized by two parameters, the mean (average expression) and the variance (difference in expression between set members). Within the set of four faces, there were four different expressions, centered on a particular average (linearly computed) and with a given variance (separation between face expressions). The set of 16 faces was identical to the four-face set, except that four instances of each face were presented. Biased variance estimates were used, making the variance levels between set sizes 4 and 16 identical.
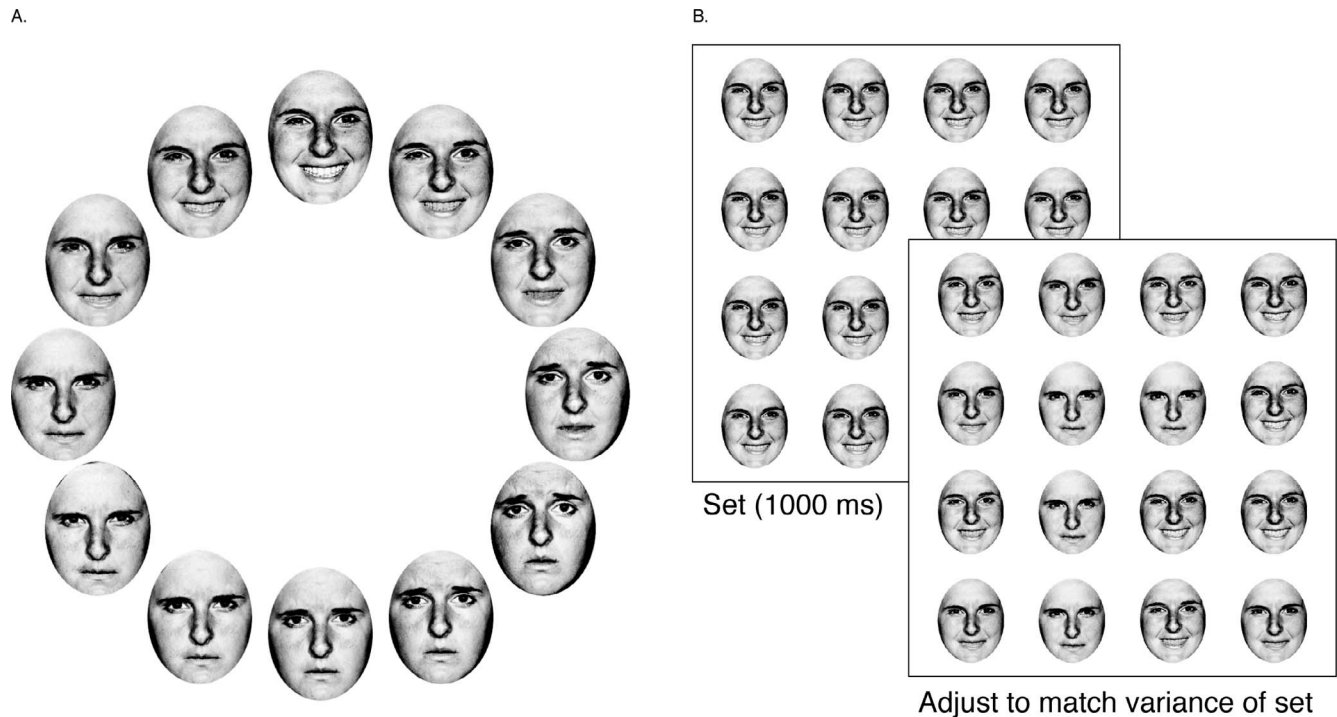
A.

B.

Figure 1. Example stimuli (A) and task (B) in Experiment 1. In the first experiment, a sample set of 16 (or four) faces was presented for 1 s. The average expression of the set was randomly selected on every trial. The faces could be identical (homogeneous expressions) or could vary by one of eight amounts ranging from relatively homogeneous to highly variable. An adjustment set was then presented, with a randomly determined average expression, and a randomly determined amount of variance among the faces. The observer's task was to adjust the *variance* of the set of faces to match the previously viewed sample set. The randomly chosen average expression of the adjustment set did not change during the adjustment phase of the trial; the subject only had control over the variance between expressions (from homogeneous to variable expressions).

Procedure

Observers saw a "sample" set of four or 16 faces for 1000 ms, followed by a second "adjustment" set of faces. The average expressions of the sample and adjustment sets of faces were randomly determined and independent from each other (e.g., the sample set might be relatively happy and the adjustment set somewhat sad). The variance, the parameter of interest, in the first "sample" set was fixed at one of eight possible values (0, 2.5, 5, 11.25, 20, 80, 180, 320).[1] The variance in the second "adjustment" set was randomly assigned, and observers were asked to adjust the variance of the set to match the previously seen sample set by advancing the mouse wheel up or down. The set variance had a range that looped continuously—from homogeneous, the set became more heterogeneous when the mouse wheel was moved in either direction, until the most heterogeneous set was presented, after which continued mouse wheel movement in the same direction made the set became more homogeneous (all faces within a set were altered a fixed amount with every mouse wheel movement). Because the adjustment looped back on itself, there were no clearly detectable "edges" to the adjustment (i.e., the mouse wheel could be moved infinitely up or down without reaching an edge). The adjustment task allowed observers to cycle through the variances (Figure 2) and choose any of 17 degrees of set variance that they thought matched the sample set. Observers pressed the left mouse button to indicate their choice, and the next trial began 500 ms later. Each run had 200 trials, and observers performed from four to six runs (800 to 1,200 total trials).

The method-of-adjustment task allowed us to measure an error distribution that reflects how far observers were from the actual set variance on every trial. Because the distribution is signed (no negative variances), we calculated average error—the deviation of the adjusted variance from the true sample set variance—as an estimate of the observer's sensitivity to the set variance. Higher average error indicates poorer sensitivity. We chose an adjustment because it provides a broad picture of the full error distribution, and does not require an assumption about the underlying discrimination function.

Results and discussion

We measured explicit sensitivity to the variance among facial expressions in groups of faces, asking
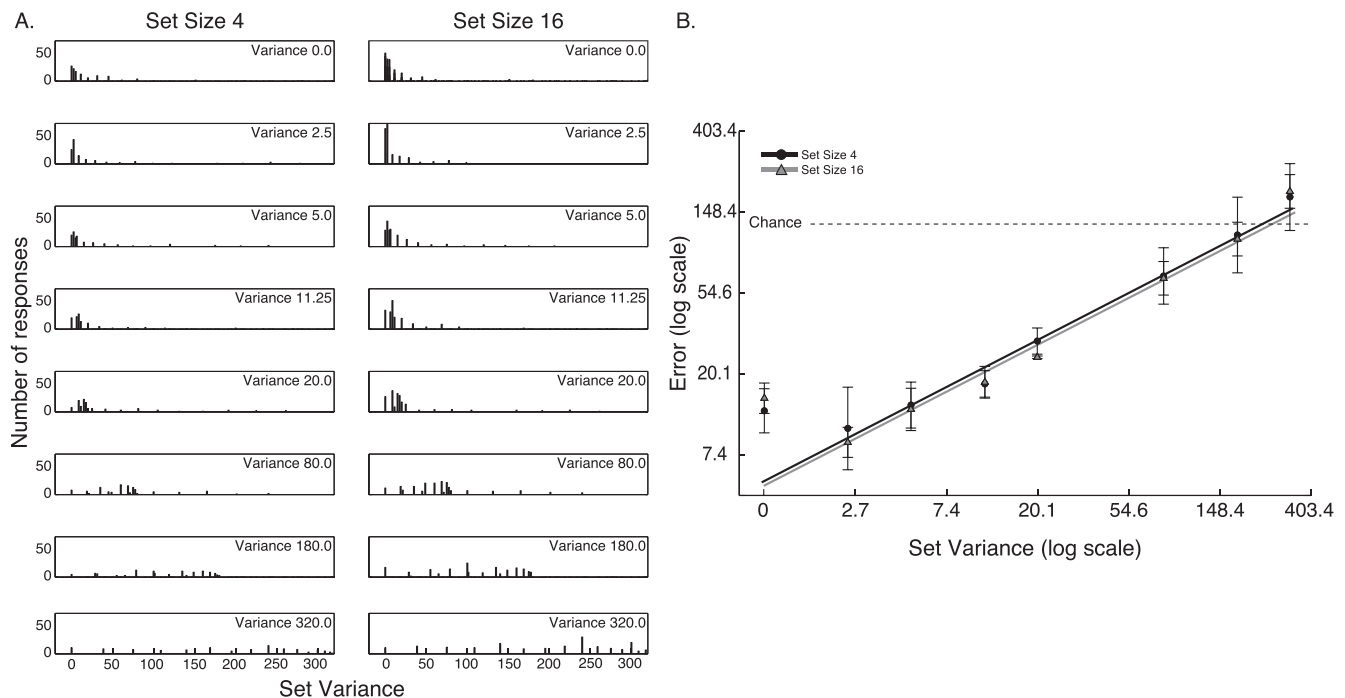
Figure 2. Experiment 1 results. (A) Error distribution for set sizes 4 and 16 at each variance condition, collapsed across observers. Notice that as set variance increases, observer responses become increasingly uniform (i.e., random). (B) The abscissa shows the variance in the sample set, where zero is a homogeneous set of identical faces. The ordinate shows the average deviation of the response in the adjustment set compared to the sample set. These data are modeled by the best fitting line. Chance was determined by 10,000 simulated trials in which random responses were selected. Higher numbers indicate poorer sensitivity. Note that for visualization purposes the variances are shown on a log scale. Error bars are within subject *SEM*.

observers to adjust the variance in an array of faces to match a previously seen sample set. Figure 2A and B shows that observers were able to adjust the variance of the set of facial expressions to match the previously seen set, but only when the variance of the sample set was small. That is, the two sets might have different average expressions (e.g., happy and sad) but observers were able to adjust the variance of the second set to match the first set with relative precision as long as the test set did not contain too much heterogeneity.

Figure 2 shows a clear effect of set variance on the subjects' abilities to match the variance in the set—sample sets of faces with sufficiently high facial expression variance are harder to match. Chance performance was determined by simulating an observer who simply guessed on each of 10,000 trials. Observers appear to be guessing when the test set becomes too heterogeneous (i.e., variance of 180 emotional units or higher). Within-subject ANOVAs, with set size (4 vs. 16) and set variance (eight levels) as within subject factors, confirmed a significant main effect of sample set variance, $F(7, 14) = 53.6$, $p < 0.01$. There was no main effect of set size, and the interaction between variance and set size was not significant ($p > 0.5$ in both cases). Notably, although we used a linear fit to characterize these data, it appears that performance remains fairly similar from a homogeneous set (zero

variance) to a slightly heterogeneous one. This is reminiscent of what has been found in other low-level texture discrimination tasks (Dakin, Bex, Cass, & Watt, 2009; Morgan, Chubb, & Solomon, 2008).

To characterize the discriminability of sample set variance, we fit a line to each set size condition, which fits the data here as well (set size 4: $r^2 = 0.93$, set size 16: $r^2 = 0.89$). Consistent with the ANOVA, there was no significant difference (determined by bootstrapping the parameter estimates) in the slopes between set size 16 and set size 4, which were essentially identical at 0.51, suggesting that variance sensitivity was unaffected by set size. We also explored the interesting possibility that performance remains constant at low levels of heterogeneity. To examine this, we fit a two-phase linear regression to the data (Hinkley, 1969). This model captures texture discrimination functions well since it estimates the "break point" between internal and external noise (i.e., the intersection of the two lines; Yeshurun & Rashal, 2010). Although the model fit the data well (set size 4: $r^2 = 0.97$, set size 16: $r^2 = 0.99$), given the additional parameters we cannot justify using it over a linear model. Overall, these results demonstrate that subjects were able to perceptually match the variance of a set at low levels of heterogeneity, independent of its mean expression, but that this ability broke down at high levels of heterogeneity.
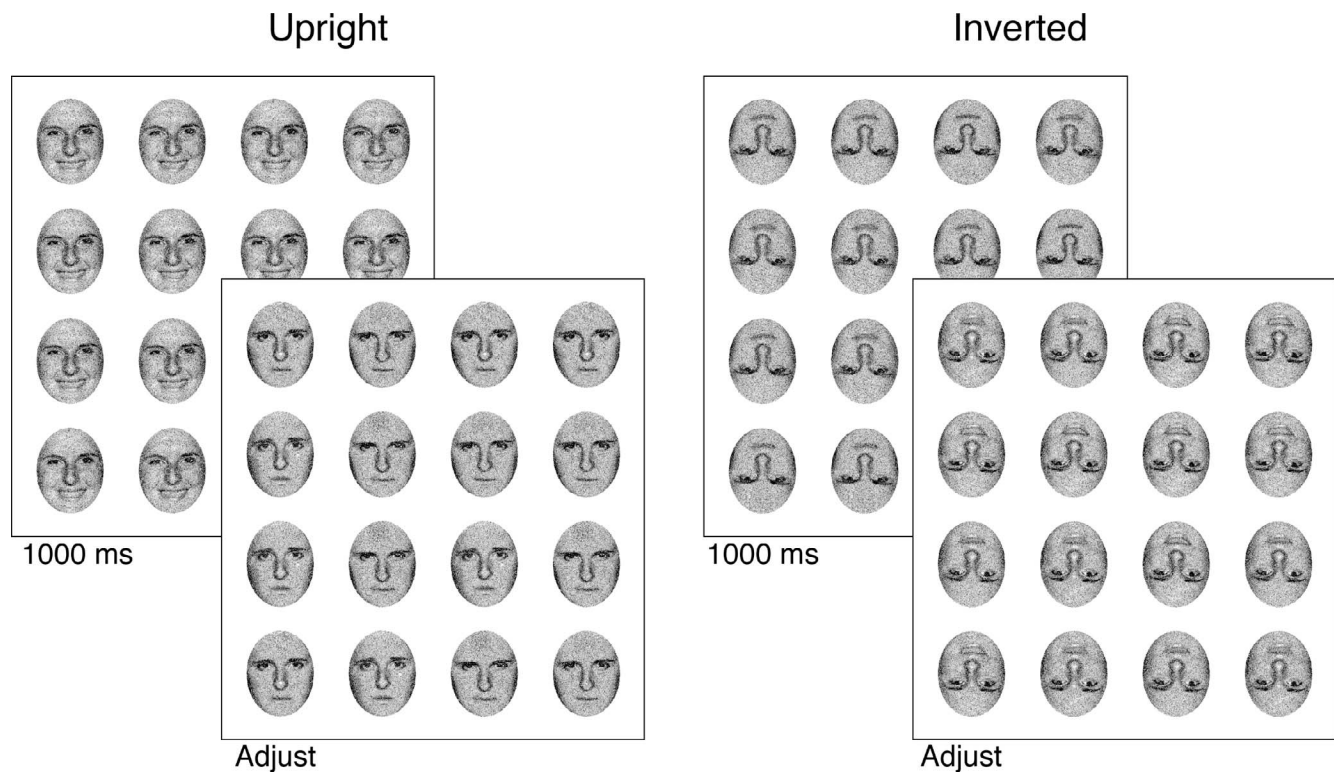
Figure 3. Stimuli used in Experiment 2. Sample sets of upright or inverted faces were presented for 1 s, followed by a set of faces subjects matched to the sample. The average expression in each set was random and independent on each trial.

Experiment 2

Experiment 1 demonstrated that observers are sensitive to variance between expressions in a set of faces. One concern is that observers might have used low-level information such as brightness, orientation, or other non-face specific cues to perform the adjustment task. In the following experiment, we tested for an inversion effect in the set variance sensitivity. If observers used a low-level cue, independent of face orientation, then we would expect the same degree of sensitivity to sets of upright and inverted faces. In fact, the experiment will show that observers are more sensitive to variance in sets of upright than inverted faces.

Method

The methods in Experiment 2 were nearly identical to the first experiment, with the exceptions that the sets of faces (both sample and adjustment sets) were presented in upright or inverted orientations (Figure 3), and that only a set size of 16 was used. Additionally, random noise was added to each face to help control for possible local contrast differences across the set.

Four observers participated in the experiment (three females, mean age = 23 years). The method, design, and task were identical to those in the first experiment.

Results and discussion

Figure 4 shows the results of the second experiment. The precision with which observers matched the variance between upright faces was similar to Experiment 1. Observers were better at adjusting to sets with lower expressive variance, and guessing increased as set variance increased.

A 2 (orientation) $\times$ 8 (set variance) ANOVA confirmed a main significant effect of set variance on discrimination thresholds, $F(7, 21) = 39.0$, $p < 0.01$, a main effect of inversion, $F(1, 3) = 30.9$, $p = 0.01$, and a significant interaction, $F(7, 21) = 6.9$, $p = 0.02$. There is a vertical shift in sensitivity to variance as a function of inversion—higher discrimination thresholds for sets of inverted faces. Slopes of the fitted lines were also significantly different from one another, determined by bootstrapping the slope estimates of each line 10,000 times (set size 4: slope = 0.40, $r^2 = 0.94$; set size 16: slope = 0.20, $r^2 = 0.66$), $p < 0.001$. Therefore, whatever cues present in the inverted faces are not sufficient to account for the (better) peak sensitivity in the upright
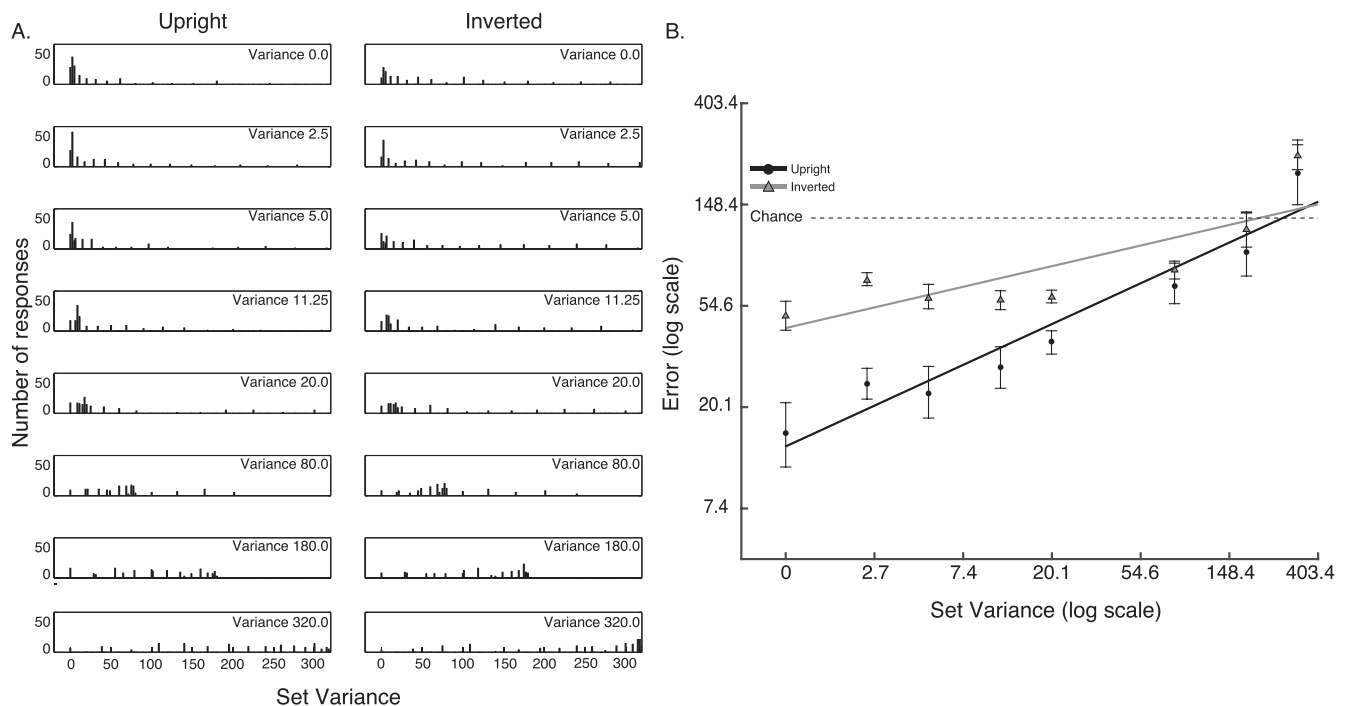
Figure 4. Experiment 2 results. (A) Error distributions (histograms) from the second experiment, for upright and inverted faces. The graph formats are identical to those in Figure 2. (B) Discrimination versus set variance, for upright faces (gray triangles) and inverted faces (black circles). The shapes of the discrimination functions are similar to those in Experiment 1 (Figure 2). There was lower sensitivity to the inverted sets of faces, suggesting that the perception of facial variance is not mediated by low-level nonfacial features. Error bars are within subject *SEM*.

sets of faces. Individual subject analyses were performed on each subject, and confirmed the pattern of results was consistent within each observer.

The inversion effect was not driven entirely by the lower levels of set variance. A follow-up 2 (orientation) $\times$ 3 (three highest levels of variance) ANOVA also revealed a main effect of orientation ($M = 124$ vs. $M = 146$), $F(1, 18) = 4.38$, $p = 0.05$, suggesting that, even at the more difficult-to-discriminate levels of variance, observers were using variance information unique to upright face processing.

Experiment 3

The first two experiments suggested that observers are sensitive to variance between expressions in a set of faces, as measured by an explicit matching task. However, the data do not effectively rule out the possibility that participants are making a binary choice between homogeneous and heterogeneous. Namely, observers may have reported all sets above a specific variance threshold as heterogeneous, without necessarily distinguishing among sets of higher variance. The goal of Experiment 3 was to explicitly test whether

observers can discriminate among different levels of variance. To examine this, we measured set variance discrimination using a method-of-constant-stimuli, two-interval forced choice (2IFC) task. If observers are sensitive to expression variance in sets of faces, it should be measurable and replicable using multiple methods.

Method

The stimuli in Experiment 3 were identical to those used in the first experiment, but the task differed. In this method-of-constant-stimuli experiment, observers viewed two sequential sets of faces. One of the sets, randomly determined, served as a standard or pedestal and had either 0 variance (homogeneous) or a variance of 20 (run in a separate block). The other set had one of seven amounts of variance (always greater than the pedestal), randomly determined on each trial. The variance levels of the sets for the 0 variance pedestal condition were identical to Experiment 1, and the average expression in every set was randomly generated. Therefore, in a single trial, a sequence of two sets could have very different average expressions. Observers were required to pick the interval that contained
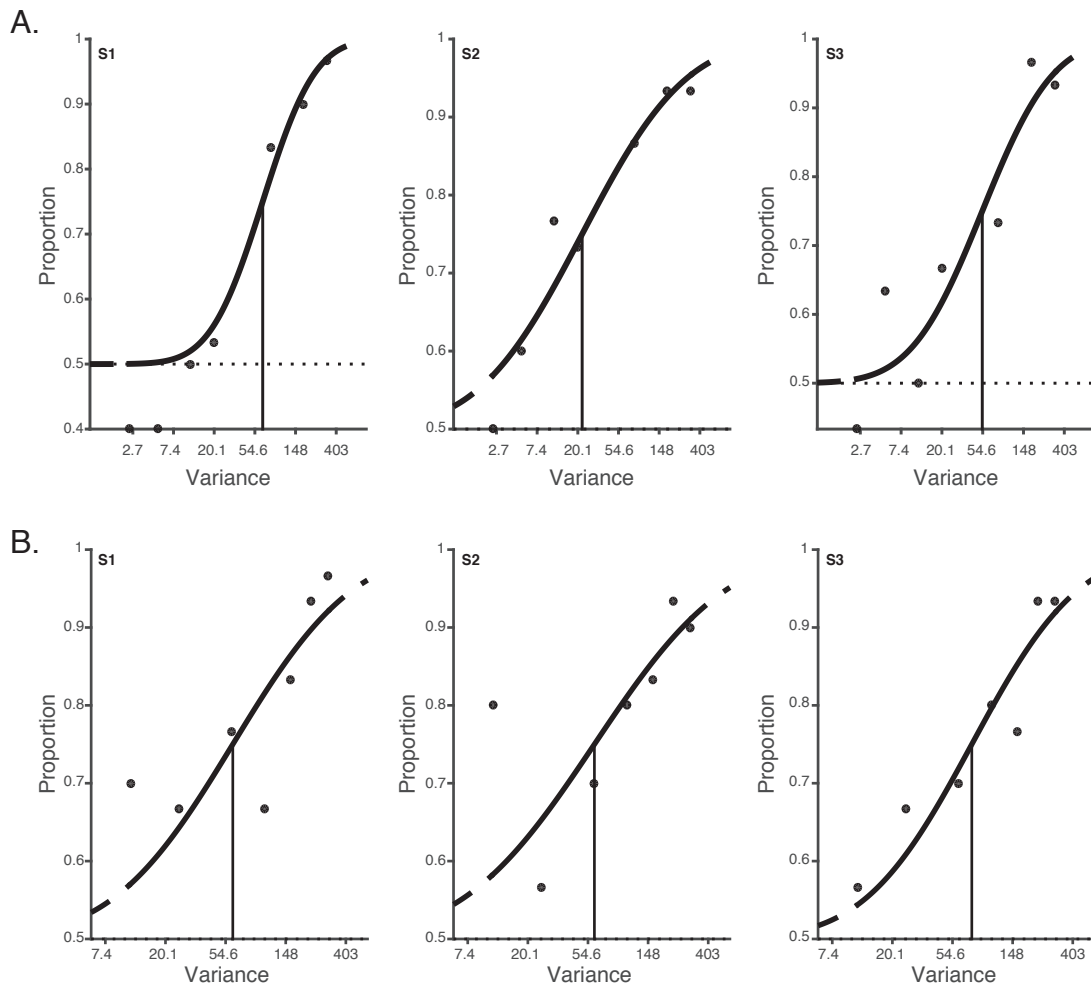
Figure 5. Results of Experiment 3. Subjects performed a 2IFC experiment, discriminating the set of faces with higher variance (A). Psychometric functions for three subjects in the 0 variance pedestal condition (A) and the 20 variance pedestal condition (B).

more expression variance (2IFC task). Three observers from the Rhodes College community (two females, average age: 26.7 years; IRB approved) completed 30 trials at each of the seven levels of variance for a total of 210 trials. In the second condition (run in a separate block), the same task was repeated, but one of the sets had a fixed pedestal variance of 20 rather than 0. In this condition, all nonpedestal sets had an expression variance greater than 20 to measure the increment threshold (and therefore contained different levels of variance from the 0 variance pedestal condition). The proportion of correct responses as a function of the increment in variance between the two sets was plotted and a Weibull function, defined in Psignifit 3 as

$$g(x, m, s) = \frac{2}{\log(2)} ms\Big(\log(x) - \log(m)\Big) + \log\Big(\log(2)\Big),$$

was fit to the data (Fründ, Haenel, & Wichmann, 2011). Parameter x indicates the midpoint (i.e., 75% threshold) on the psychometric function, and parameter m indicates the slope at the midpoint.

Results and discussion

Figure 5 shows the results of the third experiment. Subjects were able to discriminate sets of faces that differed in expression variance. The results confirm the first two experiments and extend them. The 75% discrimination thresholds were 45.9 and 60.3 for pedestals with variance 0 and variance 20, respectively. This roughly corresponds to a set containing faces separated from one another by six more emotional units than the pedestal set. The psychometric functions reveal that subjects are clearly able to perceive and discriminate the expression variance in crowds of faces.

General discussion

The experiments here demonstrated that observers are sensitive to the variance—the mixture of emotion per se—in crowd facial expression. The fact that

observers are sensitive to the average expression, identity, and gender in groups of faces (de Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007; Sweeny et al., 2009) cannot explain the results here. Likewise, the perception of variance in groups of faces cannot be entirely explained by low-level cues such as brightness or orientation, because Experiment 2 demonstrated an inversion effect. The results revealed a discrimination function in which sensitivity scales with set variance.

Ruling out other cues

Observers demonstrate sensitivity to variance across a number of low-level domains (Dakin, 1999; Morgan et al., 2008; Solomon, 2010). Face inversion is often used as a control to isolate face-specific processing (e.g., configural or holistic processing; Maurer et al., 2002; McKone, Martini, & Nakayama, 2001; Yin, 1969). As mentioned previously, the results cannot be explained by low-level cues (e.g., orientation, brightness, etc.), because inverting the faces reduced set variance sensitivity significantly. Low-level cues could, in principle, contribute to the perceived variance in sets of upright (and especially inverted) faces. However, the peak sensitivity achieved for upright faces cannot be accounted for by the information extracted from inverted faces. There must therefore be some degree of sensitivity to variance in sets of upright expressions per se.

The results cannot be explained by nonlinearities in the representation of facial expressions around the wheel of expressions. Although our stimulus set has not been psychophysically linearized, any deviation from perfectly linear psychophysical discriminability would only introduce noise; if anything, this would *reduce* the estimates of facial variance sensitivity.

A strategy in which observers sample just a single item from the set (Myczek & Simons, 2008) is unlikely to account for our results. Numerous studies have already demonstrated that sampling is not the mechanism by which ensemble perception operates (e.g., Ariely, 2008; Corbett & Oriet, 2011; Fischer & Whitney, 2011; Haberman & Whitney, 2010; Parkes et al., 2001). In the case of sensitivity to variance, by definition it must be derived from more than one face in the set, making all better-than-chance performance evidence for an ensemble representation.

Observers were unlikely to access the variance of a set through indirect means, for example by first deriving the average expression and comparing it to the most extreme face. Haberman and Whitney (2010) asked observers to find the most deviant (i.e., extreme) face in a set of emotionally varying faces. Participants were at chance in locating the most

deviant face before the trial ended, suggesting that the representation of any particular individual face was relatively poor. The low-fidelity representation of any given face makes a "compare and contrast" strategy, such as the one described previously, improbable.

Finally, the results cannot be explained by a simple insensitivity to the individual faces. One might argue that if the individual faces in a set were indistinguishable, there would be no difference in sensitivity for homogeneous and slightly heterogeneous sets (performance would look artificially good simply because the individual faces were indiscriminable). This cannot explain the results because the range of expressions in nonhomogenous sets was highly discriminable (i.e., the extreme faces in each set could be easily distinguished; Haberman & Whitney, 2009).

Why sensitivity to facial variance?

Human observers are exceedingly skilled at perceiving the average expression (e.g., average emotion, age, gender, identity) in a group of faces. Why do humans bother coding the facial information about the crowd, in the first place? We judge average crowd expression because it is just like any other ensemble summary statistical percept—it provides more useful information than the individual; it more accurately reflects the true mood (or gender or race) of the individuals who compose the group. Increasing the number of samples increases the accuracy of the estimated crowd (Alvarez, 2011).

What is the value of sensitivity to facial variance? One possibility is that it gives direct access to the "mixture" of emotions in a crowd—the ambivalence of the crowd, for example. Moreover, as mentioned earlier, focused versus wavering crowds can signal important social information not available at the level of any individual: A single angry or aggressive expression in a crowd full of different benign expressions is not surprising. Facing a homogeneous crowd of angry expressions is much more poignant—and would be useful to perceive. This is not simply a matter of judging the average expression. Rather, the sensed variance in the crowd can be used to modulate the value placed on the ensemble estimate. Given the same average expression, a more homogeneous group would signal a much more reliable estimate of the group expression or feeling. Coding expression variance in groups of faces might also be important for searching for outlier expressions. For example, being sensitive to the variance in a set might modulate visual search for deviant facial expressions (Puri, Morris, Haberman, Fischer, & Whitney, 2010).

Facial textures: A uniquely high-level texture processing

An interesting characteristic of the data, especially in Figure 2, is that there is a hint of a flat segment in the error as a function of set variance. This could indicate that discriminating variance in a set of faces is as easy or easier when the set has a nonzero variance. Although thresholds increased slightly in the variance 20 pedestal condition relative to the variance 0 pedestal condition (Experiment 3), this difference did not approach significance, $t(2) = -1.0$, $p = 0.42$. This pattern of variance sensitivity resembles the familiar threshold-versus-contrast (TvC) curve, which has been found in discrimination of several features including orientation variance (Morgan et al., 2008), motion (Burr, 1980), luminance contrast (Schofield & Georgeson, 1999; Watt & Morgan, 1985), and blur (Watt & Morgan, 1985). This type of curve is sometimes interpreted as reflecting the existence of specialized mechanisms (e.g., for encoding orientation variance; Morgan et al., 2008); in our experiments, this flat portion of the function might indicate specialized encoding of facial variance, something worth investigating in the future.

As mentioned earlier, perceiving expression variance in sets of faces may facilitate the perception of average facial expression in groups of faces, something that has clear benefits. Unfortunately, the visual system has significant internal noise—there is noise or error in the estimated facial expression in any given face (in addition to the external noise in the image presented to the observer). There are many levels in visual processing at which noise may be introduced, which could modulate our sensitivity to facial expressions. For example, we are differentially sensitive to faces in different parts of the visual field (Afraz & Cavanagh, 2009), something that could in part reflect noise. Considering that all the encoded faces in a set will be perturbed by some additional internal noise, we might expect that a homogeneous set of faces should not *look* homogeneous.

However, it is reasonable that we might not want to attribute the visual system's internal noise to the outside world, and one method of dealing with this would be to set a threshold for what counts as "variable" (Solomon, 2009). One possibility is that the visual system has an estimate of its own internal noise. Below a threshold, corresponding to the estimated internal noise, everything may look "regular" or homogeneous. It has been argued that this kind of threshold can explain several low-level phenomenon including blur (Burr & Morgan, 1997), contrast (Schofield & Georgeson, 1999), and motion perception (Murakami & Cavanagh, 1998). For example, we are not sensitive to the incessant retinal motion produced by jittering eye movements (which constantly happen); this could be because our threshold for what counts as moving is higher than the amplitude of the retinal motion (Burr & Morgan, 1997; Murakami & Cavanagh, 1998). Although speculative at this point, a similar principle may operate when coding and perceiving crowds of faces; despite some degree of inhomogeneity, crowds may appear homogeneous in expression. Future studies are required to fully address this possibility. In any case, our results show that humans are sensitive to the variance between expressions in sets of faces. The information used to perceive variance in crowd expression is upright-face-specific and could underlie our perception of the homogeneity of crowds.

Keywords: ensemble perception, faces, emotion, heterogeneity

Acknowledgments

Footnote

[1] Variance was calculated based on the emotional unit separation between set items in a biased fashion (i.e., the sum of the squared deviations divided by n, not $n-1$). So, if a set contained faces 1, 2, 4, 5 (note that the mean is excluded from the set), the variance was 2.5.

References

Afraz, A., & Cavanagh, P. (2009). The gender-specific face aftereffect is based in retinotopic not spatiotopic coordinates across several natural image transformations. *Journal of Vision*, 9(10):10, 1–17, doi:10.1167/9.10.10. [PubMed] [Article]

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science*, 21(4), 560–567.

Alvarez, G. A. (2011). Representing multiple objects as

an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131.

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

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception & Psychophysics*, *70*(7), 1325–1326.

Burr. (1980). Motion smear. *Nature*, *284*(5752), 164–165.

Burr, D. C., & Morgan, M. J. (1997). Motion deblurring in human vision. *Proceedings of the Royal Society B: Biological Sciences*, *264*(1380), 431–436.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900.

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica*, *138*(2), 289–301.

Dakin, S. C. (1999). Orientation variance as a quantifier of structure in texture. *Spatial Vision*, *12*(1), 1–30.

Dakin, S. C., Bex, P. J., Cass, J. R., & Watt, R. J. (2009). Dissociable effects of attention and crowding on orientation averaging. *Journal of Vision*, *9*(11):28, 1–16, doi:10.1167/9.11.28. [PubMed] [Article]

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology*, *62*, 1716–1722.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology*, *106*(3), 1389–1398.

Fründ, I., Haenel, N. V., & Wichmann, F. A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision*, *11*(6):16, 1–19, doi:10.1167/11.6.16. [PubMed] [Article]

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision*, *9*(11):1, 1–13, doi:10.1167/9.11.1. [PubMed] [Article]

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), R751–R753.

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734.

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention Perception & Psychophysics*, *72*(7), 1825–1838.

Hinkley, D. V. (1969). Inference about intersection in 2-phase regression. *Biometrika*, *56*(3), 495.

Leib, A. Y., Fischer, J., Liu, Y., Whitney, D., & Robertson, L. (2013). Ensemble crowd perception: A viewpoint invariant mechanism to represent average crowd identity. *Journal of Vision*, *13*(9), 424, doi:10.1167/13.9.424. [Abstract]

Leib, A. Y., Puri, A. M., Fischer, J., Bentin, S., Whitney, D., & Robertson, L. (2012). Crowd perception in prosopagnosia. *Neuropsychologia*, *50*(7), 1698–1707.

Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica*, *142*(2), 245–250.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260.

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*(3), 573–599.

Milgram, S., Bickman, L., & Berkowitz, L. (1969). Note on the drawing power of crowds of different size. *Journal of Personality and Social Psychology*, *13*(2), 79–82, doi:10.1037/h0028070.

Morgan, M., Chubb, C., & Solomon, J. A. (2008). A 'dipper' function for texture discrimination based on orientation variance. *Journal of Vision*, *8*(11):9, 1–8, doi:10.1167/8.11.9. [PubMed] [Article]

Morgan, M. J., & Glennerster, A. (1991). Efficiency of locating centres of dot-clusters by human observers. *Vision Research*, *31*(12), 2075–2083.

Murakami, I., & Cavanagh, P. (1998). A jitter after-effect reveals motion-based stabilization of vision. *Nature*, *395*(6704), 798–801.

Myczek, K., & Simons, D. J. (2008). Better than

average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, *70*(5), 772–788.

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition*, *128*(1), 56–63.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), 739–744.

Post, R., Haberman, J., Iwaki, L., & Whitney, D. (2012). The frozen face effect: Why static photographs may not do you justice. *Frontiers in Psychology, 3,* 22.

Puri, A., Morris, S., Haberman, J., Fischer, J., & Whitney, D. (2010). Effects of high-level ensemble representations on visual search. *Journal of Vision*, *10*(7): 1305, doi:10.1167/10.7.1305. [Abstract]

Rhodes, G., Neumann, M. F., Ewing, L., & Palermo, R. (2015). Reduced set averaging of face identity in children and adolescents with autism. *The Quarterly Journal of Experimental Psychology*, *68*(7), 1391–1403.

Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Schofield, A. J., & Georgeson, M. A. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: Separate mechanisms with similar behaviour. *Vision Research*, *39*(16), 2697–2716.

Solomon, J. A. (2009). The history of dipper functions.

Attention, Perception, & Psychophysics, *71*(3), 435–443.

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, *10*(14):19, 1–16, doi:10.1167/10.14.19. [PubMed] [Article]

Sweeny, T. D., Grabowecky, M., Paller, K., & Suzuki, S. (2009). Within-hemifield perceptual averaging of facial expression s predicted by neural averaging. *Journal of Vision*, *9*(3):2, 1–11, doi:10.1167/9.3.2. [PubMed] [Article]

Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention ensemble perception of a crowd's gaze. *Psychological Science*, *25*(10), 1903–1913.

Watamaniuk, S., & Duchon, A. (1992). The human visual-system averages speed information. *Vision Research*, *32*(5), 931–941.

Watt, R. J., & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research*, *25*(11), 1661–1674.

Whitney, D., Haberman, J., and Sweeny, T. D. (2013). From textures to crowds: Multiple levels of summary statistical perception. In J. S. Werner & L. M. Chalupa (Eds.), *The new visual neurosciences* (pp. 695–710). Cambridge, MA: MIT Press.

Yeshurun, Y., & Rashal, E. (2010). Precueing attention to the target location diminishes crowding and reduces the critical distance. *Journal of Vision*, *10*(10):16, 1–12, doi:10.1167/10.10.16. [PubMed] [Article]

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145.