**Title**
Perceiving crowd attention: ensemble perception of a crowd's gaze.

**Journal**
Psychological science, 25(10)

**ISSN**
0956-7976

**Authors**
Sweeny, Timothy D
Whitney, David

**Publication Date**
2014-10-01

**DOI**
10.1177/0956797614544510

Psychological Science
http://pss.sagepub.com/

---

# Perceiving Crowd Attention: Ensemble Perception of a Crowd's Gaze

Timothy D. Sweeny and David Whitney

*Psychological Science* published online 14 August 2014

DOI: 10.1177/0956797614544510

Research Article

Psychological Science
1–11

DOI: 10.1177/0956797614544510
pss.sagepub.com
SAGE

# Perceiving Crowd Attention: Ensemble Perception of a Crowd's Gaze

**Timothy D. Sweeny**[1] **and David Whitney**[2,3]
[1]Department of Psychology, University of Denver; [2]Department of Psychology, University of California, Berkeley; and [3]Vision Science Group, University of California, Berkeley

## Abstract

In nearly every interpersonal encounter, people readily gather socio-visual cues to guide their behavior. Intriguingly, social information is most effective in directing behavior when it is perceived in crowds. For example, the shared gaze of a crowd is more likely to direct attention than is a single person's gaze. Are people equipped with mechanisms to perceive a crowd's gaze as an ensemble? Here, we provide the first evidence that the visual system extracts a summary representation of a crowd's attention; observers rapidly pooled information from multiple crowd members to perceive the direction of a group's collective gaze. This pooling occurred in high-level stages of visual processing, with gaze perceived as a global-level combination of information from head and pupil rotation. These findings reveal an important and efficient mechanism for assessing crowd gaze, which could underlie the ability to perceive group intentions, orchestrate joint attention, and guide behavior.

**Keywords**
ensemble coding, summary statistical perception, joint attention, eye gaze, social perception

Received 5/17/13; Revision accepted 6/27/14

In nearly every human interaction, the visual system gathers a wealth of social information that people use to understand each other's behaviors and intentions (Allison, Puce, & McCarthy, 2000). This social perception, in turn, can rapidly and automatically guide behavior (e.g., Dimberg, Thunberg, & Elmehed, 2000; Friesen & Kingstone, 1998). Intriguingly, when social information is available in crowds, these reactions are amplified. For example, a crowd's direction of looking more effectively guides one's attention than does an individual's gaze (e.g., Gallup et al., 2012; Milner, Bickman, & Berkowitz, 1969). How do people gain access to social information at the group level, such as a crowd's attention? Can people engage the socio-crowd information at the core of group attention (e.g., Itier & Batty, 2009) all at once, at the level of the collective? Perceiving crowds as an ensemble would enable rapid and efficient engagement with groups of people. Or is a crowd's behavior understood only after an inferential process or complex cognitive deliberation?

We tested these competing hypotheses by determining whether humans use a visual process of summary representation—*ensemble coding*—to perceive the direction in which a crowd is looking. In ensemble coding, information about multiple objects is compressed into a statistic—a singular visual representation of the collective properties of the group (for reviews, see Alvarez, 2011; Whitney, Haberman, & Sweeny, 2014). The visual system's ability to compress information offers many benefits, including increased processing efficiency and reduction of noise, which can enable humans to perceive the gist of a group's appearance with greater speed and precision than would be possible by inspecting each member sequentially (Alvarez, 2011; Sweeny, Haroz, & Whitney, 2013). In fact, with ensemble coding, precise information about individuals is actually lost in favor of the group percept (Ariely, 2001; Haberman & Whitney, 2007). Moreover, ensemble codes need not be drawn from every member in a crowd—averaging across a subset of members provides surprisingly high sensitivity (Dakin, Bex, Cass, & Watt, 2009).

---

**Corresponding Author:**
Timothy D. Sweeny, University of Denver, Department of Psychology, 2155 S. Race St., Frontier Hall, Denver, CO 80210
E-mail: timothy.sweeny@du.edu

If efficient and rapid perception of crowd attention is important for behavior, then collective gaze information should be represented and experienced through ensemble coding. In the experiments reported here, we asked what occurs when people see many faces briefly and at once: Do they determine the collective gaze of the crowd by rapidly pooling information from many faces (i.e., ensemble coding), or do they instead make coarse judgments about the crowd based on a single face, as would be expected if understanding crowd gaze is a relatively slow cognitive process (e.g., Myczek & Simons, 2008)? We studied gaze cues because, unlike other ensemble-coded information, they uniquely amplify a person's reaction when seen in groups of more than two (Gallup et al., 2012; Milner et al., 1969). Perceiving gaze is also vital for joint attention and social interaction (Baron-Cohen, 1995; Driver et al., 1999; Friesen & Kingstone, 1998; Itier & Batty, 2009). Finding that a crowd's gaze is represented as a summary statistic would thus provide an important insight into perceptual mechanisms that may contribute to several important social behaviors.

## Experiment 1: Perception of Crowd Gaze With Upright Faces

In our first experiment, observers viewed crowds of computer-generated faces and estimated where each group was looking, on average. Each crowd had an average direction of looking (ranging from leftward to rightward), and, most important, each crowd member had a unique gaze direction. Each crowd was shown for only 1 s, and observers were encouraged not to look directly at the faces. On some trials, observers viewed the full crowd of four faces and estimated their average gaze. On other trials, full crowds were generated, but observers viewed only a subset of the faces from the full crowd and estimated the average gaze of this subset. Whether observers viewed all four faces or a subset of these faces, we recorded the average gaze of the full set of four for subsequent comparisons. We hypothesized that if observers use an ensemble code to perceive crowd gaze, then gaze estimates of subsets should approach the actual gaze of the full crowd when more faces are included in the subset. That is, when more gaze information is available to observers, even in a short period of time, they will use it. Alternatively, if ensemble representation is not engaged for perceiving gaze, then observers should base estimates of a crowd's gaze on a single person's direction of looking rather than on an ensemble code, and their gaze estimates should not change even when more faces from the full crowd are visible in the subset.

### *Method*

***Observers.*** Eight psychophysical observers gave informed consent to participate. We used this number as our sample size and our stopping rule because in a previous investigation with a nearly identical design and number of trials, we had sufficient power to detect and replicate a similar effect with a different dependent variable (Sweeny et al., 2013). All observers had normal or corrected-to-normal visual acuity.

***Stimuli.*** We manipulated gaze using a striking visual interaction (Wollaston, 1824), in which the direction toward which a person appears to be looking is determined by integrating local pupil information with the rotation of the head (Anstis, Mayhew, & Morley, 1969; Gibson & Pick, 1963). For example, Figure 1 illustrates how two separate pairs of pupils gazing in identical directions will appear to have leftward or rightward gazes depending on whether they are superimposed onto heads with subtle leftward or rightward rotations, respectively (Cline, 1967; Langton, Honeyman, & Tessler, 2004). We created a set of 16 computer-generated faces by independently manipulating head rotation and pupil rotation (FaceGen Modeller, Version 3.5.5; Singular Inversions, Toronto, Ontario, Canada). First, we created heads with −8°, −4°, +4°, and +8° horizontal rotations (turned toward the observer's left or right, respectively). Next, we used a head with a straightforward rotation (0°) to generate pupils with −15°, −5°, +5°, and +15° rotations around a vertical axis. We then used Adobe Photoshop (Creative Suite 5, Version 12.0) to merge each pair of these rotated pupils (and the surrounding eye contours) with each rotated head. Combining the four head rotations with the four pupil rotations produced 16 faces with unique gazes. This procedure of merging identical pupil rotations with differing head rotations was crucial because it allowed us to manipulate perceived gaze without changing local pupil information. Put another way, each face's perceived gaze was unique and was the result of a global-level interaction between head rotation and pupil rotation.

We schematized the faces using a four-step process in Photoshop. First, we eliminated the contour of the head and chin. Next, we applied a high-pass filter with a 4-pixel radius. Then, we applied a threshold to the image (at a level of 120 in the thresholding tool), rendering pixels either black or white. Last, we applied a Gaussian blur with a 0.4-pixel radius. This procedure eliminated shading information, equated all faces in terms of low-level visual information, and ensured that only geometric information conveyed rotation. Each face subtended 2.56° × 2.31° of visual angle.

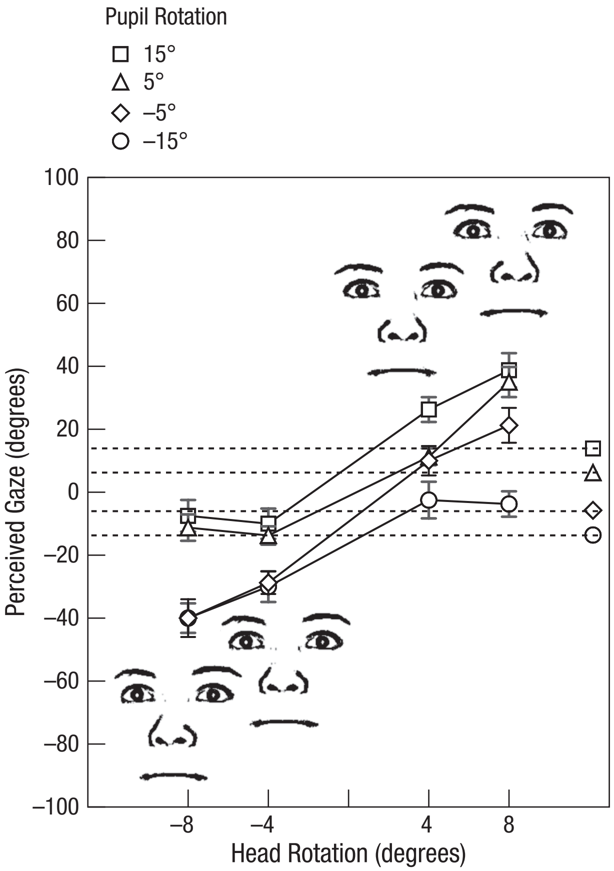

**Fig. 1.** Results from a norming experiment conducted prior to Experiment 1 to measure the perceived gazes that resulted from viewing 16 unique combinations of head and pupil rotations. The faces pictured here have identical pupil rotations (–5°, slightly leftward), but the differing leftward and rightward head rotations (–8°, –4°, +4°, and +8°) make their gazes appear heterogeneous. The graph shows mean perceived gaze across all 16 stimuli as a function of the faces' head rotation and pupil rotation. The dashed lines depict the null results—flat lines—that would have occurred if head rotations had no effect on perceived gaze. We utilized this visual interaction to ensure that the variability in each crowd's gaze was the result of global-level integration of head and pupil rotations and not just an analysis of head or pupil position alone. Note that the faces in this figure constitute just one of the crowds that observers viewed. Other crowds had the same head rotations combined with different pupil rotations.

***Preliminary norming experiment.*** The images in Figure 1 illustrate how combining head and pupil information can change the apparent direction of a person's gaze. But in order to proceed with Experiment 1, we first had to precisely measure the perceived gazes that resulted from each of the 16 combinations of head and pupil rotations. We thus conducted a norming experiment with a separate group of 8 observers. Each observer viewed each of the 16 head-pupil combinations on a test face presented on the top half of a screen. Observers were told to imagine that the test face was looking out toward a point in space and to adjust the pupil position on a response face with straightforward features (the head had 0° of horizontal rotation) so that its direction of gaze appeared to match that of the test face. The response face was presented simultaneously on the bottom half of the screen. Only the pupil positions of the response face could be rotated, in 10° increments between −95° and +95°. The starting pupil position on the response face was randomly selected on each trial from a uniform distribution between −95° and +95°. The response face remained on the screen until the observer pressed the space bar. Observers were allowed to look at both of the faces, although they were instructed to fixate only the bridge of the nose, and they had an unlimited amount of time to respond. We recorded the average pupil position on the response face (e.g., −5°, looking slightly toward the observer's left) as the perceived gaze direction for each of the 16 head-pupil combinations.

As expected, head rotations were effective in modulating perceived gaze direction (Fig. 1). This was confirmed by a main effect of head rotation in a repeated measures analysis of variance (ANOVA; four head rotations × four pupil rotations), $F(3, 21) = 30.628$, $p < .001$, $\eta_p^2 = .813$. On average across all head-pupil combinations, 1° of head rotation pulled perceived gaze by 2.97°. That is, even if the pupils stayed in a fixed position, head rotations strongly pulled the apparent gaze direction. We used the average norming values for each head-pupil combination to obtain the average gaze directions (calculated as the linear means) of the subsets and full crowds of faces in the main experiment.

***Crowds.*** By independently varying head rotation and pupil rotation, we ensured that the gaze of each crowd member, and also the collective gaze of each crowd, appeared to be unique (Fig. 2). No two heads in any crowd faced exactly the same direction; every crowd contained a head with one of four distinct rotations (−8°, −4°, +4°, and +8°) on each trial. Every pair of pupils in a crowd gazed in exactly the same direction on a given trial (e.g., −5°, as in Fig. 1, or +15°, as in Fig. 2c), but across trials, the crowd's pupils could have four different rotations (−15°, −5°, +5°, and +15°). These combinations ensured that low-level pupil information did not vary across each member of the crowd. Any perceived variability across the crowd members' gazes was the result of the Wollaston interaction and occurred at a stage of visual processing in which head and pupil information are globally integrated. Given the values obtained from our norming procedure, these combinations produced a gaze range of 47.5° in our crowds, on average. The center of each possible face location in the crowd (upper left, upper right, bottom left, bottom right) was 3.21° diagonally from the fixation point.

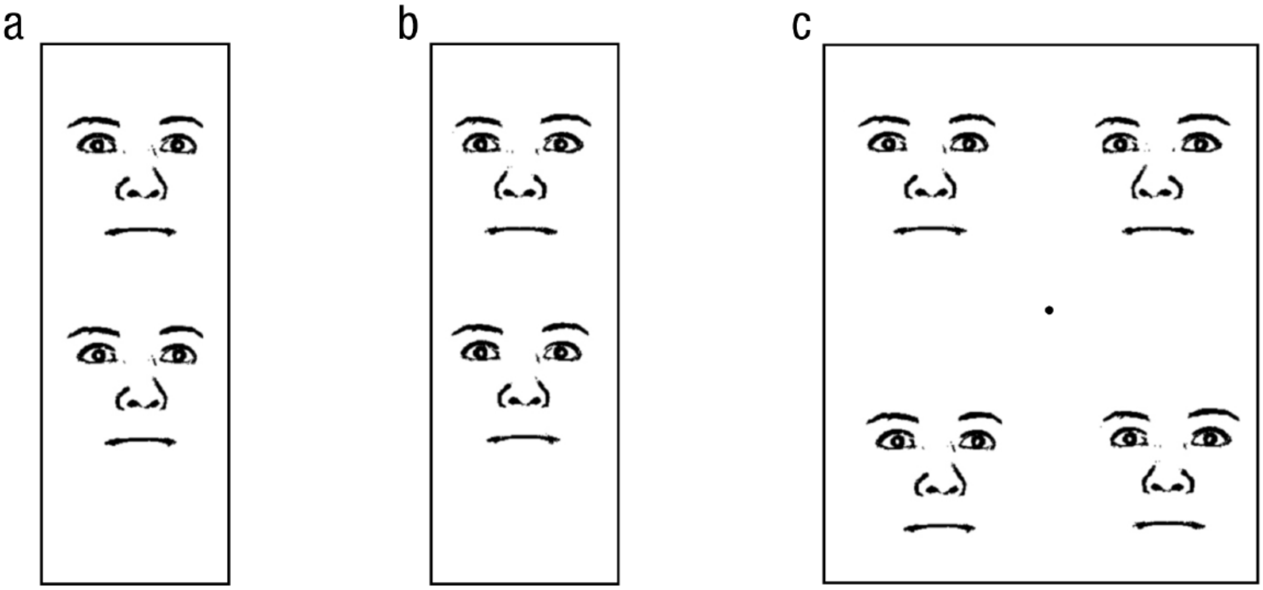

**Fig. 2.** Illustration showing that the perceived gazes of the stimuli used in the present experiments did not rely on head or pupil rotations alone, but instead relied on global integration of information across multiple features. As shown in (a), two faces with identical head rotations and different pupil rotations appear to have different gaze directions. As shown in (b), two faces with identical pupil rotations and different head rotations also appear to have unique gaze directions. The image in (c) shows one of the crowds from the experiments, drawn to scale. The black rectangle was not present in the experiments.

***Procedure.*** Observers were individually tested in a dimly lit room. They initiated each trial by pressing the space bar, whereupon a white screen with a central fixation point immediately appeared for a randomly selected duration of 500, 700, or 900 ms. Next, a randomly selected subset of faces (one, two, or three) or the full set of four appeared for 1,000 ms, followed by a white screen for 1,000 ms. The fixation point remained visible during both of these intervals. Observers were instructed never to look directly at any of the faces in the crowd. Next, a blank white screen was shown for 300 ms, followed by a single response face at the center of the screen. Observers indicated the subset's (or full set's) average direction of gaze by adjusting the pupils on the response face using the left and right arrows on a keypad. The response face always had straightforward features (the head had 0° of horizontal rotation); only the pupil positions could be rotated, in 10° increments between −95° and +95°. The starting pupil position on the response face was randomly selected on each trial from a uniform distribution between −95° and +95°. The response face remained on the screen until the observer pressed the space bar. A given set of faces (a subset of one, two, or three faces or the full set of four faces) was paired with each of the four pupil rotations six times. This yielded a total of 96 trials run across two blocks. All stimuli were presented on a 20-in. CRT monitor while observers' heads were secured in a chin rest at a viewing distance of 47 cm.

## *Results*

We used our norming data to determine the perceived gaze direction (the linear average) of both the full crowd and the subset on each trial. Then, we calculated the difference between observers' estimates and both of these values on each trial. Finally, for each observer, we calculated the variance across these difference scores as our dependent variable. We used a nearly identical approach in a previous investigation of crowd perception (Sweeny et al., 2013).

An ideal-observer analysis illustrates the different patterns of results that would occur from integrating gaze information from different numbers of faces in each crowd or from guessing (Fig. 3). Note that the purpose of the ideal-observer analysis is to illustrate these patterns and facilitate understanding of the empirical data. It is not intended to provide estimates of the exact numbers of gazes integrated or the amount of noise in any stage of the averaging process. Details of the ideal-observer analysis can be found in the Supplemental Material available online, but to summarize, this conservative simulation included the following steps. First, we randomly generated a crowd with different head rotations and gazes. Next, we added early-stage noise to the gaze of each crowd member. We then obtained the linear mean of each noise-perturbed subset (one, two, or three randomly selected faces) or the full crowd of gazes. Next, we perturbed this mean with late-stage noise. Finally, we

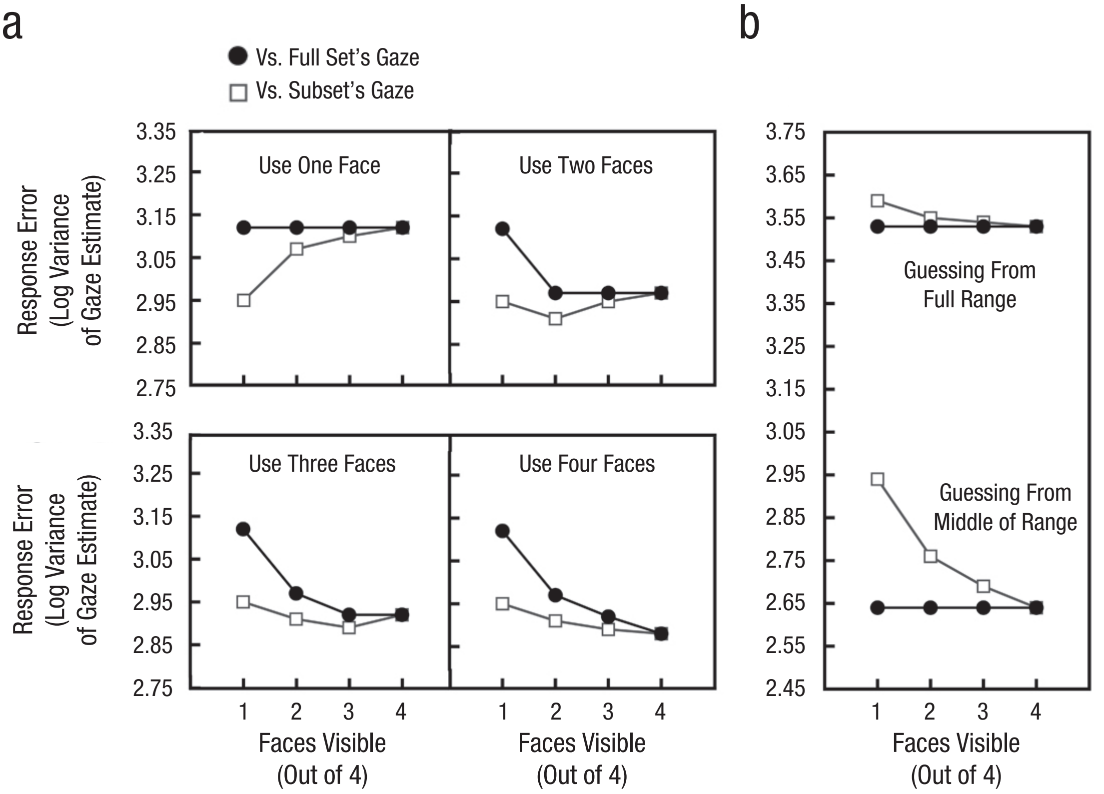

**Fig. 3.** Results of an ideal-observer analysis conducted to show the general patterns of results that would occur from (a) integrating the gazes of one, two, three, or four faces in a crowd or (b) simply guessing and making a random response from the entire range of response options and the middle third of the response range. In (a), each graph shows results for a different maximum number of faces that the ideal observer is using. For all panels, the simulated variance in errors of gaze estimates was computed against both the actual gaze direction for the full set and the actual gaze direction for the subset.

recorded the difference between this simulated gaze estimate and the actual gazes of the full set and the subsets. Calculating the variance of these differences across thousands of trials allowed us to visualize the patterns of response error that would emerge with different amounts of integration. The parameter values in this simulation were chosen simply because they produced a baseline level of performance roughly consistent with our results. Changing these parameter estimates produces quantitative, but not qualitative, changes to these patterns.

If, in a brief glimpse, observers can extract information from only a single face's gaze to estimate where a crowd is looking, errors against the full crowd's gaze should remain the same even when more faces from the crowd are visible (black circles, top left panel of Fig. 3a). At the same time, errors against the subset's gaze (open squares, top left panel of Fig. 3a) should increase because as the number of visible faces becomes larger, the single gaze observers use to make their estimate will become less representative of the subset's direction of looking. A qualitatively different pattern will emerge if observers integrate multiple gazes (e.g., three; bottom left panel of Fig. 3a) into an ensemble code. Errors against the full crowd's gaze should decrease when more faces from the crowd are visible. Errors against the subset's gaze should also decrease, but only slightly—the result of redundancy gain from averaging noisy signals. This pattern will occur only if observers integrate multiple faces, and it cannot occur if observers always respond randomly (Fig. 3b). Note that when the number of visible crowd members (e.g., two) exceeds integration capacity, error relative to the full set's gaze plateaus, and error relative to the subset's gaze increases. These qualitative predictions from our ideal-observer analysis clearly illustrate the patterns of results that we might expect to obtain from our actual observers, which we now describe.

Gaze estimates from larger subsets were closer to the gaze of the full crowd, $F(3, 21) = 9.59$, $p < .01$, $\eta_p^2 = .578$ (Fig. 4a). This pattern indicates that observers integrated gaze information from multiple faces in a crowd. Measured against the subset's gaze, response errors also tended to decrease with larger subset sizes, although this trend was not significant, $F(3, 21) = 1.07$, n.s. (Fig. 4a). The interaction between crowd size and comparison

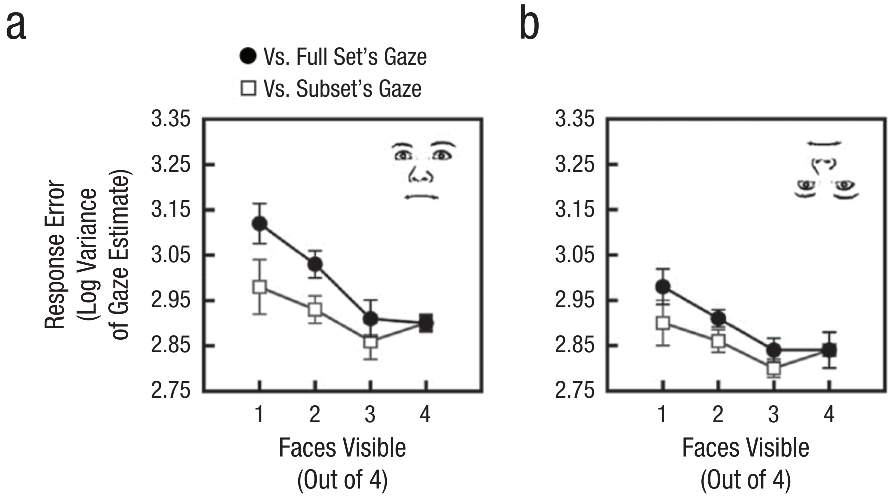

**Fig. 4.** Ensemble perception of a crowd's collective gaze when faces were (a) upright in Experiment 1 and (b) inverted in Experiment 2. For each experiment, the mean response error is shown as a function of the number of faces that were visible on a given trial (a subset of one, two, or three faces, or the full set of four). Response error was calculated on each trial by computing the difference between the observer's estimate and the actual gaze of the full set or subset. Error bars reflect ±1 *SEM* (adjusted for within-observer comparison).

gaze (full set vs. subset) was significant, $F(3, 21) = 3.95$, $p < .05$, $\eta_p^2 = .361$. Although a significant interaction is possible even when ensemble coding does not occur (e.g., top left panel of Fig. 3a and bottom panel of Fig. 3b), it is meaningful in this case, when errors measured against the subset were lower than errors measured against the full set. Specifically, the interaction shows that our primary effect was not simply the result of redundancy gain, in which multiple faces are easier to perceive than a single face (Won & Jiang, 2013). That is, the reduction in error measured against the full set's gaze was greater than the reduction that would have occurred simply by viewing more faces (as measured by errors relative to the subset), which had the effect of increasing signal-to-noise ratio (Alvarez, 2011). Measuring the interaction is also important because, hypothetically, it may not have materialized if observers had used a mix of strategies (e.g., using three faces on some trials and guessing from the middle of the response range on other trials).

Although the previous analyses did not directly reveal the number of faces observers used to estimate the gaze of the crowd, the striking similarity between our results and the patterns from our ideal-observer analysis suggests that observers integrated at least two, and probably three, faces from the crowd. This conservative estimate is consistent with the suggestion that an effective sample size tends to be around $\sqrt{n}$ (Dakin et al., 2009). Pairwise comparisons of error relative to the full set's mean are also consistent with this conclusion. Compared with viewing just a single face, viewing two faces tended to reduce error relative to the full set's gaze, $t(7) = 2.12$, $p = .07$, $d = 0.75$. A similar improvement occurred when observers viewed three faces rather than two, $t(7) = 2.46$, $p < .05$, $d = 0.869$, but the improvement gained from viewing four faces instead of three did not reach significance, $t(7) = 0.131$, n.s.

It was important to verify that the ensemble percept of crowd gaze was constructed using global interactions across facial features and not just parts of the faces. To do this, we compared the perceived crowd gazes with the actual crowd gazes that emerged from combining pupil and head rotations, the heads alone, or the pupils alone. Specifically, we calculated the variance of the differences between gaze estimates and (a) the average gazes of the subsets of visible faces in terms of head and pupil combinations (the same values from the analysis above), (b) the average physical gaze of the visible faces in terms of head rotation, and (c) the average physical gaze of the visible faces in terms of pupil rotation. The ensemble code was constructed using emergent crowd gaze; estimates followed the gazes determined from combining pupils and heads more closely than the gazes determined from heads alone, $t(7) = 2.56$, $p < .05$, $d = 0.908$, or pupils alone, $t(7) = 2.62$, $p < .05$, $d = 0.927$.

## Experiment 2: Perception of Crowd Gaze With Inverted Faces

The images and the norming data in Figures 1 and 2 clearly confirm that perception of a single person's gaze relies on global-level interactions between facial features and is not based on head or pupil information alone. This is consistent with decades of research showing that face representation and perception occur more at the level of grouped features and less at the level of individual features (e.g., Maurer, Le Grand, & Mondloch, 2002; Suzuki & Cavanagh, 1995). The final analysis from Experiment 1 showed that the ensemble perception of a crowd's gaze is also rooted in a stage of visual analysis at which gaze information is represented globally. In our second experiment, we sought to provide converging evidence to support this conclusion. Specifically, our goal was to determine whether weakening of the global feature interactions underlying the perception of a single face's gaze would produce a concomitant weakening of the ensemble gaze effect with crowds. To accomplish this, we repeated Experiment 1 with inverted faces.

Inversion does not eliminate the Wollaston interaction—the global interactions underlying perception of emergent gaze—although previous work suggests that it does produce modest reductions in its strength (Langton et al., 2004; Maruyama & Endo, 1984). This is consistent with recent suggestions that face-inversion effects reflect a quantitative change in encoding in which interactions among facial features persist, albeit to a lesser extent (Farah, Wilson, Drain, & Tanaka, 1995; Loffler, Gordon, Wilkinson, Goren, & Wilson, 2005; Perrett, Oram, & Ashbridge, 1998; Riesenhuber, Jarudi, Gilad, & Sinha, 2004; Sekuler, Gaspar, Gold, & Bennett, 2004). If inverting a face reduces a rotated head's pull on perceived gaze, and if the collective gazes in our crowds were the result of these global-level feature interactions, then inverting the faces in a crowd should produce a similar weakening of our pattern of ensemble integration. Note that we did not predict that inversion would eliminate the pattern of gaze integration we found with upright faces. Rather, we expected the same pattern to persist, but with a weaker benefit from pooling multiple gazes. Such a finding would converge with our previous analysis (see Results from Experiment 1) to demonstrate that summary representation of a crowd's gaze occurs in high-level visual processing.

## Method

***Observers.*** The same 8 observers who participated in Experiment 1 gave informed consent to participate.

***Stimuli and procedure.*** The stimuli and procedure were identical to those in Experiment 1, except that the stimuli were inverted.

***Preliminary norming experiment.*** The prediction of reduced integration in a crowd of inverted faces is based on the reasonable assumption that inverting a single face will reduce the attractive effect of head rotation on perceived gaze. If this were true, then the gazes in our crowds would appear more homogeneous, which would reduce the benefit of integrating multiple faces to estimate a crowd's gaze. We verified this assumption before running Experiment 2 by repeating our norming experiment, but this time with inverted faces. The response face, which appeared below the test face, remained upright.

Inverted head rotations attracted perceived gaze direction. This was confirmed by a main effect of head rotation in a repeated measures ANOVA (four head rotations × four pupil rotations), $F(3, 21) = 18.57$, $p < .001$, $\eta_p^2 = .726$. Compared with the attraction from upright head rotations in Experiment 1, equivalent rotations of inverted heads tended to be less effective at modulating perceived gaze, although this difference was not significant, $F(3, 21) = 1.95$, $p = .15$, $\eta_p^2 = .213$. On average across all head-pupil combinations, 1° of inverted head rotation pulled perceived gaze by 2.24°. Given this reduced but still substantial influence of inverted head rotation on perceived gaze, we predicted that ensemble integration would still occur in a crowd of inverted faces, but to a weaker extent than with upright faces. Consistency between the subtle effect of inversion in our norming experiment and a subtle reduction in ensemble integration with inverted faces would lend further support to our claim that the ensemble code is constructed from emergent gaze representations.

## Results

We used the norming data from Experiment 1, with upright faces, to calculate the difference between observers' estimates and the gazes of the subset and full set of inverted faces on each trial. Then, for each observer, we calculated the variance across these difference scores as our dependent variable. We calculated errors against upright norming values rather than against inverted norming values in order to make direct comparisons against the data with upright faces. If the perceived gazes used to estimate a crowd of upright faces' gaze are present in a crowd of inverted faces, then results should be identical regardless of the choice of reference values.

Observers recovered gaze information from crowds of inverted faces, but the benefit gained from doing so was not as pronounced as with upright faces. Gaze estimates from larger subsets approached the gaze of the full crowd, $F(3, 21) = 3.28$, $p < .05$, $\eta_p^2 = .319$ (Fig. 4b). Measured against the subset's gaze, response errors did not change as a function of subset size, $F(3, 21) = 0.974$, n.s. (Fig. 4b). Unlike with upright faces, the interaction between subset size and the comparison gaze (full set vs. subset) was not significant, $F(3, 21) = 1.48$, n.s.

It may seem paradoxical that overall response error was lower with inverted faces than with upright faces. This is, in fact, exactly what one would expect if the magnitude of the gaze shift from a rotated head were reduced. Our norming experiment showed that inverted faces would have been slightly more likely than upright faces to be perceived as looking straight ahead, which would lead, reports of crowd gaze to be clustered more around the middle of the response range. As is clear from Figure 3b, responses from the middle of the range do, in fact, reduce overall error. Note, however, that because inverting a face does not completely eliminate the influence of head rotation on perceived gaze (i.e., there is still a Wollaston interaction), we still found a pattern of gaze integration, albeit on top of a slight reduction in overall error.

To determine whether the benefit of ensemble integration was stronger with upright faces than with inverted faces, we compared the interactions across our first two experiments. That is, for each number of faces visible (one through four), we calculated the difference between the variability of estimates measured against the full set's gaze and against the subset's gaze (i.e., the differences between the black and gray lines in Figs. 4a and 4b). A main effect of face orientation in a within-subjects ANOVA revealed that observers tended to experience a greater improvement from integrating multiple faces when they were upright than when they were inverted, $F(1, 7) = 5.47$, $p = .051$, $\eta_p^2 = .439$. This is consistent with our prediction that the emergent single-face gazes measured in our norming experiments were the same gazes being integrated in our crowds.

## Experiment 3: Perception of Crowd Gaze With Very Brief Presentation

Although our first two experiments clearly demonstrate that observers rapidly integrated multiple gazes to perceive a crowd's direction of looking, it is nevertheless possible that with a display duration of 1,000 ms, observers could have serially attended to individual faces and cognitively computed their average gaze direction. Such a strategy would not necessarily qualify as ensemble coding, in which the average percept is achieved through rapid integration in parallel, bypassing the need for focused attention. Therefore, in Experiment 3, we tested whether the integration we observed in Experiments 1 and 2 could occur even when observers viewed each crowd for only 200 ms. This extremely brief duration prevented observers from making saccades to multiple faces or initiating serial shifts of attention.

### Method

***Observers.*** Sixteen new observers gave informed consent to participate. We doubled the number of observers because we expected that the data might be noisier with the reduction in presentation time.

***Stimuli and procedure.*** The stimuli were identical to those used in Experiments 1 and 2. The procedure was also identical, except that faces were shown for only 200 ms, and each observer completed two blocks with upright faces and two blocks with inverted faces, in alternating order. Half the observers started with an upright block, and half started with an inverted block.

### Results

The procedures for calculating the difference between each observer's estimates and the gazes of the subset and full set were identical to those used in Experiments 1 and 2. Even with only 200 ms to view the displays, observers integrated gaze information from multiple upright faces in a crowd. Gaze estimates from larger subsets were closer to the gaze of the full crowd, $F(3, 45) = 12.06$, $p < .01$, $\eta_p^2 = .445$ (Fig. 5a). Measured against the subset's gaze, there was a trend for response errors to decrease with larger subsets sizes, $F(3, 45) = 2.43$, $p = .077$, $\eta_p^2 = .139$ (Fig. 5a). The interaction between crowd size and comparison gaze (full set vs. subset) was significant, $F(3, 45) = 28.96$, $p < .01$, $\eta_p^2 = .658$. Compared with viewing just a single face, viewing two faces reduced error relative to the full set's gaze, $t(15) = 3.04$, $p < .01$, $d = 0.76$. Although improvements when observers viewed three faces rather than two, $t(15) = 1.51$, $p = .15$, $d = 0.75$, and four faces rather than three, $t(15) = 1.72$, $p = .10$, $d = 0.43$, were not significant, viewing four faces was clearly better than viewing two faces, $t(7) = 2.43$, $p < .05$, $d = 0.61$.

Observers also recovered ensemble gaze information from crowds of inverted faces seen for only 200 ms, but the benefit gained in doing so was not as pronounced as with upright faces. Gaze estimates from larger subsets approached the gaze of the full crowd, $F(3, 45) = 13.41$, $p < .01$, $\eta_p^2 = .472$ (Fig. 5b). Measured against the subset's gaze, response errors decreased as a function of subset size, $F(3, 45) = 6.06$, $p < .01$, $\eta_p^2 = .287$. The interaction between subset size and the comparison gaze (full set vs. subset) was significant, $F(3, 45) = 7.881$, $p < .01$, $\eta_p^2 = .344$.

Most important, we compared the interactions across the two face orientations to determine whether ensemble integration was stronger with upright faces than with inverted faces. A main effect of face orientation in a within-subjects ANOVA revealed that observers experienced a greater improvement from integrating multiple

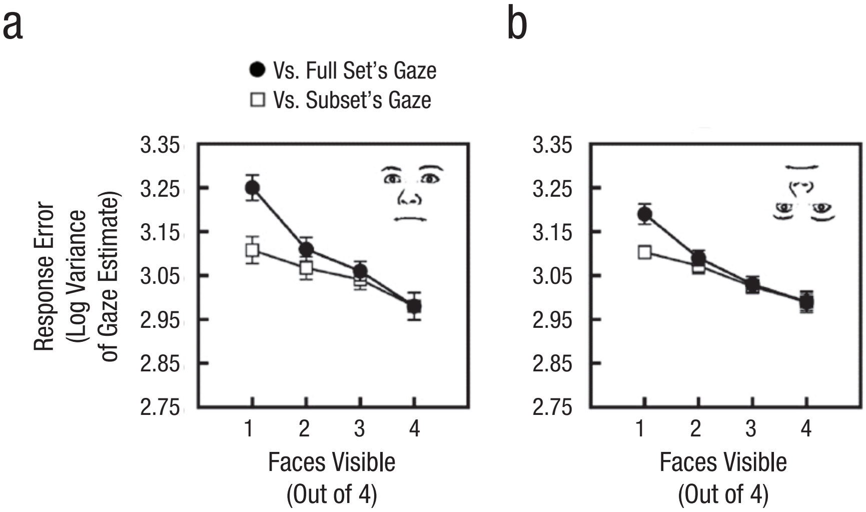

**Fig. 5.** Ensemble perception of a crowd's collective gaze in Experiment 3 when (a) upright and (b) inverted faces were shown for only 200 ms. For each panel, the mean response error is shown as a function of the number of faces that were visible on a given trial (a subset of one, two, or three faces, or the full set of four). Response error was calculated on each trial by computing the difference between the observer's estimate and the actual gaze of the full set or subset. Error bars reflect ±1 *SEM* (adjusted for within-observer comparison).

faces when they were upright than when they were inverted, $F(1, 15) = 68.21$, $p < .01$, $\eta_p^2 = .819$.

## General Discussion

We showed that the visual system pools the gazes of individual faces into an ensemble code that allows humans to rapidly and efficiently perceive where a crowd is looking. The gazes in the faces in our experiments were emergent—relying on the global integration of facial features and eyes—and inversion diminished the benefit of ensemble integration. These facts converge to suggest that ensemble perception of crowd gaze is achieved by integrating global-level facial representations (e.g., Maurer et al., 2002; Suzuki & Cavanagh, 1995) and that this integration is rooted in high-level visual areas where head rotation and eye gaze are jointly represented (De Souza, Eifuki, Tamura, Nishijo, & Ono, 2005; Perrett et al., 1985). Although it is possible that observers accessed the crowd's gaze by separately extracting average pupil rotation and average head rotation and then combining these values, this possibility seems unlikely, at least when the faces were upright, because facial organization is known to block access to individual features (Suzuki & Cavanagh, 1995). Our results add to growing evidence (Haberman & Whitney, 2007; Sweeny, Haroz, & Whitney, 2012; Sweeny et al., 2013) that ensemble coding operates at multiple levels of the visual hierarchy, enhancing not only the way people see groups of objects and textures (Ariely, 2001; Dakin et al., 2009), but also their extraction of social information in crowds of people.

Although several investigations have demonstrated the importance of perceiving a single person's gaze (Baron-Cohen, 1995; Driver et al., 1999; Friesen & Kingstone, 1998), sensitivity to individual gaze does not necessarily or inevitably lead to sensitivity to crowd gaze, nor do these previous studies begin to account for how gaze might be perceived in groups. Our results help to bridge this gap and may be useful for understanding other perceptual phenomena that occur in crowds. For example, ensemble representation of a crowd's gaze may provide the underlying metric of similarity behind pop-out of gaze in visual search (Doi & Kazuhiro, 2007).

Most important, our results may break new ground in understanding behaviors and attributions that occur only in groups (Waytz & Young, 2012). Our results characterize the perceptual aspect of a sort of joint attention that is most potent at the level of the crowd—an emergent joint attention. When people see social information, such as gaze, in a crowd, their tendency to join in is strikingly amplified compared with when they view the same information in an individual (Gallup et al., 2012; Milner et al., 1969). This pull toward conformity is widespread in the animal kingdom (for a review, see Sumpter & Pratt, 2008) and is crucial for the maintenance of group cohesion and consensus decision making (i.e., the "wisdom of the crowd"). Our results show, for the first time, that visual mechanisms are capable of representing the collective crowd properties involved in this amplified joint attention. This type of summary encoding could also be especially useful for quickly evaluating and responding to other social cues uniquely conveyed by groups, such as panic and rioting (Granovetter, 1978; Helbing, Farkas, & Vicsek, 2000). More generally, our results show that in order to understand visual processing, it is vital to consider the social pressures and group behaviors with which humans evolved.

## Author Contributions

T. D. Sweeny and D. Whitney developed the study concept. T. D. Sweeny designed the study with feedback from D. Whitney. T. D. Sweeny performed testing and data collection. T. D. Sweeny analyzed and interpreted the data with feedback from D. Whitney. T. D. Sweeny drafted the manuscript, and D. Whitney provided critical revisions. Both authors approved the final version of the manuscript for submission.

## Acknowledgments

The authors thank Erika Gonzalez for creating the stimuli at the University of California, Berkeley, and Alex McDonald for collecting and analyzing data at the University of Denver.

## Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

## Funding

This study was supported by National Institutes of Health Grant R01 EY018216 and National Science Foundation Grant NSF 1245461.

## Supplemental Material

Additional supporting information can be found at http://pss.sagepub.com/content/by/supplemental-data

## References

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*, *4*, 267–278.

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

Anstis, S. M., Mayhew, J. W., & Morley, T. (1969). The perception of where a face or television 'portrait' is looking. *The American Journal of Psychology*, *82*, 474–489.

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

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Cline, M. G. (1967). The perception of where a person is looking. *The American Journal of Psychology*, *80*, 41–50.

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), Article 28. Retrieved from http://www.journalofvision.org/content/9/11/28.full

De Souza, W. C., Eifuki, S., Tamura, R., Nishijo, H., & Ono, T. (2005). Differential characteristics of face neuron responses within the anterior superior temporal sulcus of macaques. *Journal of Neurophysiology*, *94*, 1252–1266.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, *11*, 86–89.

Doi, H., & Kazuhiro, U. (2007). Searching for a perceived stare in the crowd. *Perception*, *36*, 773–780.

Driver, J., Davis, J., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, *6*, 509–540.

Farah, M. J., Wilson, K. D., Drain, H. M., & Tanaka, J. R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perception mechanisms. *Vision Research*, *35*, 2089–2093.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*, 490–495.

Gallup, A. C., Hale, J. J., Sumpter, D. J., Garnier, S., Kacelnik, A., Krebs, J. R., & Couzin, I. D. (2012). Visual attention and the acquisition of information in human crowds. *Proceedings of the National Academy of Sciences, USA*, *109*, 7245–7250.

Gibson, J. T., & Pick, A. D. (1963). Perception of another person's looking behavior. *The American Journal of Psychology*, *76*, 386–394.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, *83*, 1420–1443.

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

Helbing, D., Farkas, I., & Vicsek, T. (2000). Simulating dynamical features of escape panic. *Nature*, *407*, 487–490. doi:10.1038/35035023

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience & Biobehavioral Reviews*, *33*, 843–863.

Langton, S. R. H., Honeyman, H., & Tessler, E. (2004). The influence of head contour and nose angle on the perception of eye-gaze direction. *Perception & Psychophysics*, *66*, 752–771.

Loffler, G., Gordon, G. E., Wilkinson, F., Goren, D., & Wilson, H. R. (2005). Configural masking of faces: Evidence for high-level interactions in face perception. *Vision Research*, *45*, 2287–2297.

Maruyama, K., & Endo, M. (1984). Illusory face dislocation effect and configurational integration in the inverted face. *Tohoku Psychologica Folia*, *43*, 150–160.

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

Milner, S., Bickman, L., & Berkowitz, L. (1969). Note on the drawing power of crowds of different size. *Journal of Personality and Social Psychology*, *13*, 79–82.

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.

Perrett, D. I., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: An account of generalisation of recognition without mental transformations. *Cognition*, *67*, 111–145.

Perrett, D. I., Smith, A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society B: Biological Sciences*, *223*, 293–317.

Riesenhuber, M., Jarudi, I., Gilad, S., & Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. *Proceedings of the Royal Society B: Biological Sciences*, *271*, S448–S450.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, *14*, 391–396.

Sumpter, D. J., & Pratt, S. C. (2008). Quorum responses and consensus decision making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 743–753.

Suzuki, S., & Cavanagh, P. (1995). Facial organization blocks access to low-level features: An object inferiority effect. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 901–913.

Sweeny, T. D., Haroz, S., & Whitney, D. (2012). Reference repulsion in the categorical perception of biological motion. *Vision Research*, *64*, 26–34.

Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 329–337. doi:10.1037/a0028712

Waytz, A., & Young, L. (2012). The group-member mind trade-off: Attributing mind to groups versus group members. *Psychological Science*, *23*, 77–85.

Whitney, D., Haberman, J., & Sweeny, T. D. (2014). 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.

Wollaston, W. H. (1824). On the apparent direction of eyes in a portrait. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *114*, 247–256.

Won, B. Y., & Jiang, Y. V. (2013). Redundancy effects in the processing of emotional faces. *Vision Research*, *78*, 6–13.