Research Article

Psychological Science
2017, Vol. 28(2) 193–203

DOI: 10.1177/0956797616678188
www.psychologicalscience.org/PS

Ensemble Perception of Dynamic Emotional Groups

Elric Elias[1], Michael Dyer[2], and Timothy D. Sweeny[1]
[1]Department of Psychology, University of Denver, and [2]Department of Psychology, Hamilton University

Abstract

Crowds of emotional faces are ubiquitous, so much so that the visual system utilizes a specialized mechanism known as ensemble coding to see them. In addition to being proximally close, members of emotional crowds, such as a laughing audience or an angry mob, often behave together. The manner in which crowd members behave—in sync or out of sync—may be critical for understanding their collective affect. Are ensemble mechanisms sensitive to these dynamic properties of groups? Here, observers estimated the average emotion of a crowd of dynamic faces. The members of some crowds changed their expressions synchronously, whereas individuals in other crowds acted asynchronously. Observers perceived the emotion of a synchronous group more precisely than the emotion of an asynchronous crowd or even a single dynamic face. These results demonstrate that ensemble representation is particularly sensitive to coordinated behavior, and they suggest that shared behavior is critical for understanding emotion in groups.

Keywords

visual perception, emotion, social perception, ensemble coding, summary perception, open data

Received 10/7/15; Revision accepted 10/17/16

Observing, evaluating, and reacting to crowds is a staple of daily life. Crowds are not just obstacles people must navigate, though; they add significance to social and emotional interactions. For example, what would a football game or surprise party be without the crowd, cheering in unison or laughing together? Crowds also exert unique influences on perception and behavior. People appear more attractive in a crowd (Walker & Vul, 2013), people gaze longer at groups that move together (Woolhouse & Lai, 2014), and groups that gaze in the same direction more strongly direct joint attention than do individuals (Milgram, Bickman, & Berkowitz, 1969). How is it that crowds, especially those that exhibit shared behavior, exert such perceptual potency? To see crowds, humans utilize a visual mechanism known as *ensemble coding* (Alvarez, 2011; Whitney, Haberman, & Sweeny, 2014). This mechanism compresses information across a group's constituents into a summary statistic, which allows people to see a group in terms of its collective attributes, such as an audience's average emotion (Haberman & Whitney, 2007). Although it is clear that ensemble coding is flexible enough to summarize social information, the extent to which it is tailored for the features that define social groups is still unclear. One central feature of social groups is that they often share behavior; people who form a group are not just proximally close to one another, they behave together (Ip, Chiu, & Wan, 2006). If ensemble coding is truly useful for summarizing social information, then it should be most strongly engaged when group information is at its most potent—that is, when a group is behaving together. Here, we tested this hypothesis by evaluating ensemble coding in the context of dynamic emotional groups.

Ensemble coding enables a kind of rapid "gist" perception, which allows people to judge the characteristics of a large crowd with a mere glance (Haberman & Whitney, 2007). Ensemble coding is thus computationally efficient, allowing the visual system to bypass bottlenecks of attention (e.g., Chong & Treisman, 2005) and working memory (e.g., Awh, Barton, & Vogel, 2007), which results in a pooled summary percept that is fast and accurate.

Corresponding Author:
Elric Elias, University of Denver, Department of Psychology, 2155 S. Race St., Frontier Hall, Denver, CO 80210
E-mail: elric.elias@du.edu

Ensemble coding is also flexible: It operates across a range of visual features. Not only does it allow people to perceive summary information about relatively basic visual features, such as motion (Watamaniuk, Sekuler, & Williams, 1989), orientation (Ross & Burr, 2008), and size (Ariely, 2001), but it also allows people to see the gist of more complex visual features, such as static faces (Haberman & Whitney, 2007) and moving bodies (Sweeny, Haroz, & Whitney, 2013). Such summary representation is also remarkably accurate, allowing people to perceive the characteristics of a group with more precision than they see the characteristics of an individual (Sweeny, Haroz, & Whitney, 2013).

The precision and efficiency of ensemble coding come at a cost, however, in that the perceiver loses access to information about individuals (Haberman & Whitney, 2007). Thus, it seems reasonable to expect that ensemble coding is not activated indiscriminately on any set of proximal features but instead is utilized exclusively, or at least more strongly, when a person encounters objects or people with shared attributes. Here, we tested this hypothesis by examining ensemble coding in the context of one of the most socially important group behaviors in which people engage—the dynamic expression of emotion (Sy, Cote, & Saavedra, 2005).

People are adept at perceiving subtle expressions of dynamic emotion on a single face (Ambadar, Schooler, & Cohn, 2005). Ensemble coding is also capable of summarizing dynamic emotion on a single face (Haberman, Harp, & Whitney, 2009; Hubert-Wallander & Boynton, 2015). It is thus reasonable to expect ensemble coding, at the very least, to be capable of summarizing information from dynamic crowds of emotional faces. More important, if ensemble coding is sensitive to the shared properties that elevate *crowds* to *groups*, then it should be especially efficient when people view groups that act collectively. People should be better at perceiving the average emotion of a group that moves together compared with a crowd composed of individuals behaving independently. In Experiment 1, we tested whether ensemble coding is more sensitive to the shared behavior of groups than to the erratic behavior of crowds. In Experiment 2, we refined our understanding of the specific visual information that ensemble coding acts on when summarizing dynamic groups of faces. In Experiment 3, we tested whether ensemble coding is sensitive to shared behavior even when emotional variability within a crowd is carefully controlled.

Experiment 1

Method

Observers. Thirty students from the University of Denver participated in Experiment 1. Observers granted informed consent and had normal or corrected-to-normal visual acuity. In a previous investigation with a similar design, number of trials, and analysis, we had sufficient power to detect and replicate an ensemble-coding effect using different stimuli with only 8 observers (Sweeny & Whitney, 2014). In anticipation of a potentially smaller effect size, we more than tripled the number of observers.

Stimuli. When designing our stimulus set, we wanted the flexibility to create crowds with multiple identities. Additionally, we needed a sufficient number of stimuli so that when different emotional intensities were displayed in succession, each identity would appear to dynamically express emotion. We selected images of neutral, fearful, happy, and angry facial expressions portrayed by eight actors from the NimStim face set (Tottenham et al., 2009). We selected closed-mouth expressions to minimize ghosting effects during morphing and because teeth can mislead observers discriminating emotional categories (Sweeny, Suzuki, Grabowecky, & Paller, 2013). For the purposes of a separate investigation, our face set included Black and White faces; however, racial identity was not central to the current investigation and will not be discussed further. For each face in our initial set of 32 exemplars, we replaced the background of each image with uniform gray (RGB value = 170, 170, 170).

To create the emotional face space, we used morphing software (Abrosoft FantaMorph Version 5.4.2; www.fantamorph.com) to linearly interpolate 48 morphs (i.e., "morph units") between each actor's neutral expression and that same actor's fearful, angry, and happy expressions. This produced a total of 1,184 unique faces (1 neutral, 49 fearful, 49 angry, 49 happy faces × 8 actors). As a result, no face in the resulting face space (Fig. 1) portrayed two emotional categories simultaneously (e.g., happy and angry). Additionally, each transition from neutral toward an emotional exemplar could be mirror-reversed (e.g., from neutral to happy and from happy back to neutral). Mirror-reversing allowed us to create face spaces for each actor that could then be smoothly navigated by observers during a method-of-adjustment response (Haberman et al., 2009); end points and potential response compression were thus avoided.

We ran a pilot study with a separate group of 8 observers to evaluate the face spaces of the eight actors. On each trial, an observer viewed a single face from our stimulus set for 506 ms and then rated its expressiveness on a scale from 1 (*least expressive*) to 10 (*most expressive*). Each observer completed 400 trials with the test face randomly drawn from the entire stimulus set on each trial. We then obtained linear fits for the relationship between physical intensity (1–50) and perceived intensity (1–10) for each actor and emotion. Crucially, each emotion range was comparably expressive—the slopes of the

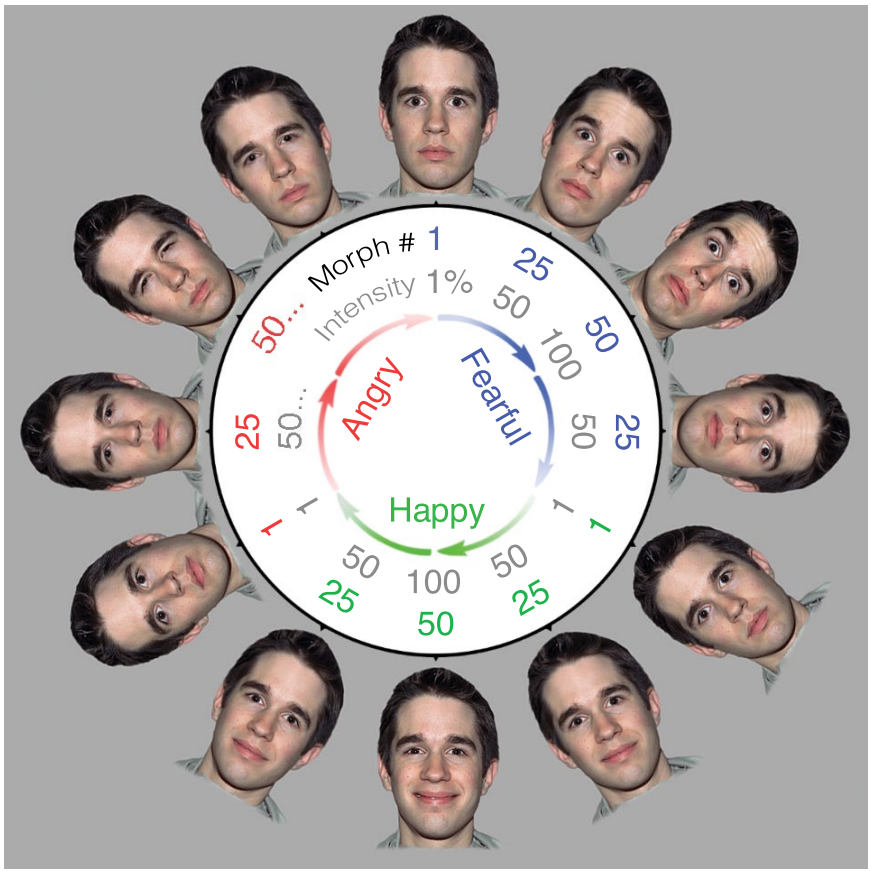

Fig. 1. Visualization of the emotional face space for one actor. We used similar face spaces from several actors to create dynamic displays for single-face and multiface trials and to create response faces that observers adjusted to match the average expression of the face or faces in each trial. For each actor, 50 stimuli were created for each of three emotion categories (fearful, happy, and angry) by morphing from neutral (Morph Unit 1, 1% intensity) to most expressive (Morph Unit 50, 100% intensity). During each trial, these dynamic faces smoothly became either more or less expressive within a single emotion category without ever abruptly changing from maximally expressive to neutral. The circular arrangement of faces in this visualization reflects the continuous nature of the face space, especially as it pertained to the adjustments during the response stage—as observers moved the cursor, the emotional intensity of the response face changed smoothly, even from one emotion category to the next. This allowed observers to traverse the face space clockwise or counterclockwise continuously and indefinitely, selecting both the emotion and intensity for each trial without reaching any end points. Faces are reprinted with permission from the NimStim face database (Tottenham et al., 2009).

linear fits we obtained did not differ across emotion, $F(2, 23) = 1.39$, n.s. We then identified the two faces with the widest range of emotional intensity, which we used exclusively as the response faces in our main experiments. This pilot study also provided us with a means to analyze our data based on perceived emotional intensity (see Results of Experiment 1 for details).

All faces subtended a visual angle of 4.88° × 5.5°. Experiments were conducted on a CRT monitor with a refresh rate of 85 Hz at a viewing distance of 55 cm. Stimuli were presented against a uniform gray background (RGB value = 170, 170, 170; luminance = 27.5 cd/m^2). Experiments were coded and run using MATLAB (Release 2014b; The MathWorks, Natick, MA) with the Psychophysics Toolbox (Brainard, 1997).

Procedure. The experiment consisted of multiface trials and single-face trials. In multiface trials, 12 faces appeared scattered around a fixation point. Each multiface trial was composed of six unique identities, with each identity appearing twice. In each single-face trial, just one face appeared randomly in 1 of the 12 possible crowd-member locations (Fig. 2). The 12 positions were not fixed; the centroid of each position varied randomly on each trial by

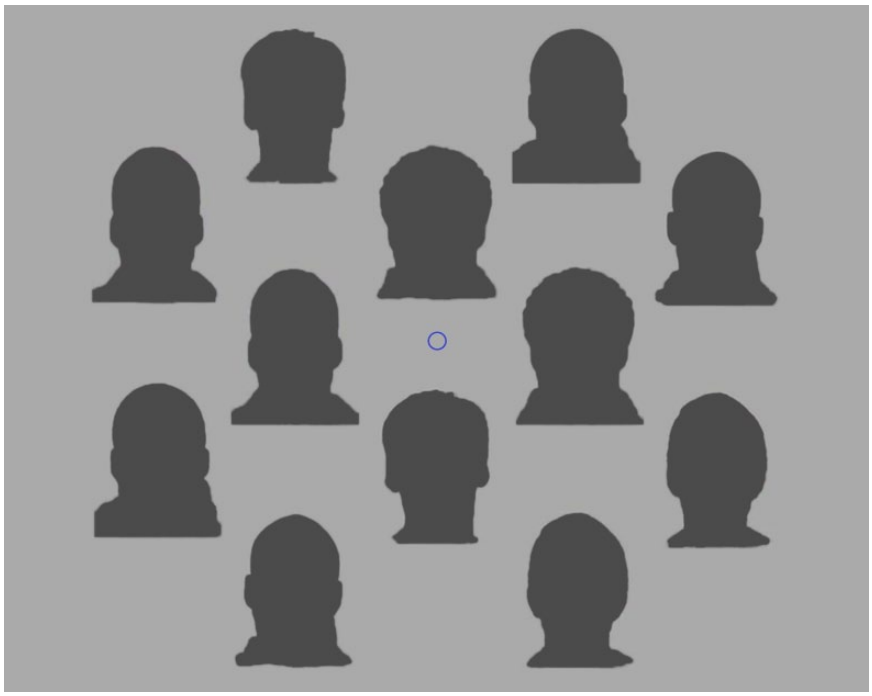

Fig. 2. Example spatial layout of the 12 possible face positions. In multiface trials, faces appeared in all positions; there were six unique facial identities, with each identity appearing twice in a crowd (silhouettes are presented here for purposes of illustration, although actual faces were used in the study). The 12 faces shared the same category of expression (e.g., fearful) and then either changed expression together as a group (synchronous-group trials) or erratically as individuals (asynchronous-crowd trials). In single-face trials, one face was displayed randomly at 1 of these 12 locations.

1 to 15 pixels in either direction along the horizontal and vertical axes. Without taking into account this random position variation, the centroid of adjacent faces was 10.8° away from each other along the horizontal axis and 8.1° away from each other along the vertical axis.

Target faces were displayed for either 500 or 1,000 ms,[1] randomly determined on each trial. Although ensemble representation of a single face's average dynamic emotion can occur in as little as 250 ms (Hubert-Wallander & Boynton, 2015), it also uniquely builds over time (Hubert-Wallander & Boynton, 2015), improving even up to durations of 800 ms (Haberman et al., 2009). Using durations of 500 and 1,000 ms thus allowed us to test whether ensemble representation is subject to the same temporal constraints with crowds as it is with individuals—specifically, whether it improves throughout the 800-ms integration window.

Each face in a multiface trial displayed an emotion from the same randomly determined emotion category. The mean emotion of the crowd, at each trial's start, was randomly chosen from a uniform distribution of intensity values ranging from 20% to 80% (equivalent to 10 to 40 morph units; see Fig. 1). All multiface trials contained emotional variability, with each face displaying a unique amount of emotional intensity at the start of a trial. These intensity values were randomly selected from a normal sampling distribution with a standard deviation of 8% (i.e., 4 morph units), centered on that trial's randomly selected emotional mean at the start of the trial. For example, faces' mean emotion on a multiface trial might initially have been 25% happy, and because of the variability from random sampling, the initial intensities of individual faces would be normally distributed between 17% and 33%. Each single-face trial began with one randomly selected identity displaying a random emotional intensity drawn from a similar sampling distribution, with that distribution's mean between 20% to 80% intensity, and a standard deviation of 8%.

Each face was dynamic in all trials, changing in emotional intensity within one emotion category by single morph units. In half of the multiface trials, the faces changed intensity in a coordinated way—the *synchronous condition*. For example, a synchronous group could have started with an average fearful intensity of 65%. As the

trial progressed, the group could have initially become less fearful (or more fearful—initial direction of intensity change was random in all trials), with all members synchronously approaching a neutral intensity at the same rate. As soon as any member of the group reached the end of the range of emotional intensities for that category (e.g., neutral or maximally fearful), the entire group would then reverse direction and gradually become more fearful (or neutral) until the end of the trial.

In the other half of the multiface trials, each face randomly and independently became more or less emotionally intense—the *asynchronous condition*. For example, an asynchronous crowd might have begun a trial with an average emotional intensity of 35% happy. A random number of faces within the crowd could have then become happier, while the rest of the faces decreased in intensity toward neutral. Each face in such an asynchronous crowd acted independently, reversing direction on reaching either end of that trial's particular emotional category (e.g., neutral or maximally happy) until the trial's end.

In the remaining trials, observers viewed only a single face. On a single-face trial, the single face's initial emotional intensity was randomly selected (e.g., 30% fear) from the same uniform sampling distribution described above, and this single face became randomly more or less intense before changing direction. Neither multiface nor single-face trials necessarily ended with the same average emotion intensity with which they began.

To prevent residual visual processing, we immediately followed each face with a mask at the same location, regardless of whether it was part of a multiface or single-face trial (Rolls, Tovée, & Panzeri, 1999). Each mask was generated by dividing an emotional face from our face space into 70 rectangular pieces and then randomly shuffling the locations of these pieces. This approach ensured that the emotional faces and scrambled masks resembled each other in terms of low-level image characteristics.

Immediately after the mask was presented, observers moved a cursor to the left or right to adjust the emotional intensity of a response face, presented at the center of the screen, until it reflected the perceived average emotion of the previous crowd or single face across the duration of the trial. The identity of the response face was never present in any multiface or single-face trial. This ensured that each observer's response reflected his or her perception of the average emotional expression and did not reflect an emotion-irrelevant response strategy, such as matching the position of a freckle, which could occur if the response face had been previously seen. Observers had unlimited time to respond. All 30 observers completed 300 trials, each of which was randomly determined to be a synchronous, an asynchronous, or a single-face trial.

Results

To evaluate how sensitive observers were at perceiving emotion, we had to first measure the objective average expression across time in both single-face and multiface trials. For multiface trials, we recorded the median of each crowd or group member's expressions across the duration of each trial and then averaged these 12 values. For each single-face trial, we recorded the median facial expression. We then compared each observer's response to the actual average emotion over time on each trial. Across all trials, we compiled these signed-difference scores into separate error distributions, one for each condition (synchronous, asynchronous, and single-face). We then calculated the standard deviation of each distribution separately for each observer. Observers with greater sensitivity were expected to produce error distributions with smaller standard deviations. This approach has been used in previous investigations of ensemble coding (e.g., Haberman et al., 2009; Sweeny, Haroz, & Whitney, 2013; Sweeny & Whitney, 2014). For each observer, this analysis yielded overall error scores (*SD*s) for synchronous, asynchronous, and single-face trials, separately for both the 500-ms and 1,000-ms durations. Numerical error values were not meaningful when observers made categorical response errors (e.g., responding in the fearful expression range when the faces were happy). We thus eliminated categorical errors when computing overall error scores in all analyses.

A repeated measures 3 (trial type: synchronous, asynchronous, single-face) × 3 (emotion: fearful, angry, happy) × 2 (duration: 500 ms, 1,000 ms) analysis of variance (ANOVA) revealed main effects of trial type, $F(2, 28) = 8.81$, $p < .01$, $d = 0.39$, and emotion, $F(2, 28) = 14.36$, $p < .01$, $d = 0.51$, but not duration, $F(1, 29) = 2.99$, n.s. The interaction among trial type, emotion, and duration was not significant, $F(4, 26) = 0.13$, nor was the interaction between trial type and duration, $F(2, 28) = 3.04$, or the interaction between trial type and emotion, $F(4, 26) = 0.51$. The pattern between performance on synchronous and asynchronous trials did not change between the 500-ms and 1,000-ms durations. Thus, we conducted planned comparisons among the synchronous, asynchronous, and single-face conditions using data collapsed across the 500-ms and 1,000-ms durations.

Observers were better at perceiving the mean of synchronous groups than asynchronous crowds, $t(29) = 4.61$, $p < .01$, $d = 0.86$ (Fig. 3). Although not central to our interests, comparisons also showed that observers were not more accurate on single-face trials than on asynchronous trials, $t(29) = 0.44$, n.s. Observers were better at perceiving the emotion of a synchronous group than the emotion of a single face, $t(29) = 4.07$, $p < .01$, $d = 0.75$.

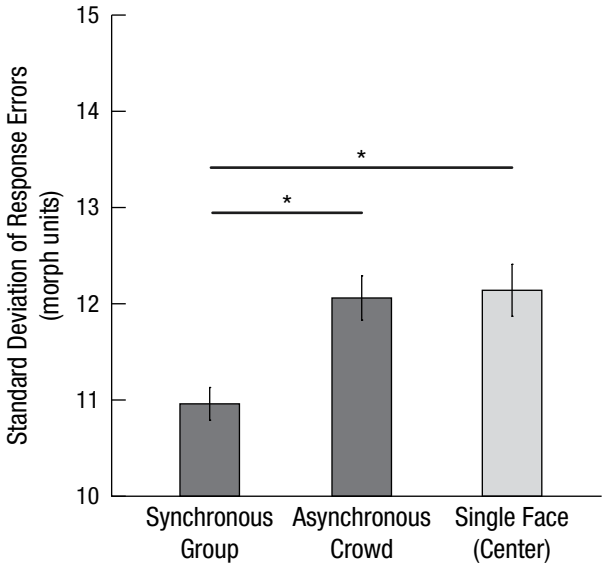

Fig. 3. Standard deviation of response errors from Experiment 1 as a function of trial type (synchronous group, asynchronous crowd, and single face in one of four center locations). Error bars represent ±1 *SEM*, adjusted for within-observers comparisons. Asterisks indicate significant differences between conditions ($p < .01$).

This surprising sensitivity in the synchronous condition held even when compared with trials in which the single face appeared in one of the four positions adjacent to fixation, $t(29) = 2.89$, $p < .01$, $d = 0.54$ (Fig. 3). The main pattern of results—more precise estimates of the emotion of synchronous groups than asynchronous crowds or even a single dynamic face—occurred for every emotion (ps < .05 for all synchronous-vs.-asynchronous and synchronous-vs.-single-face comparisons within each emotion). This demonstrates a boost in perception beyond simple differences in visual acuity, one based instead on the efficiency of ensemble representation.

To evaluate whether observers displayed sensitivity to the intensity of emotion in a dynamic array of faces above and beyond sensitivity to the emotion category the faces displayed, we regressed each observers' reports of perceived intensity (e.g., 30% happy) against the actual average physical intensity of the arrays of faces. If an observer were sensitive to emotional intensity, then his or her slope would be greater than zero. We thus collected slopes from each observer and compared the average of these slopes against zero. For both synchronous groups and asynchronous crowds, perceived intensity and physical intensity were significantly positive—synchronous-groups slope: $M = 0.28$, $SD = 0.20$, $t(29) = 7.76$, $p < .01$, $d = 1.42$; asynchronous-crowds slope: $M = 0.35$, $SD = 0.23$, $t(29) = 8.33$, $d = 1.52$.

All error values and comparisons in the preceding analyses were conducted using linearly interpolated morph units from each face (see Fig. 1). It is possible, however, that identical morph-unit values (e.g., Happy Morph Unit 25 of 50) from two identities could have displayed different amounts of emotional intensity. We therefore investigated whether the pattern of sensitivity across synchronous, asynchronous, and single-face trials—our main pattern of interest—persisted when we reanalyzed our data replacing numerical morph-unit values with perceived intensity values from our pilot norming study (e.g., a perceived-happiness rating of 5.5 out of 10). For example, if on a single trial, Identity 1 displayed Fear Morph Unit 25 on average, we referenced that actor's linear fit from the norming experiment and retrieved the perceived intensity for that same morph value. We completed this procedure for all faces on every single-face and multiface trial, including each observer's chosen response face. We then calculated signed-error scores and error distributions for each condition as in the previous analysis, but with these new intensity values.

A repeated measures ANOVA revealed that our main effect of trial type (synchronous, asynchronous, single-face) persisted with this new analysis, $F(2, 28) = 9.03$, $p < .01$, $d = 0.18$. Critically, the interaction between trial type and data type (unnormed error scores, normed error scores) was not significant, $F(2, 28) = 0.9$. Thus, the pattern of results across conditions did not differ when intensity data for each actor were taken into account.

Experiment 2

Individual faces are processed holistically when viewed upright, both when they are static (e.g., Farah, Wilson, Drain, & Tanaka, 1998) and dynamic (Singer & Sheinberg, 2006). When ensemble representation pools information from static faces, it tends to do so using holistic information (e.g., Haberman & Whitney, 2007; Sweeny & Whitney, 2014). It is tempting to assume that ensemble percepts of dynamic groups in Experiment 1 relied on similar information. Alternatively, ensemble representation in Experiment 1 could have occurred simply via the pooling of face parts into a summary statistic. We tested these competing hypotheses by presenting inverted crowds in Experiment 2.

At the very least, inversion delays accumulation of holistic information (e.g., Perrett, Oram, & Ashbridge, 1998), but it does not induce a change in processing style (Sekuler, Gaspar, Gold, & Bennett, 2004). This is especially evident when faces are studied for relatively long periods of time (e.g., Richler, Mack, Palmeri, & Gauthier, 2011), as in the current investigation. If judgments of upright faces were based on whole-face inputs, then according to this account of inversion, observers should produce the same pattern of results with inverted faces, albeit with lower sensitivity overall. Other investigations suggest that inversion encourages a more feature-based analysis (Maurer, Le Grand, & Mondloch,

2002; Moscovitch, Winocur, & Behrmann, 1997). According to this account, if the synchrony advantage in Experiment 1 were the result of whole-face inputs, the difference between synchronous and asynchronous trials should be eliminated or weakened with inverted faces. Alternatively, if perception of upright faces in Experiment 1 emerged via the pooling of information from face parts extracted via an image-based analysis, the exact same pattern should occur when this information is available for inverted faces.

Method

Observers. We recruited a new group of 30 undergraduate students from the University of Denver to participate in Experiment 2. All observers granted informed consent and had normal or corrected-to-normal vision.

Stimuli and procedure. All stimuli and procedures in Experiment 2 were identical to those from Experiment 1, except that all faces, including the response faces, were inverted.

Results

In Experiment 1, our main comparisons of interest were among the synchronous, asynchronous, and single-face (in one of the four center locations) trials. We thus focused on these comparisons in Experiment 2, again collapsing all analyses across the 500- and 1,000-ms durations and excluding miscategorizations. Inverted faces were perceived more accurately on synchronous trials than on both asynchronous trials, $t(29) = 4.74$, $p < .01$, $d = 0.87$, and single-face trials, $t(29) = 2.24$, $p < .05$, $d = 0.41$, but not more accurately than on single-face (center) trials, $t(29) = 0.28$, n.s. (Fig. 4), a pattern similar, but not identical, to that observed with upright faces.

We compared the results of Experiments 1 and 2 by conducting a two-way mixed ANOVA with one within-subjects factor (trial type: synchronous, asynchronous, single center) and one between-subjects factor (orientation: upright, inverted). A main effect of orientation revealed that, as expected, upright faces were perceived with more sensitivity than inverted faces $F(1, 58) = 8.1$, $p < .01$, $d = 0.12$. The interaction between trial type and orientation was not significant, $F(2, 57) = 1.92$, which suggests that although observers perceived inverted faces with less sensitivity than upright ones, the differences among trial types were comparable across Experiment 1 and 2. Overall, these results suggest that observers used the same information to evaluate upright and inverted faces. The relatively consistent pattern of results despite an overall increase in response errors, especially across the synchronous and asynchronous conditions, is consistent with the notion that inversion delays, but does not eliminate, holistic processing of faces, especially when faces are seen for a long amount of time.

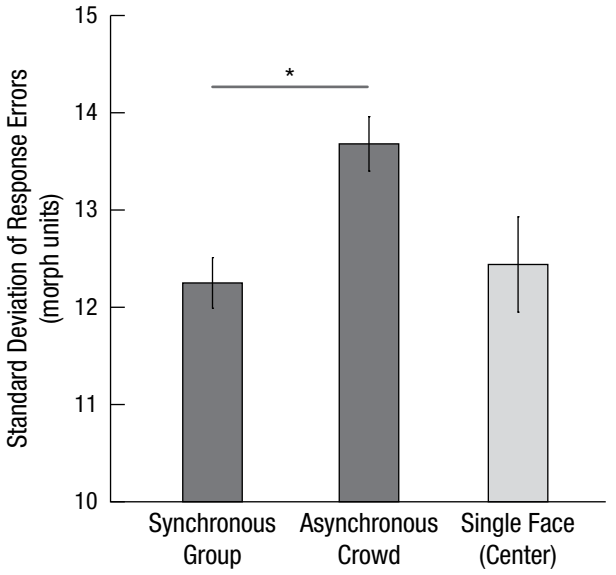

Fig. 4. Standard deviation of response errors from Experiment 2 as a function of trial type (synchronous group, asynchronous crowd, and single face in one of four center locations). Error bars represent ±1 *SEM*, adjusted for within-observers comparisons. The asterisk indicates a significant difference between conditions ($p < .01$).

Experiment 3

In Experiments 1 and 2, the variability among faces in synchronous groups never changed, since all members became either more or less happy together, at the same rate. During asynchronous trials, however, variability across crowd members fluctuated, increasing (or decreasing) as faces morphed all the way to the opposite ends (or middle) of the range of intensities. On any given frame, faces in an asynchronous crowd could have occupied a larger range of intensities than faces in synchronous groups, and this heterogeneity would have varied across the course of the trial. We confirmed this by running a 230-trial simulation of Experiment 1, which revealed that asynchronous trials did contain greater heterogeneity ($M = 5.16$ morph units, $SD = 0.75$) than synchronous trials ($M = 3.72$ morph units, $SD = 0.92$), bootstrapped $p < .01$.

Though increased heterogeneity may be a natural feature of real crowds that do not share behavior, it obscured our ability to measure the pure effect of shared behavior on ensemble coding, particularly because heterogeneous sets are known to be perceived less accurately than homogenous sets (e.g., Marchant, Simons, & de Fockert, 2013; Sweeny, Haroz, & Whitney, 2013). We thus conducted Experiment 3 to evaluate whether ensemble coding is indeed sensitive to shared behavior even when

heterogeneity across the arrays of faces was carefully controlled.

Method

Observers. We recruited a new group of 33 undergraduate students from the University of Denver. All observers granted informed consent and had normal or corrected-to-normal vision.

Stimuli and procedure. All stimuli in Experiment 3 were identical to those from Experiment 1. Several methodological details were also the same, including the locations of the faces, the display durations, the response method, and the sampling procedure used to determine the intensities of the faces in the synchronous, asynchronous, and single-face conditions at the start of each trial.

Our primary goal in Experiment 3 was to minimize the variability across faces in asynchronous crowds. To accomplish this, we restricted the dynamic range of faces in asynchronous crowds, allowing them to change their intensities only within the boundaries of their sampled group from the initial frame of each trial. In other words, no face in the crowd was ever permitted to morph beyond the minimum or maximum amount of intensity present in the initially sampled group. The amount of variability across the faces in asynchronous crowds still fluctuated as each member randomly increased or decreased in intensity, but much less so than in Experiment 1. The average emotion of any asynchronous crowd thus changed very little across a given trial despite the erratic behavior of its constituents. Synchronous groups moved in the same way as in Experiment 1; their average emotional intensity smoothly changed over time, but variability across a synchronous group's members never changed. We allowed for this difference—drifting of the mean only in the synchronous condition—because, if anything, it should have made perception in the asynchronous condition easier. That is, any single "snapshot" during an asynchronous trial would have been more representative of the group's average across the trial than a snapshot during a synchronous trial. Indeed, a preliminary simulation confirmed that the standard deviation of a crowd's mean expression across a given trial of the synchronous condition ($M = 6.71$ morph units, $SD = 3.98$) was greater than that in the asynchronous condition ($M = 1.54$ morph units, $SD = 1.10$), bootstrapped $p < .01$. Superior performance with synchronous groups in Experiment 3 would thus offer especially convincing evidence that summary perception is sensitive to shared behavior, even when differences in naturally occurring crowd heterogeneity are controlled.

Results

A repeated measures 3 (trial type: synchronous, asynchronous, single-face) × 3 (emotion: fearful, angry, happy) × 2 (duration: 500 ms, 1,000 ms) ANOVA revealed main effects of trial type, $F(2, 32) = 34.64$, $p < .01$, $d = 0.29$; emotion, $F(2, 32) = 7.462$, $p < .01$, $d = .32$; and duration, $F(1, 32) = 14.19$, $p < .01$, $d = 0.31$. The interaction among trial type, emotion, and duration was not significant, $F(4, 29) = 1.82$, nor was the interaction between trial type and duration, $F(2, 31) = 1.72$, trial type and emotion, $F(4, 29) = 0.38$, or emotion and duration, $F(2, 31) = 2.62$. We thus conducted planned comparisons among the synchronous, asynchronous, and single-face conditions using data collapsed across 500-ms and 1,000-ms durations. These comparisons revealed that summary perception was more accurate for synchronous trials than for asynchronous trials, $t(32) = 4.41$, $p < .01$, $d = 0.76$ (Fig. 5). Additionally, observers were again better at perceiving synchronous groups than single faces, even when those single faces were near the center of the group, $t(32) = 4.39$, $p < .01$, $d = 0.75$ (Fig. 5).

Variability in the synchronous and asynchronous conditions was not perfectly matched. We determined this by first obtaining the standard deviation of the intensities across faces on each frame of a trial. We averaged these values, which yielded a single value of within-frame heterogeneity on each trial. Then, for each observer, we determined the average of these standard deviations across all trials from each condition. Asynchronous trials contained slightly more heterogeneity ($M = 4.03$ morph units, $SD = 0.06$) than synchronous trials ($M = 3.72$ morph units, $SD = 0.06$), $t(32) = 17.26$, $p < .01$.

There are several reasons why this difference is not problematic. First, despite strikingly different amounts of heterogeneity in asynchronous trials in Experiment 1 ($M = 5.16$; see simulated results above) and Experiment 3 ($M = 4.03$), performance across these asynchronous conditions did not differ, $t(62) = 0.315$, n.s. More important, on a trial-by-trial basis, performance in Experiment 3 was unrelated to heterogeneity. For each observer, we regressed response error against within-frame heterogeneity on each trial, separately for each trial type. We then obtained the slope of a linear fit to this relationship and compared the distribution of slopes from all observers against a null value of zero. If performance had deteriorated when heterogeneity was high, we would have observed negative slopes for most observers. Response precision was not related to heterogeneity in synchronous trials (slope: $M = -0.11$, $SD = 0.86$), $t(32) = -0.73$, n.s., or in asynchronous trials (slope: $M = 0.07$, $SD = 1.19$), $t(32) = 0.34$, n.s. The results of Experiment 3

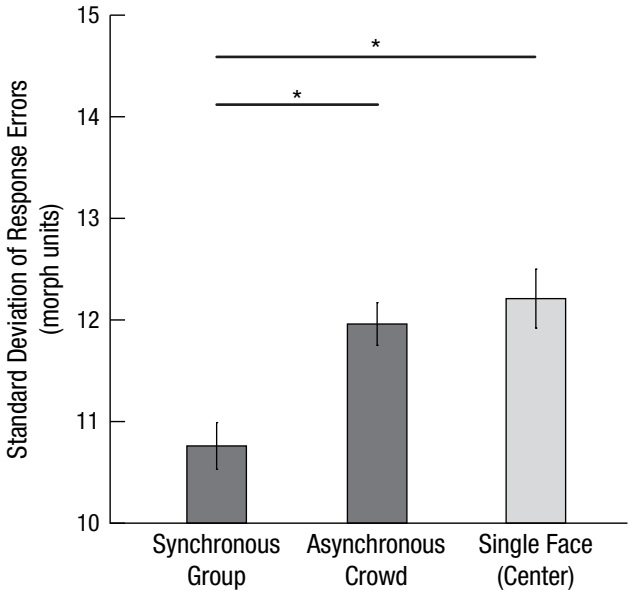

Fig. 5. Standard deviation of response errors from Experiment 3 as a function of trial type (synchronous group, asynchronous crowd, and single face in one of four center locations). Error bars represent ±1 *SEM*, adjusted for within-observers comparisons. Asterisks indicate significant differences between conditions ($p < .01$).

strengthen our conclusion that ensemble coding is sensitive to shared behavior. These results suggest again that the improved performance of observers in synchronous trials, compared with asynchronous or single-face trials, was specifically due to shared behavior and was not a result of discrepancies in heterogeneity among trial types.

Discussion

We showed that people perceive the average emotion of a dynamic array of faces more accurately when its members act collectively rather than independently. This finding demonstrates that ensemble representation is especially sensitive to shared social behavior, so much so that it even enables people to perceive the emotions of a synchronized group more accurately than those of a single dynamic face. This is surprising because the crowds and groups in our investigation were heterogeneous, and they were seen with less resolution than any individual because of diminished peripheral acuity, crowding (Whitney & Levi, 2011), and limitations of focused attention and working memory (e.g., Awh et al., 2007; Chong & Treisman, 2005). Our results add to growing evidence that ensemble representation can offset these issues by pooling across noisy signals to produce an accurate summary percept (Alvarez, 2011; Sweeny, Haroz, & Whitney, 2013), especially if the group is displaying shared behavior.

Reduced heterogeneity is almost certainly a natural feature of groups that share behavior. It is also well known that people perceive the mean of homogeneous sets more accurately than the mean of heterogeneous sets (Marchant et al., 2013). However, this cannot explain our results—even after controlling for heterogeneity between synchronous groups and asynchronous crowds, we found that the advantage of shared behavior remained. Nor can our results be explained as the result of a serial search process (Myczek & Simons, 2008). Shared behavior is, by definition, a group-level feature. Thus, the benefit of synchrony could have emerged only when the expressions of individuals were processed simultaneously and not one at a time.

Waytz and Young (2012) showed that the more a person attributes a mind to a group, the less a person then attributes a mind to that group's constituents. This tradeoff is consistent with the perceptual consequences of ensemble representation, in which information about a group is extracted at the expense of information about individuals (e.g., Haberman & Whitney, 2007). This correspondence hints that social and perceptual approaches to understanding groups (a) may be tapping into a common mechanism despite having progressed relatively independently of one another and (b) could provide new answers to unresolved questions when combined.

For example, why might ensemble coding be sensitive to shared behavior? Research in social psychology suggests that groups are defined, at least in part, by their collective behaviors, which may be built on shared values, beliefs, and attitudes (Gruenfeld & Tiedens, 2010). Considering the importance of groups in evolution (Cosmides & Tooby, 2005), it is reasonable to expect features that strongly define social groups to be prioritized in visual representation. When a group of people is identified, the visual system may thus recruit additional resources to represent them. By displaying shared behavior, the synchronous groups in our investigation may have been more likely than asynchronous crowds to have been "tagged" as a group and to thereby receive the benefit of enhanced ensemble representation.

Conversely, why are collections of people considered to be more grouplike when they share behavior (Lickel, Hamilton, & Sherman, 2001)? Research in vision science shows that ensemble representation imposes the appearance of homogeneity onto crowds of objects (e.g., Ross & Burr, 2008) and people (Sweeny, Haroz, & Whitney, 2013) whose members are not, in fact, identical. To the extent that social evaluations are rooted in visual representation, people may evaluate a synchronous group as more of a coherent unit than an asynchronous crowd because ensemble coding is operating powerfully. Indeed, collectives with high levels of "group mind" are judged to be more cohesive (Waytz & Young, 2012), and joint action increases the attribution of a mind to a group (e.g., Bloom & Veres, 1999). The present work hints at

the potential for visual representation to explain complex social phenomena and vice versa.

Our results also suggest that basic gestalt grouping cues (e.g., Wagemans et al., 2012) may gate the process of ensemble coding. The shared movements of the faces in the present investigation are strikingly similar to shared luminance changes used to demonstrate a grouping principle known as synchrony (Palmer, 1999; Sekuler & Bennett, 2001). Furthermore, our asynchronous crowds exemplify what grouping terminology defines as element aggregations—collections of weakly grouped objects whose individual elements maintain some independence (Palmer, 1999). Meanwhile, unit formation—the perception of a single, unified object via strong grouping of individuals—neatly describes our synchronous groups (Palmer, 1999). Future work should more directly evaluate the extent to which grouping influences the efficiency of ensemble representation.

In summary, we have shown that ensemble coding is sensitive to dynamic emotional information in groups of faces. Groups of emotional faces that behave together are perceived with more precision than crowds whose constituents behave individualistically, and synchronous groups are perceived even more precisely than individual faces. In other words, humans are good at perceiving a person but can be even better at perceiving people.

Action Editor

Edward S. Awh served as action editor for this article.

Author Contributions

T. D. Sweeny developed the study concept. The study was designed by T. D. Sweeny and M. Dyer and was programmed by M. Dyer. Testing and data collection were performed by E. Elias. E. Elias analyzed and interpreted the data under the supervision of T. D. Sweeny. E. Elias drafted the manuscript, and T. D. Sweeny provided critical revisions. All authors approved the final version of the manuscript for submission.

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.

Open Practices

All data have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/sgqsx/. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797616678188. This article has received the badge for Open Data. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

Note

1. On average, faces were replaced at rates of 33.67 Hz and 32.26 Hz in the 500-ms and 1,000-ms trials, respectively, which actually lasted 528 ms and 1,027 ms, respectively. These unusual values resulted from the strain of rendering a dynamic crowd in real time, which introduced subtle variability in the timing between frames in each face's animation. No observers reported noticing this, and it should not have systematically influenced our results. Critically, we still achieved our goal of presenting realistically dynamic crowds.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*, 122–131.

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, *16*, 403–410.

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

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*, 622–628.

Bloom, P., & Veres, C. (1999). The perceived intentionality of groups. *Cognition*, *71*, B1–B9.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*, 1–13.

Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 584–627). Hoboken, NJ: John Wiley & Sons.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review*, *105*, 482–498.

Gruenfeld, D. H., & Tiedens, L. Z. (2010). Organizational preferences and their consequences. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 2, pp. 1252–1287). Hoboken, NJ: John Wiley & Sons.

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

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

Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision*, *15*(4), Article 5. doi:10.1167/15.4.5

Ip, G. W. M., Chiu, C. Y., & Wan, C. (2006). Birds of a feather and birds flocking together: Physical versus behavioral cues may lead to trait- versus goal-based group perception. *Journal of Personality and Social Psychology*, *90*, 368–381.

Lickel, B., Hamilton, D. L., & Sherman, S. J. (2001). Elements of a lay theory of groups: Types of groups, relational styles,

and the perception of group entitativity. *Personality and Social Psychology Review*, *5*, 129–140.

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*, 245–250.

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

Milgram, 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.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience*, *9*, 555–604.

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.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

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.

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research*, *51*, 333–342.

Rolls, E. T., Tovée, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: Information analysis. *Journal of Cognitive Neuroscience*, *11*, 300–311.

Ross, J., & Burr, D. (2008). The knowing visual self. *Trends in Cognitive Sciences*, *12*, 363–364.

Sekuler, A. B., & Bennett, P. J. (2001). Generalized common fate: Grouping by common luminance changes. *Psychological Science*, *12*, 437–444.

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.

Singer, J. M., & Sheinberg, D. L. (2006). Holistic processing unites face parts across time. *Vision Research*, *46*, 1838–1847.

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.

Sweeny, T. D., Suzuki, S., Grabowecky, M., & Paller, K. A. (2013). Detecting and categorizing fleeting emotions in faces. *Emotion*, *13*, 76–91.

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

Sy, T., Cote, S., & Saavedra, R. (2005). The contagious leader: Impact of the leader's mood on the mood of group members, group affective tone, and group processes. *Journal of Applied Psychology*, *90*, 295–305.

Tottenham, N., Tanaka, J., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. A. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*, 242–249.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, *138*, 1172–1217.

Walker, D., & Vul, E. (2013). Hierarchical encoding makes individuals in a group seem more attractive. *Psychological Science*, *25*, 225–230.

Watamaniuk, S. N., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. *Vision Research*, *29*, 47–59.

Waytz, A., & Young, L. (2012). The group-member mind tradeoff 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.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*, 160–168.

Woolhouse, M. H., & Lai, R. (2014). Traces across the body: Influence of music-dance synchrony on the observation of dance. *Frontiers in Human Neuroscience*, *8*, Article 965. doi:10.3389/fnhum.2014.00965