Journal of Experimental Psychology:
Human Perception and Performance

© 2012 American Psychological Association
0096-1523/12/$12.00 DOI: 10.1037/a0028712

Perceiving Group Behavior: Sensitive Ensemble Coding Mechanisms for Biological Motion of Human Crowds

Timothy D. Sweeny
University of California – Berkeley

Steve Haroz
University of California – Davis

David Whitney
University of California – Berkeley

Many species, including humans, display group behavior. Thus, perceiving crowds may be important for social interaction and survival. Here, we provide the first evidence that humans use ensemble-coding mechanisms to perceive the behavior of a crowd of people with surprisingly high sensitivity. Observers estimated the headings of briefly presented crowds of point-light walkers that differed in the number and headings of their members (i.e., people in differently sized crowds had identical or increasingly variable directions of walking). We found that observers rapidly pooled information from multiple walkers to estimate the heading of a crowd. This ensemble code was precise; observer's perceived the behavior of a crowd better than the behavior of an individual. We also showed that this pooling provided tolerance against crowd variability and may cause a chaotic group to cohere into a unified Gestalt. Sensitive perception of a crowd's behavior required integration of human form and motion, suggesting that the ensemble code was generated in high-level visual areas. Overall, these mechanisms may reflect the prevalence of crowd behavior in nature and a social benefit for perceiving crowds as unified entities.

Keywords: biological motion, ensemble coding, crowds, summary statistics, collective behavior

Coordinated group behavior is common for many species (Sumpter, 2006) and important for survival (Bode, Faria, Franks, Krause, & Wood, 2010). Perceiving crowd behaviors may be important too. For example, some types of emergent social information, such as panic, are uniquely conveyed by crowd behavior (Helbing, Farkas, & Vicsek, 2000). Because the environment changes rapidly and the visual encoding of an individual's behavior can sometimes take a relatively long time (Cavanagh, Labianca, & Thornton, 2001; Neri, Morrone, & Burr, 1998; Thornton, Pinto, & Shiffrar, 1998) and effort (Thornton, Rensink, & Shiffrar, 2002), determining a crowd's movement by serially processing each individual would be of little value. Indeed, humans are capable of rapidly and automatically processing coarse differences in the movements of multiple people (Thornton & Vuong, 2004). This suggests that humans may use a specialized ensemble coding mechanism that pools the movements of individuals into a precise summary representation in order to rapidly perceive the movement of a crowd.

Ensemble coding has been characterized as an efficient mechanism for perceiving the "gist" of complex environments (e.g., Alvarez, 2011; Haberman & Whitney, 2009), and it is particularly striking when one considers the extensive visual attention literature showing that visual processing becomes slower and less accurate when multiple objects are viewed at once (e.g., Franconeri, in press). Ensemble coding is widespread in visual processing—it has been demonstrated for the perception of orientation (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), location (Alvarez & Oliva, 2008), size (Ariely, 2001; Chong & Treisman, 2003; Im & Chong, 2009; Joo, Shin, Chong, & Blake, 2009), and facial expression (Haberman, Harp, & Whitney, 2009; Haberman & Whitney, 2007, 2009) and motion direction (Atchley & Andersen, 1995; Williams & Sekuler, 1984), and speed (Watamaniuk & Duchon, 1992; Watamaniuk, Sekuler, & Williams, 1989). Efficient gist perception would be especially helpful for perceiving biological motion, which is both socially meaningful (de Gelder, 2006) and visually complex, with specialized neural mechanisms incorporating form and motion (e.g., Giese & Poggio, 2003; Grossman, Battelli, & Pascual–Leone, 2005; Grossman & Blake, 2001; Oram & Perrett, 1994; Vaina, Lemay, Bienfang, Choi, & Nakayama, 1990). These mechanisms are likely intact at birth (Simion, Regolin, & Bulf, 2008) and are quick to develop (Bertenthal, 1993; Blake, Turner, Smoski, Pozdol, & Stone, 2003). Furthermore, abnormalities in these mechanisms (Waiter et al., 2004) may be related to impairments in social function that accompany autism (Blake et al., 2003; Hubert et al., 2007). Determining if humans use ensemble coding for perceiving the behavior of crowds would be a critical step

Timothy D. Sweeny and David Whitney, Department of Psychology, and Vision Science Group, University of California – Berkeley; Steve Haroz, Department of Computer Science, University of California – Davis.

This study was supported in part by the National Institutes of Health Grants T32 EY007043 and R01 EY018216, and National Science Foundation Grant NSF 0748689.

Correspondence concerning this article should be addressed to Timothy D. Sweeny, University of California, Berkeley, Vision Science Group, 5327 Tolman Hall, Berkeley, CA 94720. E-mail: timsweeny@berkeley.edu

toward understanding fundamental mechanisms involved in normal and atypical social–behavioral development.

Our primary goal was to determine if humans use an ensemble code to perceive the average heading (i.e., the direction of walking) of a crowd of people. We directly tested this in the first experiment by restricting the number of walkers that observers could use to estimate the heading of a large crowd of people heading in different directions. We also determined whether an ensemble code was formed rapidly, and we characterized the sensitivity of the ensemble code by determining if perception of a crowd could be better than perception of an individual.

Our second goal was to determine if, by pooling information across multiple features, ensemble coding could cause a heterogeneous crowd of people to appear homogeneous. This is known to occur for perception of low-level visual features (Dakin, 2001; Dakin, Bex, Cass, & Watt, 2009; Dakin, Mareschal, & Bex, 2005; Morgan, Chubb, & Solomon, 2008; Ross & Burr, 2008; Watamaniuk & Sekuler, 1992). We directly tested this hypothesis by using a technique reminiscent of an equivalent noise analysis (e.g., Dakin et al., 2009; Morgan et al., 2008; Ross & Burr, 2008) to measure variability in perceived heading as a function of the actual heading variability within the crowd. This “crowd variability” analysis also provided an alternative demonstration of the remarkable sensitivity with which ensemble information about crowds is perceived, and it allowed us to estimate the number of individuals integrated into the ensemble code.

Experiment 1A: Ensemble Coding of a Crowd’s Heading

Observers estimated the average headings of crowds of point-light walkers (see Figure 1; e.g., Johansson, 1973). Each crowd of 12 walkers had an average heading (ranging from leftward to rightward) and a controlled amount of heading variability (i.e., walkers within a crowd had different headings). Only subsets of 2, 6, or 10 walkers were visible from the full crowd (see Methods section, below), and observers used these smaller subsets of visible walkers to make heading estimates. We hypothesized that if observers use an ensemble code that integrates information from multiple walkers, heading estimates should approach the true average of the full crowd when more walkers from the full crowd are visible in the subset. Alternatively, if observers based estimates on a single walker’s heading and not on an ensemble code, heading estimates should not change (and should remain poor) even when more walkers from the full crowd are visible in the subset.

Methods

Observers. Five experienced psychophysical observers (two naïve) gave informed consent to participate. All had normal or corrected-to-normal visual acuity and were tested individually in a dimly lit room.

Stimuli. Point-light walkers were composed of configurations of 12 white dots (each dot: 0.11° × 0.11°, 149.5 cd/m^2) presented against a black background (0.36 cd/m^2). The dots were placed at the locations of the major joints and the head such that

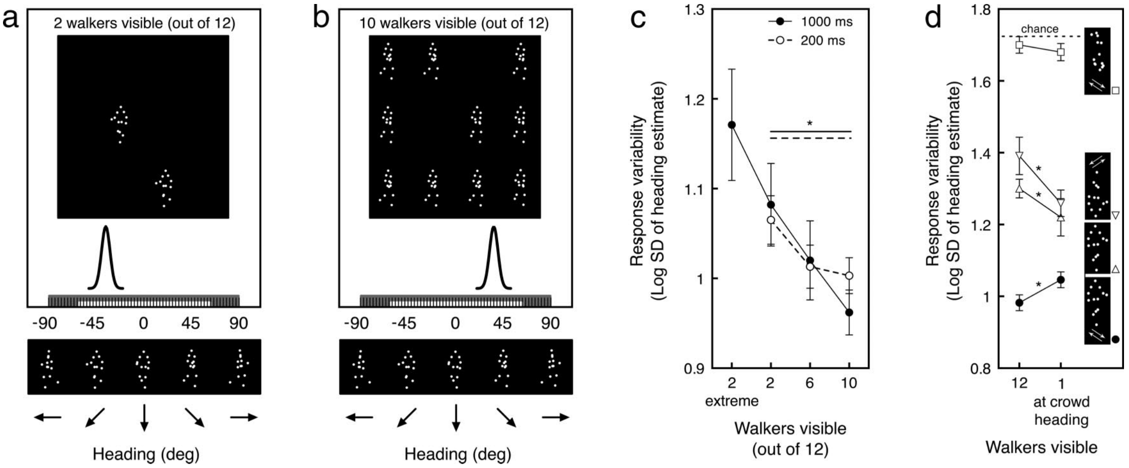

Figure 1. Conditions and results from Experiment 1. (**a** and **b**) A heterogeneous crowd of walking people was generated by sampling 12 individual headings from a Gaussian distribution centered at one of 43 headings. The white and gray bars along the x-axes indicate the range of possible crowd headings and individual headings, respectively. Only a subset of (**a**) 2, 6 (not shown), or (**b**) 10 walkers from the full crowd of 12 was visible for estimating the full crowd’s heading. (**c**) Response variability as a function of subset size with 1000-ms (closed circles, solid line) and 200-ms (open circles, dashed line) presentations. (**d**) Response variability when a full crowd of 12 was presented, or a single walker was presented at the crowd heading; walkers were either upright and coherent (closed circles), inverted (open inverted triangles), static (open triangles), or scrambled (open squares), with 200-ms presentations. Unlike the inverted, static, and scrambled conditions, the crowd of upright walkers was perceived more precisely than an individual walker heading in the same direction, suggesting that the ensemble coding of biological motion is uniquely strong and special. Error bars represent ± 1 bootstrapped *SD*. $^*p < .05$.

the overall configuration would be perceived as a human body (Johansson, 1973). We generated "videos" from sets of 21 static frames in which the local position of each dot changed from frame-to-frame in a manner which was consistent with a natural human gait (see Vanrie & Verfaillie, 2004 for a complete description of these stimuli). Each gait cycle (i.e., one step by each foot) lasted 800 ms. The application to generate the videos was written in C# and interfaced with OpenGL via the Open Toolkit Library (http://www.opentk.com). We generated 43 videos, each with a distinct heading, by rotating the 3D positions of the dots in each frame by a distinct angle around the vertical axis (i.e., the direction of walking). The headings ranged from leftward (−90°) to rightward (90°) in 3° increments (see Figure 1). The 3° increment between headings was less than the average just noticeable difference (5.778°, determined from a pilot experiment in which observers estimated the heading of a single walker presented for 1,000 ms). We limited the range to forward headings because backward headings can appear ambiguous (perceived as forward or backward; Vanrie, Dekeyser, & Verfaillie, 2004). A dot configuration with a completely leftward (−90°) or completely rightward (90°) heading subtended 1.9° × 2.91° of visual angle at the full extension of the gait cycle and 0.56° × 3.06° at the minimum extension of the gait cycle. A dot configuration with a completely forward heading (0°) subtended 1.03° × 3.06° of visual angle. Our displays did not include any depth cues; the size and surface illumination of each dot remained uniform, and overlapping dots did not provide occlusion cues. Consequently, our displays conveyed heading cues in the simplest way possible.

Crowd and walker heading selection. On a given trial, we randomly selected 12 headings from a continuous Gaussian distribution centered at one of 43 headings (ranging from −63° [strongly leftward] to 63° [strongly rightward] in 3° increments, Figure 1). The peak of the distribution determined the average heading of the crowd and the width of the distribution determined the heading variability within the crowd. The standard deviation of the sampling distribution was always the same (4°). The resulting amount of variability was noticeable; the average range (13.2°) was much larger than the average just-noticeable difference (5.78°). We used a truncated range of average crowd headings so that values from the tails of a distribution centered at −63° or 63° would not exceed −90° or 90°.

Because our stimulus set contained walkers with discrete headings (e.g., −63°, −60°, −57°, etc.), we sorted each of the 12 outputs from the continuous Gaussian distribution into 3° bins centered at the 60 possible walker headings between −90° and 90°. For example, a sampled heading of −58.6° would generate a walker with a −60° heading, and a sampled heading of −58.4° would generate a walker with a −57° heading.

Crowd and walker configurations. We presented walkers randomly placed among 12 nonoverlapping locations in a 4 × 3 grid subtending 15.6° × 8.36° of visual angle (measured from the center of each walker) with an average horizontal interwalker distance of 2.34° and an average vertical interwalker distance of 0.612°. We used an orthographic projection (i.e., discounting linear perspective such that a given walker would appear identical at any position in the grid).

Only randomly selected subsets of 2, 6, or 10 walkers were visible from the full sampled crowd of 12 on a given trial. We also included trials in which the most leftward and rightward headed walkers were shown—the *2-extreme* condition. We presented these walkers in randomly selected locations in the grid (see Figure 1). We note that although the random assignment of walkers to locations within the grid resulted in more empty space between walkers in smaller subsets, pilot results with spatially contiguous subsets confirmed that this spacing was not responsible for our findings in Experiment 1 (i.e., we obtained identical results when visible walkers were always adjacent).

Procedure. Observers initiated each trial by pressing the space bar, followed immediately by a crowd of walkers presented for 1,000 ms. Next, a black screen appeared for 1,000 ms followed by a single dynamic response walker at the center of the screen. The initial heading of the response walker was randomly chosen on each trial from a range of −90° to 90°. Observers adjusted the heading of the response walker to a value between −90° and 90° in 3° increments to match the average heading of the crowd using the right and left arrows on the keypad. The response walker remained on the screen until the observer pressed the spacebar to end the trial. This response-walker adjustment procedure smoothly altered the heading without breaking the walker's stride. An adjustment spanning the entire range of headings would have taken at least 3,200 ms, although no response required such a large adjustment. We note that although the time from the offset of the group to the end of the adjustment procedure may have introduced variability from a degraded memory trace into the recorded response, this added variability should have affected each condition equally. Furthermore, perceptual averaging has been shown to be unaffected by delays much longer than those used in this experiment (Chong & Treisman, 2005). We paired each subset size (2-extreme, 2, 6, 10) with each mean heading (43 values) for a total of 172 trials. Two observers ran in 860 trials to test the reliability and significance of the results on an individual observer level. All stimuli were presented on a 61-cm liquid crystal display monitor at a viewing distance of 102 cm.

Results

For each trial, we calculated the difference between the estimated heading of the visible subset and the actual heading of the crowd of 12. We then measured response variability as the standard deviation of the distribution of these differences. Because the walkers in each crowd varied in their individual headings, estimates based on a single walker should not change (and should remain poor) even when more of the 12 walkers from the full crowd are visible in the subset. In contrast, estimates based on an ensemble code should become more precise—that is, variability should decrease—when more of the 12 walkers from the full crowd are visible in the subset. We used a bootstrapping method (Manly, 2007) for making planned comparisons between response variability from the conditions.

Increasing the number of visible walkers clearly reduced response variability (see solid line in Figure 1C). Planned comparisons confirmed that response variability with 10 walkers was significantly less (i.e., better) than with two walkers ($p < .05$). A control condition ruled out the possibility that observers cognitively computed the average of a few select walkers; estimates using the most leftward and rightward walkers from the full crowd (the 2-extreme condition) were no better than those using two randomly selected walkers, *ns.* We found the same pattern with the two observers who completed a longer version of the experiment, both of whom showed significantly lower response variability for a crowd of 10 as compared with a crowd of 2 ($p < .05$).

Experiment 1B: Precise Ensemble Coding of a Crowd's Heading With Very Brief Presentations

To conclusively rule out a serial search strategy, we replicated Experiment 1A with extremely brief durations (200 ms). This brief duration prevented observers from making saccades to multiple walkers or using a serial search strategy to estimate the average heading. To further characterize the precision of the ensemble coding mechanism, we also compared sensitivity for a full heterogeneous crowd of 12 to that of a single walker. If integrating multiple walkers into an ensemble code averages out noise in the encoding of each walker, then perception of a crowd may be more precise than perception of an individual who is heading in the same direction as the crowd.

Methods

Stimuli and procedure. Walkers were presented for 200 ms and were backward masked by a random pattern of white dots within a rectangle with an aspect ratio comparable to that of a human configuration. We included trials showing only a single walker (with a heading equivalent to the mean of the crowd) or the full crowd of 12 intermixed with trials from each of the subset sizes (2, 6, 10). We paired each condition with each mean heading (43 values) once for a total of 215 trials.

We also tested perception of a single walker or a full crowd with static, inverted, and scrambled walkers. These control conditions allowed us to determine if any potential differences between perception of a crowd and an individual were unique to perception of upright–intact biological motion. Perception of biological motion may rely largely on the configuration of the human form (Beintema & Lappe, 2002; McLeod, Dittrich, Driver, & Zihl, 1996; Vaina et al., 1990). We thus included trials where we presented only a single randomly selected frame from each walker's gait cycle (static, with no motion—the *human configurations without motion* condition). The stage of each walker's gait within a crowd was identical in this condition. Perception of biological motion has also been suggested to rely heavily on local motion cues (Chang & Troje, 2009; Mather, Radford, & West, 1992; Thurman & Grossman, 2008; Troje & Westhoff, 2006). Inversion is known to disrupt both low-level and global–configural information in a walker (Beintema & Lappe, 2002; Gurnsey, Roddy, & Troje, 2010; Troje & Westhoff, 2006). We thus included trials with inverted walkers to determine if any potential differences between perception of a crowd and an individual would occur with a nonbiological stimulus with carefully matched low-level visual information. We also included trials where we presented moving clusters of dots without human configurations (scrambled walkers, the motion without human configurations condition). To create these scrambled walkers, we randomly positioned the location of each dot in a 3D bounding box with an aspect ratio comparable to that of a human configuration. We generated these scrambled dot locations separately for each heading (−90° through 90°) for each observer. It was crucial that the local motion of each dot in the motion without human configurations condition was centered about its randomly selected location (rather than a location on the walker's body) and was identical to the local motion with upright–coherent walkers. This preserved the local motion but disrupted the configuration of the dynamic information. Each of these control conditions was run in a separate block of 86 trials. All other experimental details were identical to those in Experiment 1A.

Because perception of biological motion most likely requires integration of form and motion (Giese & Poggio, 2003), presumably in high-level visual areas (Bonda, Petrides, Ostry, & Evans, 1996; Grossman et al., 2005; Grossman & Blake, 2001; Oram & Perrett, 1994; Vaina et al., 1990), we predicted that response variability would be greater in each of these control conditions.

Results

As with the longer presentations, response variability decreased when more walkers from the full crowd were visible for only 200 ms (see dashed line in Figure 1C). Planned comparisons confirmed that response variability with 10 walkers was significantly less (i.e., better) than with two walkers ($p < .05$). Remarkably, heading estimates of a heterogeneous crowd of 12 were even better than estimates of a single walker with a heading identical to the mean of the crowd ($p < .05$, see closed black circles in Figure 1D). This heightened crowd sensitivity was unique for perception of upright–coherent biological motion. We found the opposite pattern both for perception of inverted walkers ($p < .05$; interaction against upright walkers, $p < .05$), and for static walkers ($p < .05$, interaction, $p < .05$; Figure 1D). There was no difference between perception of a single walker and a heterogeneous crowd with point-scrambled walkers ($p = .59$), for which heading perception was near chance-level (chance log standard deviation [*SD*] = 1.72, confirmed using Monte Carlo methods).

Experiment 2: Ensemble Coding of Crowds With Different Amounts of Variability

We further characterized the sensitivity of the ensemble code by measuring response variability as a function of actual variability within the crowd [similar to an equivalent noise analysis (e.g., Dakin et al., 2009; Ross & Burr, 2008)]. The motivation for this approach is simple. Internal noise in the encoding of each walker could cause a physically homogeneous crowd to appear heterogeneous. But by pooling signals from multiple walkers into a single value that represents the entire crowd (i.e., an ensemble code), the visual system may be able to average out variability and make a crowd appear homogenous. In other words, if observers use an ensemble code to perceive a crowd's heading, then equivalent increases in the heading variability within a crowd should only cause increases in response variability after a threshold of internal noise is surpassed. This nonlinear pattern should be well fit by a crowd variability analysis, which would allow us to estimate the number of walkers integrated into the ensemble (see Crowd Variability Analysis section below for more details). This kind of approach allowed us to directly determine if ensemble coding imposed perceptual unity onto a heterogeneous crowd. This design is more direct and sensitive than an alternative approach in which we could have asked observers how variable the crowd appeared. The latter approach could be susceptible to individual differences or ambiguity in interpreting what variability means, whereas in our design, observers simply indicated the crowd's heading.

Methods

Crowd variability analysis. Random variability (i.e., noise) in neural population activity can distort the appearance of individual features (Baldassi, Megna, & Burr, 2006; Suzuki & Cavanagh, 1998; Sweeny, Grabowecky, Kim, & Suzuki, 2011). Described in signal-detection terminology, the appearance of a physically unchanging feature (i.e., a constant signal) can vary from trial to trial due to noise. To prevent this noise in individual-feature coding from distorting the appearance of a group of features, the visual system may optimize signal detection by setting an internal threshold (i.e., a criterion) based on the expected level of internal variability, and then disregarding variability in feature coding (arising from signal or noise) below this threshold. To determine encoding variability across a group of features, the visual system must generate a summary statistical representation that describes the entire group (i.e., an ensemble code). This thresholding mechanism may cause a slightly variable group of features to cohere into a perceptually homogeneous ensemble, with each feature assuming a value representative of the group [that is, involuntary summary perception (Morgan et al., 2008; Murakami & Cavanagh, 1998; Ross & Burr, 2008; Watamaniuk & Sekuler, 1992; Watt & Morgan, 1983)]. In other words, when actual feature differences are small, each member of a crowd may take on the appearance of the average of the group.

Compelling perceptual demonstrations with groups of low-level visual features verify these perceptual predictions and confirm the use of ensemble coding (Dakin, 2001; Dakin et al., 2009; Dakin et al., 2005; Morgan et al., 2008; Ross & Burr, 2008; Watamaniuk & Sekuler, 1992). For example, as long as orientation variability among an array of tilted patches remains lower than the presumed internal orientation–noise threshold, the patches appear identical even with increases in their actual orientation variability; real orientation differences are only perceived when larger than a certain amount, presumably the internal noise threshold. To fit this nonlinear increase in perceived variability, these demonstrations used the following "equivalent noise" equation:

$$\sigma^2_{\text{obs}} = \frac{\sigma^2_{\text{loc}} + \sigma^2_{\text{ext}}}{\text{N}_{\text{glo}}}$$

In an equivalent noise analysis, the observer's response variability (σ_{obs}) is limited by the local noise in feature coding (σ_{loc}), external noise in the crowd (σ_{ext}), and the number of features integrated into the ensemble code (N_{glo}; Figure 2, see the references directly above for more information).

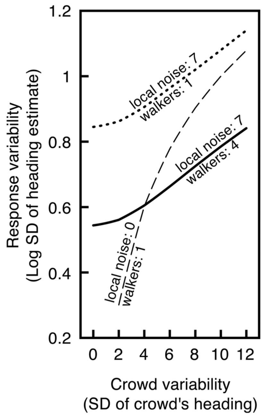

Figure 2. Illustration of response variability as a function of equivalent increases in crowd variability using the crowd variability analysis and hypothetical values of local noise and walkers incorporated into the ensemble code. Increasing local noise in heading encoding introduces an expansive nonlinearity by increasing response variability at low levels of crowd variability. Using multiple walkers to estimate a crowd's heading lowers overall variability while retaining the nonlinearity.

Here, we used a modified version of the equivalent noise analysis to measure variability in heading estimates as a function of actual variability in a crowd's heading (we refer to our modified version as a "crowd variability analysis"). Our only modification was measuring response variability across trials rather than measuring threshold for perceiving actual differences within the crowd. Consider the following illustration of our analysis on perception of a crowd's heading instead of orientation. A hypothetical observer with no local heading encoding noise (an unlikely scenario) using only a single randomly selected walker to estimate a crowd's heading would show immediate increases in response variability with equivalent increases in the crowd's actual variability (see Figure 2). This is because a single randomly selected walker is a poor representative of a heterogeneous crowd. Adding local noise to this observer's heading-encoding would increase response variability and introduce an expansive nonlinearity (i.e., a flattening of the function) for low levels of crowd variability. This nonlinearity reflects the use of an internal noise threshold (see above). In the most plausible scenario, an observer's estimate would be based on noisy encoding of multiple walkers, which lowers the overall level of response variability (by incorporating more heading information) while maintaining the nonlinearity. Of importance, fitting response variability as a function of crowd variability with this analysis allowed us to (1) demonstrate the use of an ensemble code by confirming the use of multiple walkers to make a heading estimate, (2) estimate the number of walkers pooled into the ensemble code, and (3) determine if a heterogeneous crowd appeared cohesive.

In our experiments, an observer utilizing an ensemble code should perceive the heading of a crowd with some variability as precisely as they perceive the heading of a homogeneous crowd as long as the variability in the crowd is less than their internal noise threshold. That is, adding some noise to the crowd should not impair the performance of an observer using multiple walkers, at least up to a point. This pattern of results would provide separate evidence of ensemble coding and complement our findings from Experiment 1.

Stimuli and procedure. In this experiment, all 12 walkers in a crowd were visible. Crowds had different amounts of heading variability (i.e., walkers within a crowd had identical or increasingly variable headings; see Figure 3A). The standard deviations of the sampling distribution included 0° (resulting in a homogenous group), 2°, 4°, 6°, 8°, 10°, and 12°. For visual simplicity and to allow comparisons between discrete levels of heading variability, we binned trials by requested standard deviation. Due to this

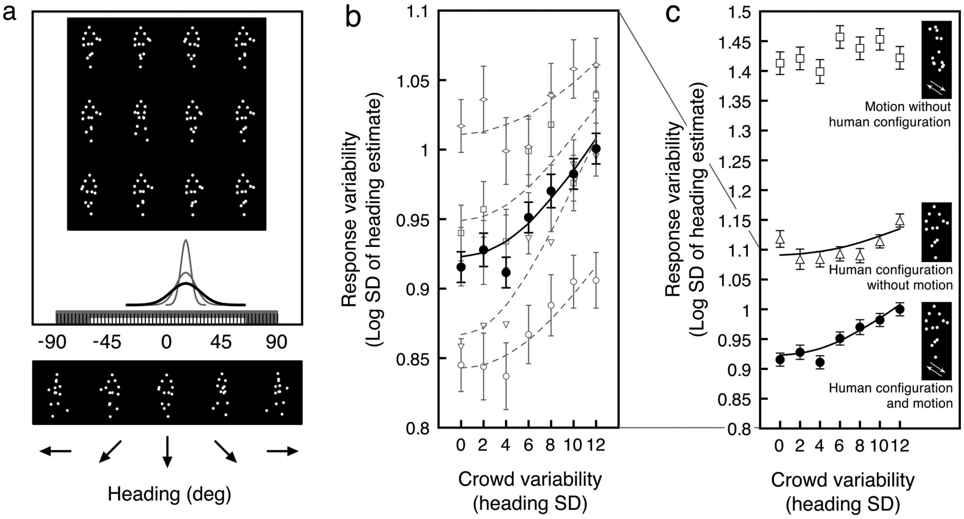

Figure 3. Conditions and results from Experiment 2 **(a)** Sampling distributions with different *SD*s (0°–12°) were used to generate crowds with increasing heading variability. Sampling distributions with *SD*s of 4° and 8° (gray distributions), and 12° (black distribution) are shown below a crowd drawn from a distribution with an *SD* of 12°. **(b)** Response variability as a function of crowd variability for perception of crowds of upright–coherent walkers. **(c)** Response variability as a function of crowd variability for perception of walkers with human configurations without motion (static crowds) or with motion without human configurations (scrambled crowds) as compared with walkers with human configurations and motion (upright–coherent crowds). The solid black lines represent the best fits of the crowd variability analysis for the average of all observers (a fit was not possible for the motion without human configurations condition). Dashed gray lines in panel b represent the best fits for individual observers. Error bars represent ± 1 bootstrapped *SD*.

sampling and binning, the means of the sampled crowds of walkers were close to, but did not exactly match the requested means. However, there was little bias in the difference of the sampled means from the requested means (e.g., mean [*M*] = −0.014, *SD* = 2.25, based on trials from a typical observer). Also because of the sampling and binning, the standard deviations of the sampled groups of walkers were close to, but did not exactly match the requested standard deviations. The sampled standard deviations for crowds with high variability tended to be slightly less than the large requested standard deviations (e.g., sampled *SD* = 11.7° for requested *SD* = 12°, based on one typical observer), probably as a consequence of the limited range of headings. These trends should not have affected our results in any systematic way since all analyses of perceived heading error were made with respect to the actual sampled means, and sampled standard deviation differences were relatively small.

We collapsed across the range of crowd headings for all analyses (although we did separately confirm that all effects were consistent for leftward, oncoming, and rightward ranges of crowd headings). On each trial, observers viewed a crowd presented for 1,000 ms and then indicated the mean heading of the crowd by adjusting a subsequently presented dynamic response walker.

The main condition included walkers that had human configurations combined with motion. This allowed us to obtain a baseline level of perceptual sensitivity against which to compare the results from our control conditions in which walkers were either static or scrambled. The static condition allowed us to determine how much perception of a full crowd of 12 walkers relied on human configurations alone. For this condition, we presented only a single randomly selected frame from each walker's gait cycle (the same stimuli from the human configurations without motion condition from Experiment 1B). The stage of each walker's gait was identical in this condition. This allowed us to determine if specific stages of the gait cycle were more useful than others for determining the average heading. The scrambled condition allowed us to determine how much perception of a full crowd relied on local dot motion alone. For this condition, we presented moving clusters of dots without human configurations (the same stimuli from the motion without human configurations condition from Experiment 1B).

For each condition, we paired each value of heading variability (seven values) with each mean heading (43 values) five times for a total of 1,505 trials run across five blocks. Each condition was run in separate groups of five blocks. Observers completed the upright–coherent condition first. Observers then completed blocks with the static or scrambled walkers. We ran the conditions in this order because we expected performance with static and scrambled crowds to be worse than with upright–coherent crowds. Thus, predicted poor performance in the static and scrambled conditions would occur despite greater expertise with the task. We used the same bootstrapping procedure from Experiment 1 for all planned comparisons.

Results

Upright–coherent crowds. Heading estimates of crowds with moderate variability were as precise as estimates of homogeneous crowds (Figure 3B); response variability was equivalent for crowds with 4° and 0° of heading variability (for the group average; $p > .32$, and each observer alone; p values > 0.23). This is despite the fact that the range of headings in the 4° standard deviation was over two times the just-noticeable-difference. This robustness to variance suggests that individual walkers that would be noticeably different on their own may cohere as a single Gestalt when in a crowd. We fit the pattern of response variability with the crowd variability analysis using the Matlab curve fitting toolbox (MathWorks, Ltd). The crowd variability analysis provided an excellent fit to the nonlinear pattern of response variability (for the average of all four observers, $R^2 = 0.923$). Because the number of walkers integrated into the estimate was a free parameter of the analysis (see Crowd Variability Analysis section, above), this fit allowed us to estimate the number of walkers integrated into the ensemble code. The best fit to the average performance of all observers was consistent with at least four walkers having been integrated into the ensemble code. The nonlinear pattern and the excellent fit confirm our findings from Experiment 1 that observers used multiple walkers to determine the heading of the crowd. Moreover, the number of walkers integrated (four) is consistent with the suggestion that an effective sample size tends to be around $\sqrt{n}$ (Dakin et al., 2009).

Static crowds. Response variability was higher with static form information as compared with dynamic form information ($p < .01$; Figure 3C). We performed a separate analysis to determine if perception of a crowd's heading was dependent on a particular stage of the gait cycle (e.g., response variability could have been lowest on trials where walkers had extended ankles; Chang & Troje, 2009; Mather et al., 1992; Rosenholtz, 1999; Thurman & Grossman, 2008; Troje & Westhoff, 2006). No clear pattern emerged across observers, suggesting that any use of a particular stage of the gait cycle, if at all, was idiosyncratic or irrelevant. While performance with static information did follow a nonlinear pattern, suggesting some use of an ensemble code and perceptual homogenization, performance was significantly worse overall (see above), and therefore cannot explain our main findings. Overall, these results show that the ensemble code with upright–coherent crowds could not have been based solely on the presence of human configurations.

Scrambled crowds. Response variability was higher with local motion alone as compared with when human configurations and motion were combined, illustrated by a significant increase for the average of all observers ($p < .01$) and even for the observer (TS) who showed the smallest numerical increase in response variability between these two conditions ($p < .01$; Figure 3B). The ensemble code with upright–coherent crowds could not have been based solely on the local motion of the dots.

Discussion

We showed that the visual system pools the movements of individual people into an ensemble code in order to perceive the average heading of a crowd. This summary statistical representation was formed with surprising speed and precision well beyond what would be expected from some previous investigations of crowd perception (Cavanagh et al., 2001; Neri et al., 1998; Thornton et al., 1998). Our results extend previous findings in which coarse differences across multiple walking people were encoded rapidly and incidentally (Thornton & Vuong, 2004). The ensemble code relied on integration of human form and local motion, suggesting that the summary representation was generated in high-level visual areas. This is consistent with previous findings that the visual system combines form and motion information to improve sensitivity (e.g., Atkinson, Dittrich, Gemmell, & Young, 2004; Bassili, 1979; Knight & Johnston, 1997; Lander, Christie, & Bruce, 1999). In general, these results may reflect the prevalence of collective animal behavior in nature, and they suggest that perceiving the behavior of a crowd as a singular unit is an important perceptual ability supported by specialized neural mechanisms.

Our demonstration of superior perception of heterogeneous crowds as compared with individuals is particularly novel and surprising, and it suggests that the ensemble coding of biological motion is uniquely strong and special. Although several investigations of ensemble coding with low- and high-level visual features have explicitly compared perception of heterogeneous crowds versus individuals [for example, orientation and low-level motion (Bulakowski, Bressler, & Whitney, 2007), and facial expressions (Haberman & Whitney, 2009)], ours is the first to significantly and consistently demonstrate better perception of crowds than of individuals. This could be because the encoding of a person's heading may be relatively noisy as compared with the encoding of other features (especially with brief presentation) and pooling multiple signals is most likely to sharpen perception when encoding of an individual feature is noisy. Furthermore, this crowd advantage is consistent with a recent demonstration showing that extraction of summary statistics is best with large sets (Robitaille & Harris, 2011). These findings are particularly interesting considering that the encoding of multiple features is traditionally thought to impair perception (Franconeri, in press).

We used a crowd variability analysis to show that increases in a crowd's heading variability only affected the perceived heading when the crowd was already noisy. This finding suggests that ensemble coding may minimize the salience of subtle differences to cause a chaotically moving crowd to cohere into a unified and visually appealing Gestalt. This imposition of perceptual cohesion on a crowd could have profound social consequences. For example, in the group-member mind tradeoff, cohesion increases the likelihood that people will attribute a collective mind and responsibility to a group, and it decreases the likelihood that people will attribute minds and accountability to individuals within the group (Waytz & Young, 2012).

We speculate that a linear pooling mechanism (Parkes et al., 2001) could account for the ensemble coding of biological motion in the current investigation. This mechanism could produce better performance with a heterogeneous crowd, as compared with an individual when encoding of an individual is particularly noisy and a sufficient number of samples from the crowd are included in the ensemble code (as we found in Experiment 1B). Temporal integration of biological motion signals has been suggested in the context of perceiving a single person's movement, with information accumulating across independent biological motion detectors up to 2,800 ms (Neri et al., 1998). We speculate that a similar linear pooling mechanism is operative with crowds of people, with the exception that it pools information very quickly (at least by 200 ms). Such a mechanism could mitigate the effect of noise (Murakami & Cavanagh, 1998; Ross & Burr, 2008) on the perception of crowds, and may provide a compelling explanation of the curious human attraction to coordinated movement (Gao, McCarthy, & Scholl, 2010). For example, ensemble coding may impose perceptual coherence onto disordered crowds of marching bands, dancers, and entertainers.

Our results build on previous investigations with static features (Haberman et al., 2009; Haberman & Whitney, 2007, 2009) by showing that ensemble coding can summarize very complex social cues conveyed by movement of the human body. While further research is necessary to reveal exactly where or how ensemble coding occurs (e.g., are all visual features ensemble coded in a single high-level area, or separately, in distinct stages of visual processing?), summary statistical encoding with social features suggests that possible loci should include, but are not limited to, high-level visual areas. More generally, our results demonstrate that humans are surprisingly well equipped to perceive emergent social information that occurs beyond the level of the individual (Helbing et al., 2000), and they suggest that this crowd perception mechanism may have developed to offer perceptual resolution beyond that which is possible when viewing individuals.

References

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

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

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

Atchley, P., & Andersen, G. J. (1995). Discrimination of speed distributions: Sensitivity to statistical properties. *Vision Research, 35,* 3131–3144.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception, 33,* 717–746.

Baldassi, S., Megna, N., & Burr, D. C. (2006). Visual clutter causes high-magnitude errors. *PLoS biology, 4,* e56. doi:10.1371/journal.pbio.0040056

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology, 37,* 2049–2058.

Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 5661–5663. doi:10.1073/pnas.082483699

Bertenthal, B. I. (1993). *Perception of biomechanical motion by infants: Intrinsic image and knowledge-based constraints.* Hillsdale, NJ: Erlbaum.

Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L., & Stone, W. L. (2003). Visual recognition of biological motion is impaired in children with autism. *Psychological Science, 14,* 151–157.

Bode, N. W., Faria, J. J., Franks, D. W., Krause, J., & Wood, A. J. (2010). How perceived threat increases synchronization in collectively moving animal groups. *Proceedings Biological Sciences/The Royal Society, 277,* 3065–3070. doi:10.1098/rspb.2010.0855

Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 16,* 3737–3744.

Bulakowski, P. F., Bressler, D. W., & Whitney, D. (2007). Shared attentional resources for global and local motion processing. *Journal of Vision, 7,* 10. doi:10.1167/7.10.10

Cavanagh, P., Labianca, A. T., & Thornton, I. M. (2001). Attention-based visual routines: Sprites. *Cognition, 80,* 47–60.

Chang, D. H., & Troje, N. F. (2009). Acceleration carries the local inversion effect in biological motion perception. *Journal of Vision, 9,* 19. doi:10.1167/9.1.19

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

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

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,* 28. doi:10.1167/9.11.28

Dakin, S. C., Mareschal, I., & Bex, P. J. (2005). Local and global limitations on direction integration assessed using equivalent noise analysis. *Vision Research, 45,* 3027–3049. doi:10.1016/j.visres.2005.07.037

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 18,* 1016–1026.

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience, 7,* 242–249. doi:10.1038/nrn1872

Franconeri, S. L. (in press). *The nature and status of visual resources* Oxford, UK: Oxford University Press.

Gao, T., McCarthy, G., & Scholl, B. J. (2010). The wolfpack effect. Perception of animacy irresistibly influences interactive behavior. *Psychological Science, 21,* 1845–1853. doi:10.1177/0956797610388814

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4,* 179–192. doi:10.1038/nrn1057

Grossman, E. D., Battelli, L., & Pascual–Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Research, 45,* 2847–2853. doi:10.1016/j.visres.2005.05.027

Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research, 41,* 1475–1482.

Gurnsey, R., Roddy, G., & Troje, N. F. (2010). Limits of peripheral direction discrimination of point-light walkers. *Journal of Vision, 10,* 15. doi:10.1167/10.2.15

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision, 9,* 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: CB, 17,* R751–R753. doi:10.1016/j.cub.2007.06.039

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

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

Hubert, B., Wicker, B., Moore, D. G., Monfardini, E., Duverger, H., Da Fonseca, D., & Deruelle, C. (2007). Brief report: Recognition of emotional and non-emotional biological motion in individuals with autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 37,* 1386–1392. doi:10.1007/s10803-006-0275-y

Im, H. Y., & Chong, S. C. (2009). Computation of mean size is based on perceived size. *Attention, Perception & Psychophysics, 71,* 375–384. doi:10.3758/APP.71.2.375

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14,* 201–211.

Joo, S. J., Shin, K., Chong, S. C., & Blake, R. (2009). On the nature of the stimulus information necessary for estimating mean size of visual arrays. *Journal of Vision, 9,* 7. doi:10.1167/9.9.7

Knight, B., & Johnston, A. (1997). The role of movement in face recognition. *Visual Cognition, 4,* 265–273.

Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory & Cognition, 27,* 974–985.

Manly, B. F. J. (2007). *Randomization, bootstrap, and Monte Carlo methods in biology.* Boca Raton, FL: Chapman & Hall/CRC.

Mather, G., Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proceedings Biological Sciences/The Royal Society, 249,* 149–155. doi:10.1098/rspb.1992.0097

McLeod, P., Dittrich, W., Driver, J., & Zihl, J. (1996). Preserved and impaired detection of structure from motion by a "motion-blind" patient. *Visual Cognition, 3,* 363–392.

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

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

Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature, 395,* 894–896. doi:10.1038/27661

Oram, M. W., & Perrett, D. I. (1994). Responses of anterior superior temporal (STPa) neurons to "biological motion" stimuli. *Journal of Cognitive Neuroscience, 6,* 99–116.

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

Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision, 11,* 18.

Rosenholtz, R. (1999). A simple saliency model predicts a number of motion popout phenomena. *Vision Research, 39,* 3157–3163.

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

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 809–813. doi:10.1073/pnas.0707021105

Sumpter, D. J. (2006). The principles of collective animal behaviour. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 361,* 5–22. doi:10.1098/rstb.2005.1733

Suzuki, S., & Cavanagh, P. (1998). A shape-contrast effect for briefly presented stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1315–1341.

Sweeny, T. D., Grabowecky, M., Kim, Y. J., & Suzuki, S. (2011). Internal curvature signal and noise in low- and high-level vision. *Journal of Neurophysiology, 105,* 1236–1257. doi:10.1152/jn.00061.2010

Thornton, I. M., Pinto, J., & Shiffrar, M. (1998). The visual perception of human-locomotion. *Cognitive Neuropsychology, 15,* 535–552.

Thornton, I. M., Rensink, R. A., & Shiffrar, M. (2002). Active versus passive processing of biological motion. *Perception, 31,* 837–853.

Thornton, I. M., & Vuong, Q. C. (2004). Incidental processing of biological motion. *Current Biology: CB, 14,* 1084–1089. doi:10.1016/j.cub.2004.06.025

Thurman, S. M., & Grossman, E. D. (2008). Temporal "bubbles" reveal key features for point-light biological motion perception. *Journal of Vision, 8,* 28. doi:10.1167/8.3.28

Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a "life detector"? *Current Biology: CB, 16,* 821–824. doi:10.1016/j.cub.2006.03.022

Vaina, L. M., Lemay, M., Bienfang, D. C., Choi, A. Y., & Nakayama, K. (1990). Intact "biological motion" and "structure from motion" perception in a patient with impaired motion mechanisms: A case study. *Visual Neuroscience, 5,* 353–369.

Vanrie, J., Dekeyser, M., & Verfaillie, K. (2004). Bistability and biasing effects in the perception of ambiguous point-light walkers. *Perception, 33,* 547–560.

Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc, 36,* 625–629.

Waiter, G. D., Williams, J. H., Murray, A. D., Gilchrist, A., Perrett, D. I., & Whiten, A. (2004). A voxel-based investigation of brain structure in male adolescents with autistic spectrum disorder. *NeuroImage, 22,* 619–625. doi:10.1016/j.neuroimage.2004.02.029

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

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.

Watamaniuk, S. N., & Sekuler, R. (1992). Temporal and spatial integration in dynamic random-dot stimuli. *Vision Research, 32,* 2341–2347.

Watt, R. J., & Morgan, M. J. (1983). The recognition and representation of edge blur: Evidence for spatial primitives in human vision. *Vision Research, 23,* 1465–1477.

Waytz, A., & Young, L. (2012). The group-member mind trade-off: Attributing mind to groups versus group members. *Psychological Science, 20,* 77–85.

Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research, 24,* 55–62.

Received January 3, 2012
Revision received April 2, 2012
Accepted April 5, 2012 ■