Title: Differential hemispheric and visual stream involvement in ensemble coding of emotion in facial crowds

Authors: Hee Yeon Im[1,2], Daniel N. Albohn[3], Troy G Steiner[3], Cody Cushing[2], Reginald B. Adams, Jr. [3], and Kestutis Kveraga[1,2]

1. Department of Radiology, Harvard Medical School
2. Athinoula A. Martinos Center for Biomedical Imaging, Department Radiology, Massachusetts General Hospital
3. Department of Psychology, The Pennsylvania State University

Correspondence to: Hee Yeon Im

A.A. Martinos Center for Biomedical Imaging

Department of Radiology,

Harvard Medical School

Massachusetts General Hospital

Charlestown, MA 02129

Email: him3@mgh.harvard.edu

Abstract [200 words]

The visual system takes advantage of redundancies in the scene by extracting summary statistics from groups of similar items. Similary, in social situations, we routinely make snap judgments of crowds of people. Reading "crowd emotion" is critical for guiding us away from danger (e.g., mass panic or violent mobs) and towards help from friendly groups. Scrutinizing each individual's expression would be too slow and inefficient. How the brain accomplishes this feat, however, remains unaddressed. Here we report a set of behavioral and fMRI studies in which participants made avoidance or approach decisions by choosing between two facial crowds presented in the left and right visual fields (LVF/RVF). Behaviorally, participants were most accurate for crowds containing task-relevant cues-avoiding angry crowds/approaching happy crowds. This effect was amplified by sex-linked facial cues (angry male/happy female crowds), and highly lateralized with greater recognition of task-congruent stimuli presented in LVF. In a related fMRI study, the processing of facial crowds evoked right-lateralized activations in the *dorsal* visual stream, whereas similar processing of single faces preferentially activated the *ventral* stream bilaterally. Our results shed new light on our understand of ensemble face coding, revealing qualitatively different mechanisms involved in reading crowd vs. individual emotion.

Keywords: crowd emotion | ensemble coding | face perception | facial expression | hemispheric lateralization

Introduction

We routinely encounter groups of people at work, school, or social gatherings. Perceiving crowd emotion guides an observer's own reactions to people in many social dynamics. For example, rapidly inferring intent to commit violence from the facial expressions of a mob on the street can allow us to escape in time and avoid potential danger, perhaps by seeking help from another group that looks friendly. Likewise, reading the general mood and receptiveness of an audience allows us to adjust our ongoing behavior, by explaining in more detail or deferring on a point, for a more efficient communication. Such extraction of the prevailing crowd state must occur rapidly, as rarely there is enough time to deliberate and analyze each group member's facial expression in a serial fashion. One way to achieve this level of efficiency is to represent the groups of faces as a higher-level description in the form of ensemble statistics (Alvarez, 2011; Cohen et al., 2016; Haberman & Whitney, 2012).

The emerging field of ensemble processing has revealed that people are adept at efficiently extracting a unified representation across groups of various feature domains (Alvarez & Oliva, 2008; Ariely, 2001; Chong & Treisman, 2003; Dakin & Watt, 1997; Halberda et al., 2006; Im & Chong, 2014; Maule & Franklin, 2015). While the advantage and efficiency of ensemble processing have been well established and recently incorporated into many cognitive models (Rensink, 2000; Wolfe et al., 2011), mechanisms supporting this ability have only begun to be explored. Behavioral and neural evidence suggests that extracting ensemble statistics (e.g., orientation, texture density, or mean size) from collections of objects engages a mechanism that is similar to scene perception, but distinct from individual object perception (Cohen et al., 2014; Im & Halberda, 2013; Reninger & Malik, 2004). For example, representation of ensembles does not require conscious perception or segmentation of representations of individual elements

(Choo & Franconeri, 2010; Haberman & Whitney, 2010; Im & Halberda, 2013). In addition, the parahippocampal place area (PPA), but not the object-sensitive lateral occipital area complex (LOC), appears to have a prominent role in discerning different textures and statistical properties in collections of objects (Cant & Xu, 2012; Cant et al., 2012). This evidence all converges on the idea that the visual system represents ensembles as independent from each of individual object representations. Most of this work on neural mechanisms of ensemble coding has only exploited a variety of objects (e.g., simple dots, texture patches, or cropped real-world objects) as visual stimuli. One recent behavioral study in prosopagnosia patients showed their intact ability to extract average identity of a crowd despite impairments in single face recognition (Leib et al., 2012), suggesting a possibility of distinct mechanisms subserving the processing of facial crowds versus an individual face. However, it still remains to be empirically tested whether and how the visual system processes crowd emotion and individual emotion differently. Here we aim to characterize the functional and neural mechanisms underlying crowd emotion processing, and compare them to those mediating processing of single emotional faces.

The network underlying perception of a single emotional face has been relatively well described. The brain regions believed to comprise the network supporting face perception include the fusiform face area (FFA; Kanwisher et al., 1997) located in the ventral temporal cortex and the posterior superior temporal sulcus (pSTS; Chao et al., 1999). This network operates in concert with a number of regions such as the amygdala, insula, orbitofrontal cortex (OFC), and precuneus/posterior cingulate cortex (PC/PCC) to support processing of facial emotion, face evaluations, and social cognition (Breiter et al., 1996; Decety et al., 2014; Fusar-Poli et al., 2009; Kircher et al., 2000; Rempel-Clower, 2007). Mediated by these extensively interacting neural

substrates, different information (e.g., identity, emotional expressions, eye gazes, etc.) can be processed and combined into a unified representation of a face that facilitates social behaviors (Adams & Kleck, 2005; Adams & Kveraga, 2015). For example, one evolutionary function of facial expressions is to signal and induce different intents and behaviors of both expressor and perceiver (Adams et al., 2006; Marsh et al., 2005). For an expressor, angry and happy expressions both signal a heightened likelihood of approach towards a perceiver. For a perceiver, however, an anger expression is more likely to elicit an avoidance reaction, and a happy expression elicits approach. When eye gaze direction is combined with a facial expression that shares this social signal value (e.g., anger and happy expressions combined with direct gaze, signaling approach), it enhances the processing fluency of both emotion and gaze direction in the face (Adams & Kleck, 2005). Associated with the different social motivations of approach and avoidance, angry and happy expressions are also thought to engage different processing biases across the left and right hemispheres such that positive, happy expressions (approach-related) have a left hemisphere processing bias, whereas negative, angry expressions (avoidance-related) tend to be processed more in the right hemisphere (Davidson, 1992, 1995; Davidson & Irwin, 1999; Silberman & Weingartner, 1986).

In the current study, we built on these previous findings, based on single face perception, by examining various influences on crowd emotions including: 1) stimulus-driven effects (valence, identity, and size of the crowd), 2) the effect of social motivations (e.g., task goals of avoidance and approach), and 3) the pattern of hemispheric specialization in reading crowd emotion, and the neural regions mediating it. We employed stimuli containing two groups of faces (single faces were also examined in a direct comparison in the fMRI study), one in the left and one in the right visual field. Because participants had to choose one of the two stimuli presented, their

decisions required relative comparison of the two. This allowed us to create task settings that were more representative of the type of affective appraisal that we are often involved in on a daily basis: e.g., looking for a more fun and friendly group of people to have a chat with in a cocktail party. Unlike the estimation task where the absolute value is judged, the answers and the ease of the decision in a comparison task can vary depending on the task goal. For example, the decision to choose to *approach* a happy crowd vs. an emotionally neutral crowd should be quite clear. However, the same comparison (happy vs. neutral) becomes more ambiguous when observers have to decide which crowd they would rather *avoid*. Thus, our stimuli and task design allowed us to examine how stimulus valence and perceivers' social motivations interact during crowd emotion perception in clear and ambiguous social situations.

In Experiment 1, we examined the effect of task goal on crowd emotion processing and the pattern of hemispheric specialization. Participants were presented two crowds of faces and asked to choose as quickly and accurately as possible which crowd of faces they would rather avoid (Exp.1A) or approach (Exp.1B). To characterize how the neural activity mediating crowd emotion processing may differ from that underlying individual emotion processing, we also conducted an fMRI study (Exp.2) in which participants viewed two stimulus types (containing crowds of faces and containing individual faces) during the avoidance task. In this study, we compared the patterns of brain activation when participants choose between two facial crowds vs. deciding between two single faces. If perception of crowd emotion relies on the same mechanism underlying single face perception, we would observe activations of the face network in the brain (e.g., FFA, STS, OFC, PC/PCC, and amygdala) during perception of crowd emotion, perhaps to a larger degree and larger extent than single face perception, given the greater complexity of the stimulus. Alternatively, if perception of crowd emotion relies on a qualitatively

distinct process, revealed neurally and behaviorally, we would expect to observe brain activations in a distinct set of brain areas and a different pattern of hemispheric lateralization from that found in individual emotion processing.

General Methods

Participants

In Experiment 1, a total of 42 undergraduate students participated: 21 subjects (12 female) participated in the avoidance task (Exp.1A) and a different cohort of 21 participants (11 female) participated in the approach task (Exp.1B). No subjects were excluded from the behavioral data analysis. In Experiment 2, a new group of 32 (18 female) undergraduate students participated. Two participants were excluded from further analyses because they made too many late responses (e.g., RTs longer than 2.5s). Thus, the behavioral and fMRI analyses for the Experiment 2 were done with a sample of 30 participants. All the participants had normal color vision and normal or corrected-to-normal visual acuity. Their informed written consent was obtained according to the procedures of the Institutional Review Board at the Pennsylvania State University. The participants received monetary compensation or a course credit.

Apparatus and stimuli

Stimuli were generated with MATLAB and Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). In each crowd stimulus (Fig.1A), either 4 or 6 morphed faces were randomly positioned in each visual field (right and left) on a grey background. Therefore, our facial crowd stimuli were comprised of either 8 or 12 faces. We used a face-morphing software (Norrkross MorphX) to create a set of 51 morphed faces from two highly intense, prototypical facial expressions of the same person for a set of six different identities (3 male and 3 female faces), taken from the

Ekman face set (Ekman & Friesen, 1976). The morphed face images were controlled for luminance, and the emotional expression of the faces ranged from happy to angry (Fig.1B), with 0 in Emotional Unit (EU) being neutral (morph of 50% happy and 50% angry), +25 in EU being the happiest (100% happy), and -25 in EU being the angriest (100% angry). Because the morphed face images were linearly interpolated between two extreme faces, they were separated from one another by EU intensity such that Face 1 was one EU happier than Face 2, and so on. Therefore, the larger the separation between any two morphed faces in EU, the easier it was to discriminate them.

Since the previous literature on averaging of other visual features showed that the range of variation is an important determinant of averaging performance (e.g., size or hue: Maule & Franklin, 2015; Utochkin & Tiurina, 2014), we kept the range of faces the same (i.e., 18 in emotional units) across the two set sizes. One of the two crowds in either left or right visual field always had the mean value of zero in emotional units, which is neutral on average, and the other had the emotional mean of +9 (very happy), +5 (somewhat happy), -9 (very angry), and -5 (somewhat angry). Thus, the sign of such offset between the emotional and neutral crowds in EU indicates the valence of the emotional crowd to compared to the neutral: The positive values indicate more positive (happier) crowd emotion compared with the neutral and the negative values indicate more negative (angrier) mean emotion.

Stimuli for the individual emotion condition (Fig.1C; only included in the fMRI study) comprised one emotional face (either angry or happy) and one neutral face from the same set of morphed face images randomly positioned in the same invisible frame surrounding the crowd stimuli in each visual field. The offsets between the emotional and neutral faces remained the same as

those in facial crowd stimuli. To ensure that the difference is not due to the confound of simply having more "stuff" in crowd emotion condition, compared to the individual emotion condition, we included scrambled faces in the individual emotion condition so that the same number of the face-like blobs were presented as in the crowd emotion condition. This ensured that any differences are not due to low-level visual differences in the stimulus displays, but rather to how many resolvable emotional faces participants had to discriminate on each trial (2 vs. 8 or 12).

On one half of the trials, the emotional stimulus (i.e., happy or angry: ± 5 or ± 9 EU away from the mean) was presented in the left visual field and the neutral stimulus was presented in the right visual field, and it was switched for the other half of the trials. Each face image subtended 2˚ x 2˚ of visual angle, and face images were randomly positioned within an invisible frame subtending 13.29˚ x 18.29˚, each in the left and right visual fields. The distance between the proximal edges of the invisible frames in left and right visual fields was 3.70˚.

Procedure

Participants in Experiment 1 sat in a chair at individual cubicles about 61 cm away from a computer with a 48 cm diagonal screen (refresh rate = 60 Hz). Participants in Experiment 2 were presented with the stimuli rear-projected onto a mirror attached to a 64-channel head coil in the fMRI scanner. Fig.1A illustrates a sample trial of the experiment. Participants were presented with visual stimuli for 1 second, followed by a blank screen for 1.5 second. The participants were instructed to make a key press as soon as possible to indicate which of the two crowds of faces or two single faces on the left or right they would rather avoid. They were explicitly informed that the correct answer was to choose either the crowd or the face showing a more negative (e.g., angrier) emotion for the avoidance task and a more positive (e.g.,

happier) emotion for the approach task. Responses that were made after 2.5 seconds were considered late and excluded from data analyses. Feedback for correct, incorrect, or late responses was provided after each response.

In Experiment 1, half of the participants performed the avoidance task and the other half performed the approach task. Experiment 1 had a 4 (emotional distance between facial crowds, -9, -5, 5, or 9) x 2 (visual field of presentation, LVF and RVF) x 2 (set size: 4 or 6 faces in each visual field) design, and the sequence of total 320 trials (20 repetitions per condition) was randomized. In Experiment 2 (fMRI), all the participants performed the avoidance task. Because we needed more trials for statistical power for fMRI data analyses and we observed no effect by the number of crowd members on crowd emotion perception (Fig.S1), we only used crowd stimuli containing 4 faces in Experiment 2. Thus, Experiment 2 had a 2 (stimulus type: crowd and individual) x 4 (emotional distance) x 2 (visual field of presentation) design and the sequence of total 512 trials (32 repetitions per conditions) was optimized for hemodynamic response estimation efficiency using the *optseq2* software (https://surfer.nmr.mgh.harvard.edu/optseq/).

fMRI data acquisition and analysis

fMRI images of brain activity were acquired using a 3 T scanner (Siemens Magnetom Prisma) located at The Pennsylvania State University Social, Life, and Engineering Sciences Imaging Center. High resolution anatomical MRI data were acquired using T1-weighted images for the reconstruction of each subject's cortical surface (TR = 2300 ms, TE = 2.28 ms, flip angle = 8˚, FoV = 256 x 256 mm^2, slice thickness = 1 mm, sagittal orientation). The functional scans were acquired using gradient-echo EPI with a TR of 2000 ms, TE of 28ms, flip angle of 52˚ and 64 interleaved slices (3 x 3 x 2 mm). Scanning parameters were optimized by manual shimming of

the gradients to fit the brain anatomy of each subject, and tilting the slice prescription anteriorly 30˚ up from the AC-PC line as described in the previous studies (Deichmann et al., 2003; Kveraga et al., 2007; Wall et al., 2009), to improve signal and minimize susceptibility artifacts in the brain regions including OFC and amygdala (Kringelbach & Rolls, 2004). We acquired 780 functional volumes per subject in four functional runs, each lasting 6.5 min.

The acquired fMRI mages were pre-processed using SPM8 (Wellcome Department of Cognitive Neurology). The functional images were corrected for differences in slice timing, realigned, corrected for movement-related artifacts, coregistered with each participant's anatomical data, normalized to the Montreal Neurological Institute template, and spatially smoothed using an isotropic 8-mm full width half-maximum Gaussian kernel. Outliers due to movement or signal from preprocessed files, using thresholds of 3 SD from the mean, 0.75 mm for translation and 0.02 radians rotation, were removed from the data sets, using the ArtRepair software (Mazaika et al., 2009). Subject-specific contrasts were estimated using a fixed-effects model. These contrast images were used to obtain subject-specific estimates for each effect. For group analysis, these estimates were then entered into a second-level analysis treating participants as a random effect, using one-sample t-tests at each voxel. The resulting contrasts were thresholded at $p < 0.001$ (uncorrected) and a minimal cluster size of 10 voxels. For visualization and anatomical labeling purposes, all group contrast images were overlaid onto the inflated group average brain, by using 2D surface alignment techniques implemented in FreeSurfer (Fischl et al. 2004).

Results

Experiment 1: behavioral study

In Experiment 1 (Fig.1A), participants viewed two crowds of faces, one in the left visual field (LVF) and one in the right visual field (RVF) for 1 second and had to make a key press rapidly to indicate which group of faces they would rather avoid (Exp.1A) or approach (Exp.1B). One visual field always contained an emotionally neutral crowd and the other visual field contained a crowd that was either a happier or angrier on average than the neutral crowd. The overall accuracies for both Experiments 1A and 1B were significantly higher than chance (avoidance task: 64.88% vs. 50%; approach task: 63.72% vs. 50%, all *p's* < 0.001), suggesting that the participants were able to extract the average crowd emotion from the two groups of faces and choose appropriately which group they would rather avoid or approach. Although accuracies for the avoidance task vs. approach task were not significantly different (64.88% vs. 63.72%: $t(40) = 1.330$, $p = 0.191$), the mean response time (RT) [1] was significantly slower for the avoidance task than for the approach task (1.17 vs. 0.98 second: $t(40) = 2.156$, $p < 0.05$, Cohen's d = 0.666), suggesting that it generally took longer for participants to make avoidance decision compared to approach decision (Fig.S1, first row). Neither accuracy nor RT was affected by the size of the crowd (4 vs. 6 faces in each crowd: second and third rows in Fig.S1), suggesting that extraction of crowd emotion does not require serial processing of each individual crowd member, but is processed in parallel. Because there was no effect of crowd size, we collapsed the data from the different crowd size conditions for further analyses.

In our morphing methods, the emotional distance between the morphed faces could be quantified based on the arbitrary values of the emotional unit (EU) number, with zero being

[1] We also conducted RT analyses using each participant's median RT. Just as mean RT, median RT was significantly slower for the avoidance task than for the approach task (1.16 second vs. 0.97 second on average: $t(40) = 1.995$, $p < 0.05$, Cohen's d = 0.632). We also confirmed that median RTs yielded the same results for all the other findings reported in this manuscript (see Supplementary Information).

neutral (e.g., 50% happy and 50% angry), +25 being extremely happy, and -25 being extremely angry (Fig.1B). Because the emotionally neutral crowd (EU of zero on average) was always presented on one side, the positive value of the emotional distance between the two crowds indicates that the other side to be compared contained a happier crowd than the neutral crowd (e.g., +9 vs. 0: very happy vs. neutral and +5 vs. 0: somewhat happy vs. neutral) and the negative value of the emotional distance indicates that the other side contained an angrier crowd than the neutral crowd (e.g., -9 vs. 0: very angry vs. neutral and -5 vs. 0: somewhat angry vs. neutral). Such separation seemed effective in systematically manipulating the difficulty of the task (Fig.1D): In both the avoidance and approach tasks, accuracy increased when the emotional distance between the two crowds being compared increased (e.g., accuracy for ±9 was higher than for ±5, all *p*'s < 0.005). A repeated-measures analysis of variance confirmed the significant main effect of the emotional distance (four levels: -9, -5, +5, and +9) on performance accuracy (avoidance task: $F(3,60) = 4.69$, $p < 0.01$, $\eta_p^2 = 0.29$; approach task: $F(3,60) = 4.644$, $p < 0.01$, $\eta_p^2 = 0.219$).

Furthermore, post-hoc contrast analyses revealed that participants were most accurate for the crowd emotion that was congruent with the task goal – to approach or avoid. That is, subjects were most accurate when comparing a very angry crowd versus a neutral crowd (emotional distance of +9) during the avoidance task (Fig.1D, red line: $F(1,20) = 12.659$, $p < 0.01$, $\eta_p^2 = 0.388$) and when comparing a very happy crowd versus a neutral crowd (emotional distance of -9) during the approach task, than comparing any other combinations (Fig.1D, green line: $F(1,20) = 18.318$, $p < 0.01$, $\eta_p^2 = 0.504$).

The RTs were not significantly different for these conditions, although there was a trend toward faster RTs for the most task-congruent comparisons: very angry vs. neutral for avoidance and very happy vs. neutral for approach, compared to any other comparisons (Fig.S2, first row). These results suggest that observers were most accurate and efficient when they had to compare a task-congruent crowd emotion with a neutral crowd, with facilitated processing of angrier crowds for the avoidance task, and of happier crowds for the approach task. Thus, it appears that motivational information systematically modulates observers' evaluation on crowd emotion: Comparing angry vs. neutral crowds allows for an easy and clear decision for the avoidance task whereas comparing the same pair of crowd emotions requires more difficult and ambiguous decision for the approach task.

Hemispheric asymmetry for crowd emotion processing: clear vs. ambiguous decisions

We next examined how the left and right hemispheres were specialized for comparing two crowds of faces when the crowd stimuli contained a clear, task-congruent cue vs. an ambiguous, task-incongruent cue. When participants judged which crowd they would avoid (Exp.1A), choosing an angry over a neutral crowd was a clear and easy social decision directly relevant to the nature of the task. On the other hand, choosing a neutral over a happy crowd introduces ambiguity into the decision, because the neutral crowd does not contain an explicit threat cue (although it is less friendly than a happy crowd). In order to assess the hemispheric asymmetry for these clear vs. ambiguous decisions on avoidance behavior based on crowd emotion, we compared the participants' accuracy when either a clear social threat cue (e.g., angry crowd) or an ambiguous threat cue (e.g., neutral crowd) was presented in LVF/RH vs. RVF/LH. As shown in Fig.2A, the accuracy was higher when a clear threat cue for avoidance task (e.g., an angry crowd) was presented in LVF than in RVF (red bars in Fig.2A). However, the pattern was completely

reversed for an ambiguous threat cue (e.g., a neutral crowd). The accuracy was higher when an ambiguous threat cue was presented in RVF rather than in LVF. This observation was confirmed by a two-way repeated-measures ANOVA with a significant interaction between the visual field containing a threat cue (LVF vs. RVF) and the type of the threat cue (clear vs. ambiguous: $F(1,20) = 6.133$, $p < 0.05$, $\eta_p^2 = 0.235$), although the main effect of the visual field ($F(1,20) = 0.818$, $p = 0.376$, $\eta_p^2 = 0.039$) and the main effect of the cue type ($F(1,20) = 0.033$, $p = 0.858$, $\eta_p^2 = 0.002$) were not significant. This result indicates hemispheric specialization in which LVF/RH presentations are superior for processing a clear social threat (an angry crowd over neutral) whereas RVF/LH presentations are superior for processing a more ambiguous social threat (a neutral crowd over a happy one) during the avoidance task (Exp.1A).

For the approach task (Exp.1B), unlike the avoidance task (Exp.1A), participants had to choose which of the two crowds (e.g., Angry vs. Neutral or Happy vs. Neutral) they would rather approach. It is important to note that the emotional valence of the “clear” social cue for the approach task is opposite to that for the avoidance task. For the approach task, choosing a happy over a neutral crowd is a clearer social decision whereas choosing a neutral over an angry crowd is a more ambiguous decision. Despite the emotional valence of a clear social cue being flipped (e.g., angry for avoidance and happy for approach task), we found a consistent pattern of hemispheric asymmetry for clear and ambiguous social decision for approach decisions: the LVF/RH was superior for the clear social cue (i.e., happy versus neutral), while the RVF/LH was superior for the ambiguous social cue (i.e., neutral versus angry). As shown in Fig.2B, the accuracy was higher when a clear, task-congruent cue (e.g., a happy crowd) was presented in LVF/RH than in RVF/LH. On the other hand, the accuracy was higher when an ambiguous cue (e.g., a neutral crowd) was presented in RVF/LH than in LVF/RH. This interaction between the

visual field containing the crowd to be chosen and the type of social cue (clear vs. ambiguous) conveyed by the crowd was confirmed by a two-way repeated measures ANOVA ($F(1,20) = 5.447$, $p < 0.05$, $\eta_p^2 = 0.232$).

Two-way ANOVAs (factors: visual field and type of cue) of participants' mean RT for the avoidance and approach tasks (Fig.S2, second and third rows) revealed that the main effect of type of cue (clear vs. ambiguous) was significant (avoidance task: $F(1,20) = 20.687$, $p < 0.001$, $\eta_p^2 = 0.506$ and approach task: $F(1,20) = 16.541$, $p < 0.001$, $\eta_p^2 = 0.479$), although the main effect of visual field (avoidance task: $F(1,20) = 0.470$, $p = 0.470$, $\eta_p^2 = 0.026$ and approach task: $F(1,20) = 1.317$, $p = 0.266$, $\eta_p^2 = 0.068$) and the interaction (avoidance task: $F(1,20) = 2.888$, $p = 0.105$, $\eta_p^2 = 0.126$ and approach task: $F(1,20) = 0.298$, $p = 0.592$, $\eta_p^2 = 0.016$) were not significant. These RT results indicate faster processing for crowds containing clear cues than for crowds containing ambiguous cues, for both the avoidance and approach tasks (Fig.S2). Importantly, our RT results also suggest that the differences in accuracy observed in all our studies are not due to a speed-accuracy trade-off.

Together, our results suggest that the right hemisphere (RH) is dominant for processing a clear, goal-congruent social cue in the context of extracting crowd emotion, preferring aversive and positive stimuli during avoidance and approach decisions, respectively, whereas the LH facilitates processing of more ambiguous social cues. Unlike the traditional framework of face processing, which posits right hemispheric lateralization for aversive or negative face stimuli, and left hemispheric preference for positive, approach-evoking stimuli (Davidson, 1992, 1995; Davidson & Irwin, 1999; Silberman & Weingartner, 1986), our data suggest that the pattern of the hemispheric lateralization for reading crowd emotion is modulated in a more flexible

manner depending on the task goal and decision uncertainty, rather than being based on stimulus valence.

Sex-specific facial cues that modulate crowd emotion perception

Because previous findings of individual face perception have documented that female- and male-specific facial features are perceptually confounded with happy and angry expressions, respectively (Adams et al., 2015; Becker et al., 2007), we investigated whether processing of crowd emotion was also modulated by sex-specific facial identity cues. We compared the accuracy for male and female facial crowd stimuli in the different tasks. Figures 3A and 3B show the accuracy on the avoidance task and approach task, plotted separately by the sex of the face images (female vs. male) and the valence of emotional face images (happy and angry) to be compared to a neutral crowd. We found that participants' responses were more accurate for happy female crowds than angry female while the opposite was true for angry male crowds, which were more accurately recognized than happy male crowds. The two-way repeated-measures ANOVA confirmed this observation. We observed a significant interaction between the sex and the emotion of the face images both in the avoidance task ($F(1,20) = 4.407$, $p < 0.05$, $\eta_p^2 = 0.181$) and in the approach task ($F(1,18) = 4.309$, $p < 0.05$, $\eta_p^2 = 0.255$), although the main effects of the stimulus sex or of the emotional valence of the face images were not significant (all *p's* > 0.160). This result suggests that integration of crowd emotion from emotional faces is also modulated by sex-specific identity cues.

Comparing the differing task demands for avoidance and for approach, we also observed a modulation by task demands. Participants were most accurate for comparing an angry male crowd vs. a neutral male crowd (Fig.3A) during the avoidance task (Exp.1A), suggesting that

facial anger and masculine features both conveyed threat cues and interacted to facilitate decisions to avoid a crowd. Conversely, participants were most accurate in comparing a happy female crowd vs. a neutral female crowd (Fig.3B) during the approach task. Finally, although the sex of the faces in our crowd stimuli modulated the perception of crowd emotion, we found that the sex of the *participants* did not influence perception of crowd emotion both in the avoidance task and the approach task (Fig.S3).

We further examined whether increased variability in facial identities per se interferes with reading of crowd emotion and tested the robustness of our effects in two additional replication and extensions experiments in which participants were presented with crowd stimuli containing a mix of different identities (see supplementary materials and Fig.S4). We replicated these effects using facial crowds with mixed identities and with new cohorts of participants (Fig.S5 and S6), confirming the robustness of our findings of task-goal dependent modulation and hemispheric asymmetry.

Experiment 2: fMRI study

Behavioral results

To investigate the neural substrates underlying ensemble coding of crowd emotion and mediating a social decision (avoidance) between two crowds of faces, we conducted an fMRI study. Participants were presented with stimuli containing either two facial crowds (shown in Fig.1A) or, as our replication and extensions studies, two single faces presented in a crowd of scrambled masks (Fig.1C). In the fMRI study, we scanned 30 participants, using only the avoidance task because of time and budgetary constraints. Participants were asked to choose rapidly which of the two facial crowds (crowd emotion condition) or which of the two single

faces (individual emotion condition) they would rather avoid, using an event-related design with crowd emotion and individual faces conditions randomly intermixed.

We first confirmed that the behavioral results on the crowd emotion condition from our fMRI study again replicated our behavioral results in Experiments 1A and 1B (Fig.1 and 2) and in the replication and extension experiments (see supplementary materials, Fig.S4). The participants' overall accuracy for crowd emotion condition in the fMRI study was 63.16%, not significantly different from that we observed from Experiment 1A (64.88%; $t(48) = -1.468$, $p = 0.149$). Critically, we again replicated the laterality effect on the avoidance task we found in Experiment 1A and in the replication and extension experiments (Supplementary Information and Fig.S6). The participants' accuracy was higher when a clear, task-congruent angry crowd was presented in LVF/RH than in RVF/LH (red bars in Fig.4A) and when an ambiguous neutral crowd was presented in RVF/LH than in LVF/RH (orange bars in Fig.4A). This was confirmed by a significant interaction between the visual field of presentation (LVF vs. RVF) and the type of a threat cue (clear vs. ambiguous) in a two-way repeated measures ANOVA ($F(1,28) = 6.357$, $p < 0.05$; main effects were not significant).

For the individual emotion condition, the overall accuracy was 65.92%, which was slightly but not significantly, higher than that for crowd emotion condition ($t(56) = -1.491$, $p = 0.106$). Even though only two faces were presented, and thus there was no need to extract the average crowd emotion, the level of accuracy for comparing two individual faces was similar to that for comparing two facial crowds. This confirms that the difference in our fMRI findings comparing crowd emotion vs. individual emotion conditions is not due to a difference in task difficulty, but reflects qualitative differences in neural processing patterns and substrates.

Importantly, in the individual emotion condition, we observed a different pattern of hemispheric lateralization from that found with crowd emotion perception. Unlike the crowd emotion condition, individual emotion condition showed that both clear (blue bars in Fig.4B) and ambiguous threat cues (cyan bars in Fig.4B) were more accurately identified when presented to the LVF/RH than to the RVF/LH. In addition, a clear threat cue (an angry face) was identified more accurately both in the LVF/RH and in RVF/LH presentations. A two-way repeated-measures ANOVA confirmed this observation: The main effects of the visual field of presentation (LVF vs. RVF: $F(1,28) = 8.193$, $p < 0.01$) and of the type of the threat cue (clear vs. ambiguous: $F(1,28) = 18.511$, $p < 0.001$) were significant, but the interaction between the visual field of presentation and emotional valence was not significant ($F(1,28) = 1.702$, $p = 0.203$). This pattern of hemispheric lateralization found in the individual emotion condition is consistent with the previous findings suggesting that affective processing in general is right lateralized, with more marked laterality effects for the negative valence (Becker et al., 2007; Borod et al., 1998; Davidson & Irwin, 1999; Silberman & Weingartner, 1986). Together, our behavioral data from the fMRI experiment support our hypothesis that perception of crowd emotion and individual emotion may engage different patterns of hemispheric specialization.

fMRI results

Distinct neural substrates for crowd emotion vs. individual emotion processing

Our goal in the fMRI experiment was first to characterize the neural substrates involved in participants' avoidance decision between two facial crowds vs. those mediating decisions between two individual faces. Fig.5 shows the brain regions activated when participants were comparing two crowds vs. comparing two individual faces (Fig.5A) and vice versa (Fig.5B), and

the complete list of activations is reported in the Tables 1 and 2. We found that comparing two facial crowds with the task goal of avoidance (crowd emotion condition, red labels in Fig.5A) showed highly lateralized cortical activations in the occipital, parietal, and frontal regions along the dorsal stream (e.g., IPL, IFG, dmPFC, MFG, and OFC) including anterior insula, in both hemispheres (Table 1, top panel). On the other hand, deciding which of two individual faces to avoid evoked higher activation in the regions along the ventral stream and in the cingulate cortex, vmPFC, and the bilateral posterior insula (Table 1, bottom panel). In particularly, the regions that showed greater activation for comparing two individual faces included the areas that have been shown to be involved in processing emotional faces and threat perception (e.g., FFA, STS, amygdala, PCC, precuneus and PHC; Aminoff et al., 2013; Breiter et al., 1996; Chao et al., 1999; Decety et al., 2014; Fusar-Poli et al., 2009; Kanwisher et al., 1997; Kircher et al., 2000; Kveraga et al., 2015; Rempel-Clower, 2007).

Hemispheric asymmetry for clear and ambiguous threat cues in facial crowds

We have replicated our novel finding on the clear hemispheric asymmetry for clear and ambiguous threat cues in crowd stimuli (see Fig.2A, 4A, and S5), using different experimental settings and different cohorts. Specifically, we showed that recognition of clear threat cues (an angry crowd compared to a neutral one) and ambiguous threat cues (a neutral crowd compared to a happy one) was facilitated when they were presented in the LVF and the RVF, respectively. Thus, we next compared patterns of brain activations when clear threat cues vs. ambiguous threat cues were presented in LVF (Fig.6A and 6B) and in RVF (Fig.6C and 6D). Fig.6A shows the brain regions activated preferentially when a clear threat (e.g., angry crowd) was presented in the LVF with a neutral crowd in RVF, compared to when an ambiguous threat (e.g., a neutral crowd) was presented in the LVF with a happy crowd in RVF. A number of brain regions, both in

the ventral and dorsal streams, as well as in the amygdala, insula, and cingulate cortex, were differentially activated by a clear threat over an ambiguous threat presented in the LVF (Table 1). However, none of the brain regions showed significantly greater activation for an ambiguous threat than for a clear threat. These results suggest an LVF/RH advantage for processing a clear threat cue over an ambiguous threat cue.

Contrasting this result with trials in which a clear threat was presented in the RVF vs. when an ambiguous threat was presented in the RVF, we found significantly greater activations only in the left inferior parietal lobule (IPL) and the right inferior frontal gyrus (IFG). The opposite contrast (an ambiguous threat in RVF – a clear threat in RVF) showed significantly higher activations in left frontal cortex, left temporal gyrus as well as subcortical regions (the caudate nucleus and thalamus). These results suggest that the RVF/LH is superior for processing an ambiguous threat cue over processing a clear threat cue. Together, our fMRI results reveal a hemispheric lateralization pattern in which processing of a clear threat cue conveyed by an angry crowd is facilitated when presented in the LVF/RH whereas processing of an ambiguous threat cue conveyed by a neutral crowd (when presented with a happy crowd) is facilitated in the RVF/LH, consistent with our behavioral findings in the avoidance task (Fig.2A, 4A, and S6).

No task-goal relevant hemispheric lateralization for processing individual emotion

When an angry face (e.g., clear threat) was presented in the LVF with a neutral face in the RVF (Fig.7A), compared to trials in which a neutral face was paired with a happy face (i.e., an ambiguous avoidance decision) in the LVF with a happy face in the RVF, we found greater activation in the FFA, pSTS, amygdala, precuneus/posterior cingulate cortex (PC/PCC) and parahippocampal cortex (PHC). These areas have been previously suggested to be involved in

face and affective processing (Aminoff et al., 2013; Breiter et al., 1996; Decety et al., 2014; Fusar-Poli et al., 2009; Gosselin et al., 2006; Kveraga et al., 2015; Rempel-Clower, 2007). Thus, processing a clear threat cue over an ambiguous threat cue is favored in the LVF/RH, consistent with the previous findings showing RH dominance for single face processing (Borod et al., 1998). For the opposite contrast (e.g., ambiguous vs. clear threat in the LVF; Fig.7B), we found differential SMA and insula activation in the left hemisphere.

For the comparison between a clear threat cue (e.g., an angry face; Fig.7C) vs. an ambiguous threat cue (e.g., a neutral face; Fig.7D) presented in the RVF, we did not observe significantly greater activation in these regions for face and emotion processing (Table 2). Together, the individual emotion condition in which participants compared two individual faces did not reveal a clear pattern of hemispheric asymmetry for processing clear vs. ambiguous threats in the LVF vs. RVF. Instead, our main finding was that the LVF/RH facilitates processing of a clear threat decision in comparing an angry face to a neutral one.

Discussion

The goal of this study was to characterize the functional and neural mechanisms that support crowd emotion processing. We had three main findings: 1) We found that when the crowd emotions judged were congruent to the task goal (e.g., avoiding angry male crowds and approaching happy female crowds), observers' performance accuracy increased; 2) There was a goal-dependent hemispheric asymmetry such that presenting stimuli to the left visual field/right hemisphere led to higher accuracy in processing a clear, task-congruent cue (e.g., angry crowd for avoidance, and a happy crowd for approach), whereas an RVF/LH presentation was superior for decisions involving an ambiguous cue (e.g., neutral crowd vs. one displaying a task-

incongruent emotion); 3) Crowd emotion processing yielded highly lateralized brain activations along the dorsal stream areas, from occipital and parietal to frontal regions whereas individual face emotion processing preferentially activated the brain areas in the ventral stream including the face network (e.g., FFA, STS, OFC, PC/PCC, and amygdala).

The goal-dependent modulation of crowd emotion processing suggests that the mechanism underlying the reading of crowd emotion is highly flexible and adaptive, allowing perceivers to focus most keenly on desired outcomes in dynamic social contexts (e.g., to avoid unfriendly crowds or to approach friendly ones). It is worth noting that neither the stimulus display nor the response characteristics changed between avoidance and approach tasks: the only difference was the decision (approach or avoid) that was mapped to the response. The same visual stimuli containing facial crowds appear to be biased differently depending on whether the task goal was to avoid or to approach. Gender also interacted with the processing of crowd emotion, in a manner relevant to the current goal. Thus, the current study provides new evidence that extrinsic (e.g., emotional expressions and sex of the crowds) and intrinsic (observers' motivation) factors mutually facilitate the reading of crowd emotion in a manner that is functionally related to the task at hand.

In the individual emotion condition, we observed the RH advantage manifesting in higher accuracy for emotional faces presented in the LVF, consistent with previous findings (Borod et al., 1998). Moreover, the accuracy was higher for angry faces than for happy faces in the LVF, indicating superior processing of angry faces over happy faces (Hansen & Hansen, 1988; Öhman et al., 2001). Consistent with the observers' accuracy, greater activations were found in the face network when an angry face was presented in the LVF/RH. Conversely, the lateralization effects

during perception of crowd emotion were task goal-dependent. The relative advantage of the RH and LH processing was reversed by the current task goal such that the RH was superior in identifying the angrier crowd during the avoidance task *and* in identifying the happier crowd during the approach task. For the RVF/LH presentations, we found the opposite pattern: accuracy was higher for implicit, ambiguous decisions, such as detecting the less threatening crowd during the avoidance task and the less friendly crowd during the approach task. Because the task-congruent emotions were anger and happiness for the avoidance and the approach tasks, respectively, our results suggest a right hemispheric bias for the task-congruent crowd emotion. To our knowledge, this is the first demonstration that emotional processing can be biased differently in RH and LH, depending on the task goal and viewers' intent. Critically, we replicated this goal-dependent hemispheric lateralization for crowd emotion processing in the three main behavioral experiments we report here and two replication experiments reported in the supplementary materials (Fig.S6), suggesting the robustness of this effect in different experimental settings and with new cohorts of participants.

Goal-dependent parallel processing is particularly useful when a large number of complex stimuli (such as a crowd of emotional faces) and competing cognitive goals tax the processing capacity of the visual system, as was the case in our task. Flexibility refers to the ability to adapt to changing conditions or demands, which requires selection of appropriate, and inhibition of inappropriate, responses (Scott, 1962). Therefore, complementary processing emphases of the left and right hemispheres enable flexible and adaptive responding optimized for the current task goal in dynamic environments (Rogers et al., 2013). For instance, RH is thought to dominate in attending to novelty, executing rapid responses, detecting behaviorally relevant sensory events, and extracting global features, whereas LH is suggested to control responses requiring

consideration of alternatives by inhibiting RH to process local features (Corbetta & Shulman, 2002; Robertson et al., 1988; Rogers, 2002; Rogers et al., 2013). The complementary functions of the LH and the RH are consistent with our fMRI results in which the task-congruent stimulus (e.g., an angry crowd) evoked greater brain activation than the implicit, ambiguous stimulus (e.g., a neutral crowd) when presented in the LVF/RH whereas the ambiguous stimulus evoked greater activation than the task-congruent stimulus presented in the RVF/LH. We suggest that processing of crowd emotion engages highly adaptive, flexible interplay between the RH and the LH with different biases for processing the task-congruent and clear features and for a more detailed analysis of an implicit, ambiguous stimulus.

Our finding on the different sets of brain areas activated during crowd emotion processing vs. individual emotion processing appears to have implications for two distinct visual streams at different levels of processing. The two parallel, but interacting visual pathways with different processing biases have been previously characterized (Freud et al., 2016; Goodale & Milner, 1992; Norman, 2002; Ungerleider & Mishkin, 1982). In the dorsal stream, the faster, coarser magnocellular pathway underlies rapid detection and comprises much of the 'vision for action (Milner & Goodale, 1995), but is also involved in perception (Freud et al., 2016; Kveraga, 2014; Kveraga et al., 2007; Kveraga et al., 2007, 2009). This "quick and dirty" processing of gist conveyed via the dorsal stream (LeDoux, 1996) would allow observers to initiate responses even before conscious perception (Goodale & Milner, 1992), which is particularly useful when detailed analysis is slow and delays can be costly due to a potential danger (e.g., being threatened by an angry crowd of people). Unlike the dorsal stream, the ventral stream is dominated by the parvocellular pathway (though some magnocellular inputs are present, Merigan & Maunsell, 1993), which subserves visual recognition, detailed analysis, memory, and

emotional content, comprising much of the 'vision for perception (Goodale & Milner, 1992)'. Our finding that processing of crowd emotion and individual emotion evoked stronger activations in the dorsal stream and ventral stream, respectively, suggests that crowd emotion and individual emotion are likely represented with emphases on different functions and consequences of vision that can be complementary to each other: global processing of crowd emotion to trigger a rapid decision based on gist vs. analysis vision for individual emotion relying on more fine-grained, reflective perception.

To conclude, here we have reported evidence for distinct mechanisms dedicated to processing of crowd emotion and individual face emotion, which are biased towards different visual streams (dorsal vs. ventral), and show different patterns of hemispheric lateralization. The differential engagement of the dorsal stream regions and the complementary functions of the left and right hemispheres both suggest that processing of crowd emotion is specialized for action execution that is highly flexible and goal-driven, allowing us to trigger a rapid and appropriate reaction to our social environment. Furthermore, we have shown that observers' goals – to avoid or approach - can exert powerful influences on the perception accuracy of crowd emotion, highlighting the importance of understanding the interplay of ensemble coding of crowd emotion and social vision.

References

Adams, R.B., Jr., Ambady, N., Macrae, C. N., & Kleck, R. E. (2006). Emotional expressions forecast approach-avoidance behavior. *Motivation & Emotion*, *30*, 179-188.

Adams, R. B., Jr., Hess, U., & Kleck, R. E. (2015). The Intersection of Gender-Related Facial Appearance and Facial Displays of Emotion. *Emotion Review, 7*, 5-13.

Adams, R. B., Jr., & Kleck, R. E. (2005). The effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion, 5*, 3-11.

Adams, R. B., Jr. & Kveraga, K. (2015). Social Vision: Functional forecasting and the integration of compound social cues. *Review of Philosophy and Psychology, 6,* 591-610.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15,* 3, 122-131.

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

Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences, 18,* 379-390.

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

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology, 92,* 179-190.

Borod, J. C., Cicero, B. A., Obler, L. K., Welkowitz, J., Erhan, H. M., Santschi, C., Grunwald, I. S., Agosti, R. M., & Whalen, J. R. (1998). Right hemisphere emotional perception: evidence across multiple channels. *Neuropsychology, 12*, 446-458.

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

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., & Rauch, S. L. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron, 17*, 875-887.

Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. *The Journal of Neuroscience, 32,* 7685-7700.

Cant, J. S., Sun, S. Z., & Xu, Y. (2015). Distinct cognitive mechanisms involved in the processing of single objects and object ensembles. *Journal of Vision, 15,* 1-21.

Chao, L. L., Martin, A., & Haxby, J. V. (1999). Are face-responsive regions selective only for faces? *NeuroReport, 10,* 2945-2950.

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

Choo, H., & Franconeri, S. L. (2010). Objects with reduced visibility still contribute to size averaging. *Attention, Perception & Psychophysics, 72,* 86-99.

Cohen, M.A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences, 19,* 324-335.

Cohen, M. A., Konkle, T., Rhee, J. Y., Nakayama, K., & Alvarez, G. A. (2014). Processing multiple objects is limited by overlap in neural channels. *Proceedings of the National Academy of Sciences, U.S.A. 111*, 8955-8960.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Neuroscience, 3,* 201-215.

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37*, 3181-3192.

Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. *Brain & Cognition, 20,* 125-151.

Davidson, R. J. (1995). Brain Asymmetry, Cerebral asymmetry, emotion, and affective style. In: Davidson R.J., Hugdahl, K, editors (MIT Press, Cambridge, MA), pp 361–387.

Davidson, R. J., & Irwin, W. (1999). The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Sciences, 3*, 11-21.

Decety, J., Skelly, L., Yoder, K. J., & Kiehl, K. A. (2014). Neural processing of dynamic emotional facial expressions in psychopaths. *Social Neuroscience, 9,* 36-49.

Deichmann, R., Gottfried, J. A., Hutton, C., & Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage, 19*, 430-441.

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

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., & Dale, A. M. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex, 14*, 11-22.

Freud, E., Plaut, D. C., & Behrmann, M. (2016). 'What' is happening in the dorsal visual pathway. *Trends in Cognitive Sciences, 20*, 773-784.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., & Allen, P. (2009). Functional atlas of emotional faces processing: A voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of Psychiatry and Neuroscience, 34,* 418-432.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15,* 20-25.

Gosselin, N., Samson, S., Adolphs, R., Noulhiane, M., Roy, M., Hasboun, D., Baulac, M., & Peretz, I. (2006). Emotional responses to unpleasant music correlates with damage to the parahippocampal cortex. *Brain, 129,* 2585-2592.

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

Haberman, J., & Whitney, D. (2012). "Ensemble Perception: Summarizing the scene and broadening the limits of visual processing." In J. Wolfe and L. Robertson (Eds.), From Perception to Consciousness: Searching with Anne Treisman. Oxford University Press.

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science, 17,* 572-576.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology, 54*, 917-924.

Im, H. Y., & Chong, S. C. (2014). Mean size as a unit of visual working memory. *Perception, 43,* 663-676.

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics, 75,* 278-286.

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

Kircher, T. T. J., Senior, C., Phillips, M. L., Benson, P. J., Bullmore, E. T., Brammer, M., Simmons, A., Williams, S. C., Bartels, M., & David, A. S. (2000). Towards a functional neuroanatomy of self-processing: effects of faces and words. *Cognitive Brain Research, 10,* 133-144.

Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Progress in Neurobiology, 72,* 341-372.

Kveraga, K. (2014). Threat perception in visual scenes: dimensions, action and neural dynamics. In K. Kveraga & M. Bar (Eds.) Scene Vision: making sense of what we see. Cambridge, MIT Press, pp. 291-307.

Kveraga, K., Boshyan, J., Adams, R. B., Jr., Mote, J., Betz, N., Ward, N., Hadjikhani, N., Bar, M., & Barrett, L. F. (2015). If it bleeds, it leads: separating threat from negativity. *Social Cognitive and Affective Neuroscience, 10*, 28-35.

Kveraga, K., Boshyan, J. & Bar, M. (2007). The magnocellular trigger of top-down facilitation in object recognition. *Journal of Neuroscience, 27*, 13232-13240.

Kveraga, K., Boshyan, J., & Bar, M., (2009). The proactive brain: Using memory-based predictions in visual recognition. In S. Dickinson, M. Tarr, A. Leonardis and B. Schiele (Eds.) Object Categorization: Computer and Human Vision Perspectives. Cambridge University Press, pp. 384-401.

Kveraga, K., Ghuman ,A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition, 65,* 145-168.

LeDoux, J. E. (1996). The Emotional Brain, Simon and Schuster, New York.

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

Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach- and avoidance-related behaviors. *Emotion, 5,* 119-124.

Maule, J., & Franklin, A. (2015). Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision, 15,* 1-18.

Mazaika P. K., Hoeft, F., Glover G. H., & Reiss A. L. (2009). Methods and software for fMRI analysis for

clinical subjects, in Poster Session Presented at the Meeting of Human Brain Mapping 2009 (San Fransisco, CA)

Merigan, W. H., & Maunsell, J. H. (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience, 16*, 369–402.

Milner, A.D., & Goodale, M.A. (1995). The Visual Brain in Action. Oxford University Press, Oxford.

Norman, J. (2002). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Journal of Behavioral and Brain Science, 25,* 73-96.

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology, 80,* 381-396.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial vision, 10,* 437-442.

Rempel-Clower, N. L. (2007). Role of orbitofrontal cortex connections in emotion. *Annals of the New York Academy of Sciences, 1121*, 72-86.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition, 7*, 17-42.

Reninger, L. W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision Research, 44*, 2301-2311.

Rogers, L. J., Vallortigara, G., & Andrew, R. J. (2013). Divided brains: The biology and behavior of brain asymmetries. Cambridge, UK: Cambridge University Press.

Rogers, L. J. (2002). Advantages and disadvantages of lateralization. In: L. J. Rogers & R. J. Andrew (eds.), *Comparative Vertebrate Lateralization*, Cambridge: Cambridge University Press, pp. 126-153.

Robertson, L. C., Lamb, M. R., & Knight, R. T. (1988). Effects of lesions of temporal- parietal junction on perceptual and attentional processing in humans. *Journal of Neuroscience, 8,* 3757-3769.

Silberman, E. K., & Weingartner, H. (1986). "Hemispheric lateralization of functions related to emotion". *Brain and Cognition, 5*, 322-353.

Scott, W. A. (1962). Cognitive complexity and cognitive flexibility. *American Sociological Association, 25*, 405-414.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In “Analysis of Visual Behavior” (D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield, eds.), pp. 549-586. MIT Press: Cambridge, MA.

Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: a reply to Marchant, Simons, and De Fockert. *Acta Psychologica, 146*, 7-18.

Wall, M. B., Walker, R., & Smith, A. T. (2009). Functional imaging of the human superior colliculus: an optimised approach. *Neuroimage, 47*, 1620-1627.

Wolfe, J., Vo, M. L-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and non-selective pathways. *Trends in Cognitive Sciences, 15*, 77-84.

Acknowledgments

This work was supported by the National Institutes of Health R01MH101194 to K.K. and to R.B.A., Jr.

Table captions

Table 1: The list of regions of increased activation associated showing greater activation for Clear vs. Ambiguous threat cues when presented in LVF vs. RVF, for Crowd emotion condition.

Table 2: The list of regions of increased activation associated showing greater activation for Clear vs. Ambiguous threat cues when presented in LVF vs. RVF, for Individual emotion condition.

Figure captions

Fig.1: A sample trial, sample face images, and the results from Experiment 1. (A) A sample trial of Experiment 1. (B) Some examples of 51 morphed faces from two extremely happy and angry faces of the same person, with Face +25 in Emotional Unit (EU) being extremely happy, Face 0 being neutral, and Face -25 being extremely angry. (C) A sample stimulus of individual emotion condition in Experiment 2 (included in the fMRI study). (D) The effect of the similarity in average emotion between facial crowds on crowd emotion processing: Participants' accuracies on Experiment 1A (avoidance task, red line) and Experiment 1B (approach task, green line) are plotted as a function of the emotional distance in EU between two facial crowds to be compared. The error bars indicate the standard error of the mean (SEM).

Fig.2: The task-goal dependent hemispheric asymmetry. (A) Participants' accuracy during the avoidance task in Experiment 1A is plotted for the extent of ambiguity of the social cue conveyed in an emotional facial crowd separately (red bars for Clear vs. orange bars for Ambiguous), when the emotional crowd was presented in the LVF/RH vs. in the RVF/LH. Note that an *angry* crowd conveys a clear cue and a neutral crowd conveys an ambiguous cue in the avoidance task. The error bars indicate SEM. (B) Participants' accuracy during the approach task in Experiment 1B is plotted for the extent of ambiguity of the social cue conveyed in an emotional facial crowd for Clear vs. Ambiguous separately, when the emotional crowd was presented in the LVF/RH vs. in the RVF/LH. Note that a *happy* crowd conveys a clear cue and a neutral crowd conveys an ambiguous cue in the approach task. The error bars indicate SEM.

Fig.3: The effect of the sex-specific identity cue of facial crowds on crowd emotion perception. (A) Participants' accuracy for the avoidance task (Experiment 1A) for sex of facial crowds (pink

bars for female crowds vs. blue bars for male crowds) and for the emotional valence of an emotional crowd (Angry vs. Happy). Angry female crowds were identified most accurately in the avoidance task. (B) Participants' accuracy for the approach task (Experiment 1B). In the avoidance task, happy female crowds were identified most accurately.

Fig.4: Different patterns of hemispheric asymmetry for crowd emotion processing and individual emotion processing. (A) Participants' accuracy for the crowd emotion condition in Experiment 2 (fMRI study) when they decided which crowd they would avoid. We found the same pattern of hemispheric asymmetry as in Experiment 1A: Clear threat cue (an angry crowd) was more accurately identified in the LVF whereas an ambiguous threat cue (a neutral crowd) was more accurately identified in the RVF. (B) From the individual emotion condition, different pattern of hemispheric asymmetry was observed: both clear and ambiguous cues were more accurately identified in the LVF/RH than in the RVF/LH. Moreover, both the LVF/RH and the RVF/RH showed better accuracy for a clear threat cue than an ambiguous threat cue.

Fig.5: Different sets of regions that were activated during crowd emotion processing and individual emotion processing. (A) The brain areas that showed greater activation when participants were making avoidance decision by comparing two crowds, than comparing two single faces. (B) The brain areas that showed greater activation when participants were making avoidance decision by comparing two individual faces, than comparing two single crowds.

Fig.6: The effects of the degree of ambiguity conveyed in an emotional crowd (Clear vs. Ambiguous) and the visual field of presentation of the emotional crowd (LVF/RH vs. RVF/LH). (A) Brain activation greater when the *LVF/RH* contained a *clear* threat cue conveyed in an angry

crowd than an ambiguous threat cue conveyed in a neutral crowd. (B) Brain activation greater when the *LVF/RH* contained an *ambiguous* threat cue conveyed in a neutral crowd than a clear threat cue conveyed in an angry crowd. (C) Brain activation greater when the *RVF/LH* contained a *clear* threat cue conveyed in an angry crowd than an ambiguous threat cue conveyed in a neutral crowd. (D) Brain activation greater when the *RVF/LH* contained an *ambiguous* threat cue conveyed in a neutral crowd than a clear threat cue conveyed in an angry crowd.

Fig.7: The effects of the degree of ambiguity conveyed in an emotional face (Clear vs. Ambiguous) and the visual field of presentation of the emotional face (LVF/RH vs. RVF/LH). (A) Brain activation greater when the *LVF/RH* contained a *clear* threat cue conveyed in a single angry face than an ambiguous threat cue conveyed in a neutral face. (B) Brain activation greater when the *LVF/RH* contained an *ambiguous* threat cue conveyed in a single neutral face than a clear threat cue conveyed in an angry face. (C) Brain activation greater when the *RVF/LH* contained a *clear* threat cue conveyed in a single angry face than an ambiguous threat cue conveyed in a neutral face. (D) Brain activation greater when the *RVF/LH* contained an *ambiguous* threat cue conveyed in a single neutral face than a clear threat cue conveyed in an angry face.

Table 1.

Activation location	MNI coordinates			t value
CROWD: Clear (LVF) > Ambiguous (LVF)	**x**	**y**	**z**	
L medial orbitofrontal cortex (mOFC)	-6	29	-20	3.931
	-24	29	-14	4.317
R medial orbitofrontal cortex (mOFC)	12	47	-12	5.062
L middle frontal gyrus (MFG)	-42	26	46	3.844
R middle frontal gyrus (MFG)	36	5	44	4.990
	45	38	0	3.967
	33	2	52	3.535
L dorsolateral prefrontal cortex (dlPFC)	-48	44	-2	3.850
	-45	35	22	3.686
R dorsolateral prefrontal cortex (dlPFC)	42	35	20	4.895
	48	44	-2	3.902
R postcentral gyrus	45	-25	38	4.665
L inferior temporal gyrus (ITG)	-42	-31	-18	4.611
R inferior temporal gyrus (ITG)	45	-16	-20	3.665
	60	-16	-30	4.159
R middle temporal gyrus (MTG)	69	-49	-4	4.480
L amygdala	-21	-4	-20	3.522
R amygdala	30	-4	-24	4.456
L supplementary motor area (SMA)	-36	-1	34	4.184
R supplementary motor area (SMA)	41	5	44	3.928
R angular gyrus	42	-58	34	4.168
L cerebellum	-3	-46	-10	4.087
L thalamus	-12	-28	-2	3.856
R thalamus	9	-28	-2	3.587
L inferior frontal gyrus (IFG)	-54	14	34	3.723
R inferior frontal gyrus (IFG)	51	11	32	3.449
	57	17	14	3.634
L lateral orbitofrontal cortex (lOFC)	-33	32	-4	3.120
R lateral orbitofrontal cortex (lOFC)	33	20	-24	3.959
	39	41	-4	4.524
L inferior occipital gyrus (IOG)	-48	-79	-10	3.601
R posterior cingulate cortex (PCC)	6	-40	36	3.125
L rostolateral prefrontal cortex (rPFC)	-15	68	10	4.403
L orbitofrontal cortex (OFC)	-24	29	-14	4.317
L temporal pole	-33	8	-36	4.055
L insula	-36	-16	14	3.885
R insula	33	-4	12	3.767
R inferior parietal lobule (IPL)	51	-61	44	3.420
L putamen	-27	2	2	3.619
L extrastriate cortex (BA19)	-48	-79	-10	3.601
CROWD: Ambiguous (LVF) > Clear (LVF)	**x**	**y**	**z**	
None				

Activation location	MNI coordinates			t value
CROWD: Clear (RVF) > Ambiguous (RVF)	**x**	**y**	**z**	
R inferior frontal gyrus (IFG)	36	5	28	3.571
L inferior parietal lobule (IPL)	-42	-52	64	3.459
L superior occipital gyrus (SOG)	-15	-100	14	3.404
CROWD: Ambiguous (RVF) > Clear (RVF)	**x**	**y**	**z**	
L rostolateral prefrontal cortex (rPFC)	-3	65	20	4.524
L medial orbitofrofrontal cortex (mOFC)	-4	44	-28	3.479
L inferior temporal gyrus (ITG)	-51	-10	-22	3.566
L middle temporal gyrus (MTG)	-51	14	-22	3.498
L caudate nucleus	-9	14	14	3.862
R caudate nucleus	9	11	12	3.452
L thalamus	-3	-22	-2	3.421
R thalamus	9	-1	4	3.418

Table 2.

Activation location	MNI coordinates			t value
IND: Clear (LVF) > Ambiguous (LVF)	**x**	**y**	**z**	
L amygdala	-15	2	-22	4.227
R fusiform face area (FFA)	45	-55	-12	3.789
R fusiform gyrus (FG)	36	-46	-4	3.559
L parahippocampal cortex (PHC)	-24	-28	-18	3.768
R parahippocampal cortex (PHC)	27	-21	-29	3.516
L posterior cingulate cortex (PCC)	-21	-49	5	3.639
R extrastriate cortex (BA18)	30	-100	8	3.401
L paracentral lobule	-6	-31	66	3.402
R posterior superior temporal sulcus (pSTS)	48	-49	14	3.438
L thalamus	-9	-31	12	3.524
L precentral gyrus	-60	-1	42	3.587
IND: Ambiguous (LVF) > Clear (LVF)	**x**	**y**	**z**	
L supplementary motor area (SMA)	-12	11	56	4.324
L thalamus	-9	-10	-2	3.406
L insula	-30	17	-4	3.423
IND: Clear (RVF) > Ambiguous (RVF)	**x**	**y**	**z**	
L inferior parietal lobule (IPL)	-54	-55	44	3.430
R insula	36	-4	-8	3.636
L dorsolateral prefrontal cortex (dlPFC)	-36	11	54	3.487
R posterior cingulate cortex (PCC)	12	-43	48	3.496
IND: Ambiguous (RVF) > Clear (RVF)	**x**	**y**	**z**	
L caudate nucleus	-12	20	-6	3.924
R caudate nucleus	6	11	-8	4.364
L lateral orbitofrontal cortex (lOFC)	-33	16	-22	-3.475
R postcentral gyrus	66	-16	40	3.589
R extrastriate cortex (BA18)	36	-97	-10	3.585

Fig.1

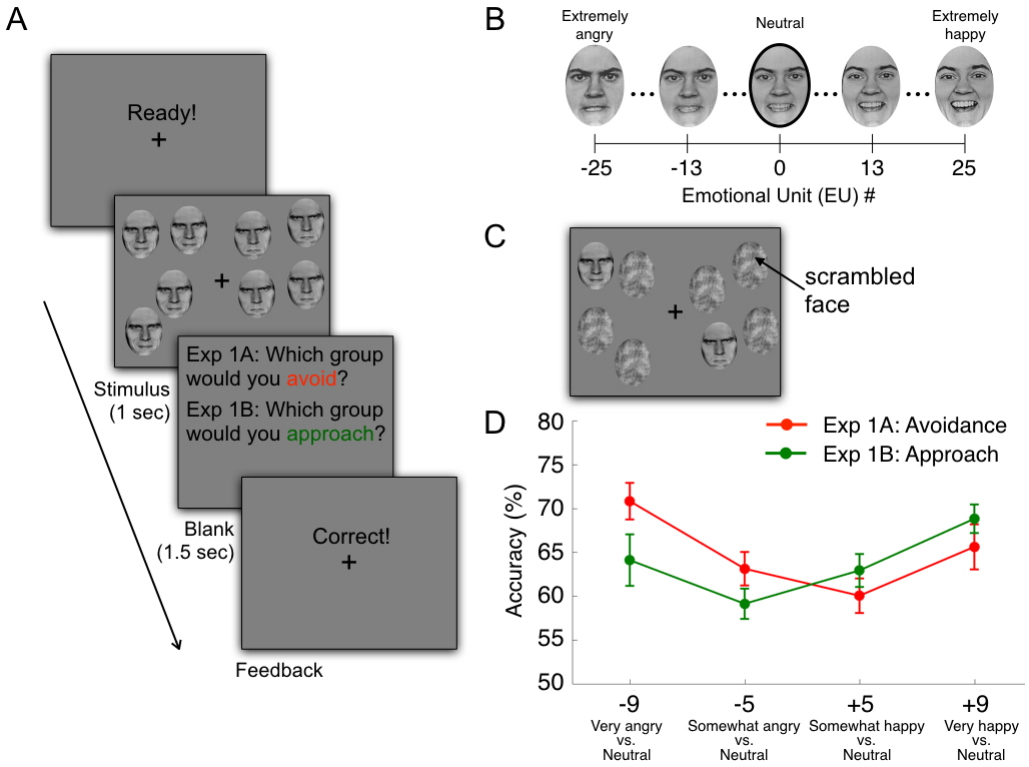

A
Ready!
+
Stimulus
(1 sec)
Exp 1A: Which group would you avoid?
Exp 1B: Which group would you approach?
Blank
(1.5 sec)
Correct!
+
Feedback
B
Extremely angry
Neutral
Extremely happy
-25
-13
0
13
25
Emotional Unit (EU) #
C
scrambled face
D
Exp 1A: Avoidance
Exp 1B: Approach
Accuracy (%)
80
75
70
65
60
55
50
-9
Very angry vs. Neutral
-5
Somewhat angry vs. Neutral
+5
Somewhat happy vs. Neutral
+9
Very happy vs. Neutral

Fig.2

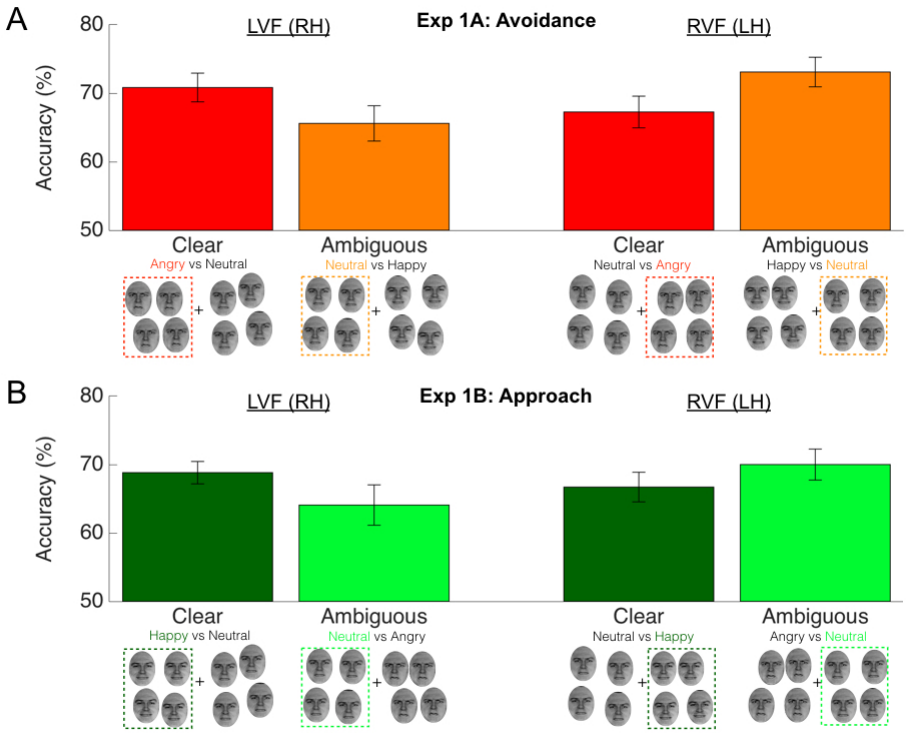

A
Exp 1A: Avoidance
LVF (RH)
RVF (LH)
Accuracy (%)
80
70
60
50
Clear
Angry vs Neutral
Ambiguous
Neutral vs Happy
Clear
Neutral vs Angry
Ambiguous
Happy vs Neutral
B
Exp 1B: Approach
LVF (RH)
RVF (LH)
Accuracy (%)
80
70
60
50
Clear
Happy vs Neutral
Ambiguous
Neutral vs Angry
Clear
Neutral vs Happy
Ambiguous
Angry vs Neutral

Fig.3

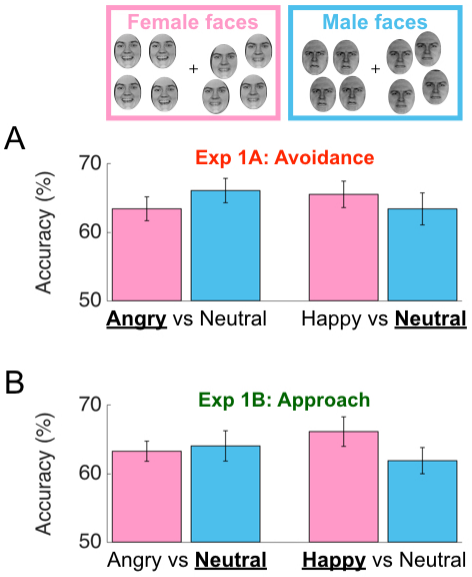

Female faces
Male faces
A
Exp 1A: Avoidance
Accuracy (%)
70
60
50
Angry vs Neutral
Happy vs Neutral
B
Exp 1B: Approach
Accuracy (%)
70
60
50
Angry vs Neutral
Happy vs Neutral

Fig.4

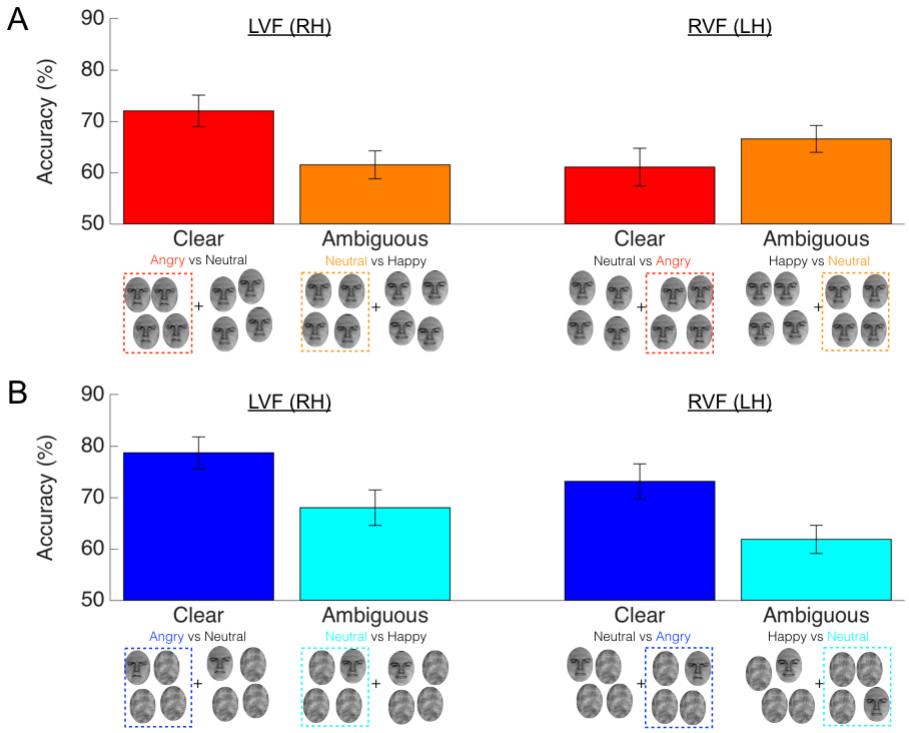

A
LVF (RH)
RVF (LH)
Accuracy (%)
90
80
70
60
50
Clear
Ambiguous
Clear
Ambiguous
Angry vs Neutral
Neutral vs Happy
Neutral vs Angry
Happy vs Neutral
B
LVF (RH)
RVF (LH)
Accuracy (%)
90
80
70
60
50
Clear
Ambiguous
Clear
Ambiguous
Angry vs Neutral
Neutral vs Happy
Neutral vs Angry
Happy vs Neutral

Fig.5

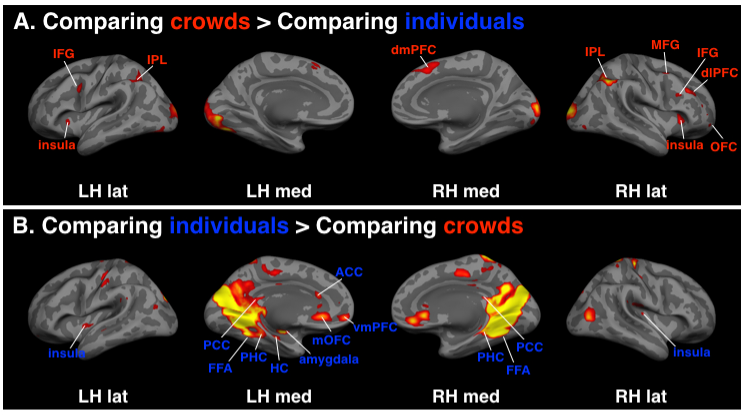

A. Comparing crowds > Comparing individuals
IFG
IPL
insula
dmPFC
IPL
MFG
IFG
dlPFC
insula
OFC
LH lat
LH med
RH med
RH lat
B. Comparing individuals > Comparing crowds
insula
ACC
vmPFC
PCC
mOFC
PHC
HC
amygdala
FFA
PHC
PCC
FFA
insula
LH lat
LH med
RH med
RH lat

Fig.6

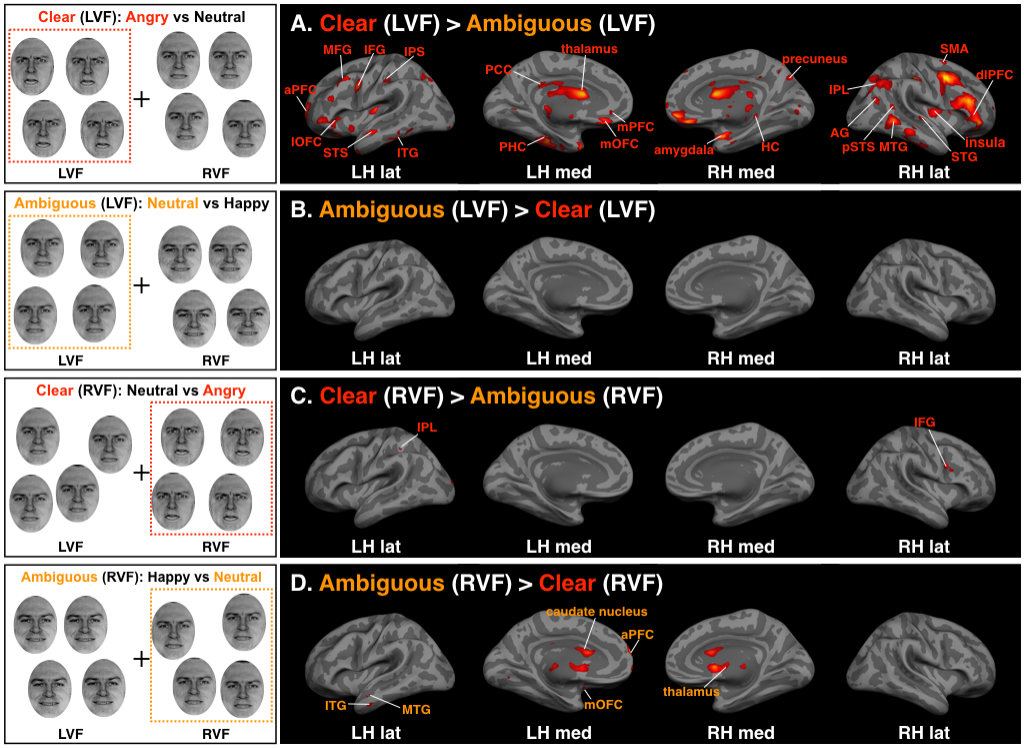
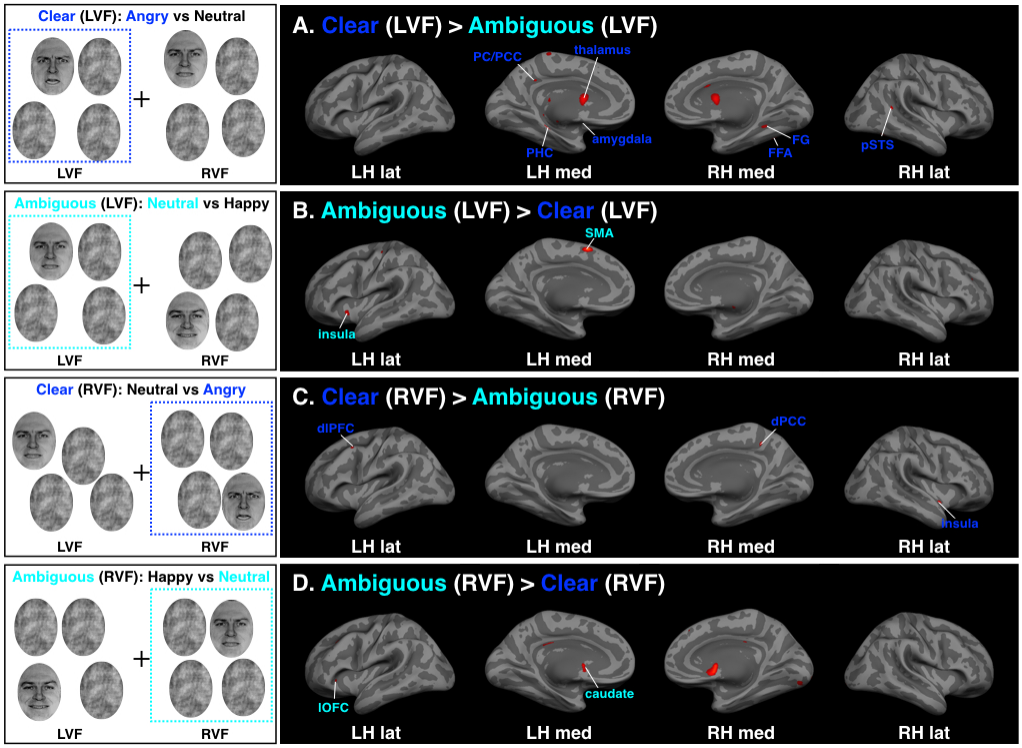

Clear (LVF): Angry vs Neutral
LVF
RVF
A. Clear (LVF) > Ambiguous (LVF)
MFG
IFG
IPS
aPFC
lOFC
STS
ITG
thalamus
PCC
mPFC
PHC
mOFC
precuneus
amygdala
HC
SMA
dlPFC
IPL
AG
pSTS
MTG
insula
STG
LH lat
LH med
RH med
RH lat
Ambiguous (LVF): Neutral vs Happy
B. Ambiguous (LVF) > Clear (LVF)
Clear (RVF): Neutral vs Angry
C. Clear (RVF) > Ambiguous (RVF)
IPL
IFG
Ambiguous (RVF): Happy vs Neutral
D. Ambiguous (RVF) > Clear (RVF)
caudate nucleus
aPFC
ITG
MTG
mOFC
thalamus

Fig.7


Clear (LVF): Angry vs Neutral
LVF
RVF
A. Clear (LVF) > Ambiguous (LVF)
PC/PCC
thalamus
amygdala
PHC
FG
FFA
pSTS
LH lat
LH med
RH med
RH lat
Ambiguous (LVF): Neutral vs Happy
LVF
RVF
B. Ambiguous (LVF) > Clear (LVF)
SMA
Insula
LH lat
LH med
RH med
RH lat
Clear (RVF): Neutral vs Angry
LVF
RVF
C. Clear (RVF) > Ambiguous (RVF)
dlPFC
dPCC
Insula
LH lat
LH med
RH med
RH lat
Ambiguous (RVF): Happy vs Neutral
LVF
RVF
D. Ambiguous (RVF) > Clear (RVF)
lOFC
caudate
LH lat
LH med
RH med
RH lat

Supplementary Information

Differential hemispheric and visual stream involvement in ensemble coding of emotion in facial crowds

Authors: Hee Yeon Im[1,2], Daniel N. Albohn[3], Troy G Steiner[3], Cody Cushing[2], Reginald B. Adams, Jr. [3], Kestutis Kveraga[1,2]

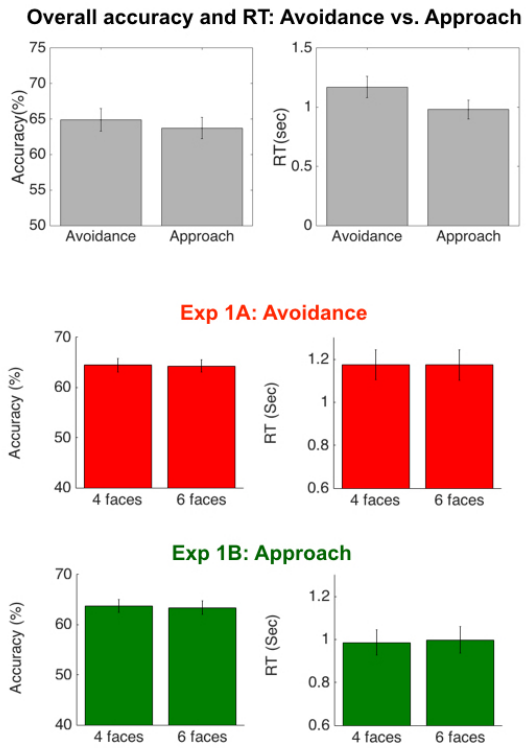
1. Department of Radiology, Harvard Medical School

2. Athinoula A. Martinos Center for Biomedical Imaging, Department Radiology, Massachusetts General Hospital

3. Department of Psychology, The Pennsylvania State University

S1. No effect of the number of crowd members on crowd emotion perception

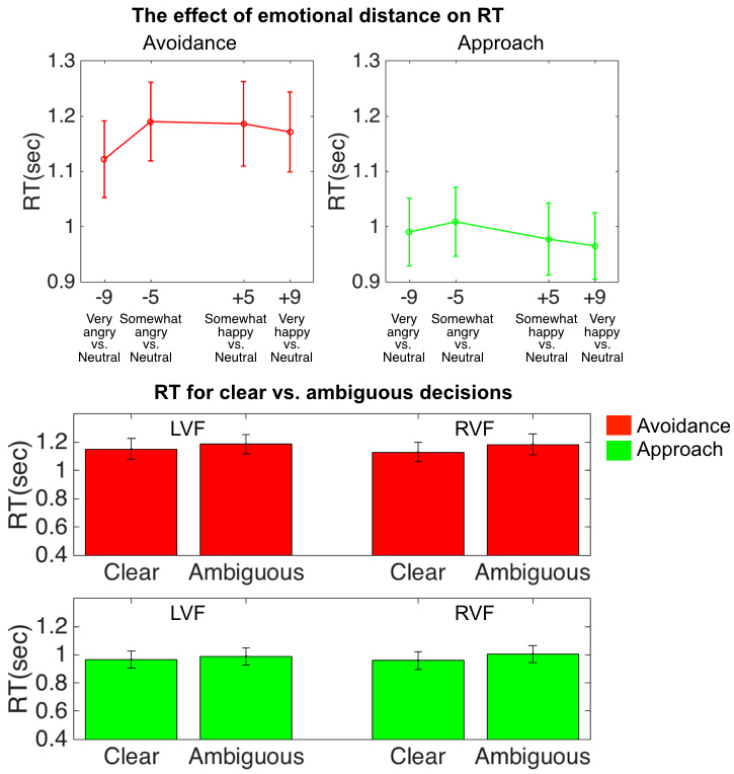
We compared the participants' accuracy and response time (RT) for crowd stimuli containing 4 faces in each visual field (8 total) vs. crowd stimuli containing 6 faces in each visual field (12 total). Although the overall accuracies for avoidance vs. approach tasks were not significantly different, RT for avoidance task was significantly slower than that for approach task (also see the main text). Furthermore, as shown in the second (Experiment 1A, avoidance) and third (Experiment 1B, approach) rows, we found that neither the accuracy nor the response time (RT) was effected by the number of crowd members (4 vs. 6 faces in each crowd) both in Experiment 1A (avoidance task: $t(40) = 0.113$, $p > .250$, Cohen's d = 0.035 for Accuracy, $t(40) = 0.010$, $p > .250$, Cohen's d = 0.003 for RT) and in Experiment 1B (approach task: $t(40) = 0.212$, $p > .250$, Cohen's d = 0.069 for Accuracy, $t(40) = -0.130$, $p > .250$, Cohen's d = -0.042 for RT)[2]. These results suggest that extraction of average crowd emotion does not necessarily require serial processing of each individual crowd member, but can be done in parallel.



[2] The median RTs showed the same pattern: the difference in median RTs between set sizes was not significant: $t(40) = 0.030$, $p = 0.998$ for avoidance task and $t(40) = 0.187$, $p = 0.953$ for approach task.

S2. RT results

We assessed the effect of the emotional distance on participants' response times (RTs) in the avoidance and approach tasks. Two-way ANOVA with factors of task type (avoidance and approach) and emotional distance (-9,-5, +5, and +9) revealed a significant main effect of the task type ($F(1,160) = 10.915$, $p < 0.01$), although neither main effect of the emotional distance ($F(3,160) = 1.397$, $p = 0.210$) and nor the interaction ($F(3,160) = 1.306$, $p = 0.274$) were significant[3]. However, we observed a trend of the RTs towards being faster for the most task-congruent emotion (i.e., clear cue, -9: very angry vs. neutral for Avoidance task and +9: very happy vs. neutral for Approach task). Two-way ANOVA analyses (factors: visual field and type of cue) on participants' mean RT for the avoidance and approach tasks revealed that the main effect of type of cue (clear vs. ambiguous) was significant (avoidance task: $F(1,20) = 20.687$, $p < 0.001$, $\eta_p^2 = 0.506$ and approach task: $F(1,20) = 16.541$, $p < 0.001$, $\eta_p^2 = 0.479$), although the main effect of visual field (avoidance task: $F(1,20) = 0.470$, $p = 0.470$, $\eta_p^2 = 0.026$ and approach task: $F(1,20) = 1.317$, $p = 0.266$, $\eta_p^2 = 0.068$) and the interaction (avoidance task: $F(1,20) = 2.888$, $p = 0.105$, $\eta_p^2 = 0.126$ and approach task: $F(1,20) = 0.298$, $p = 0.592$, $\eta_p^2 = 0.016$) were not significant[4].

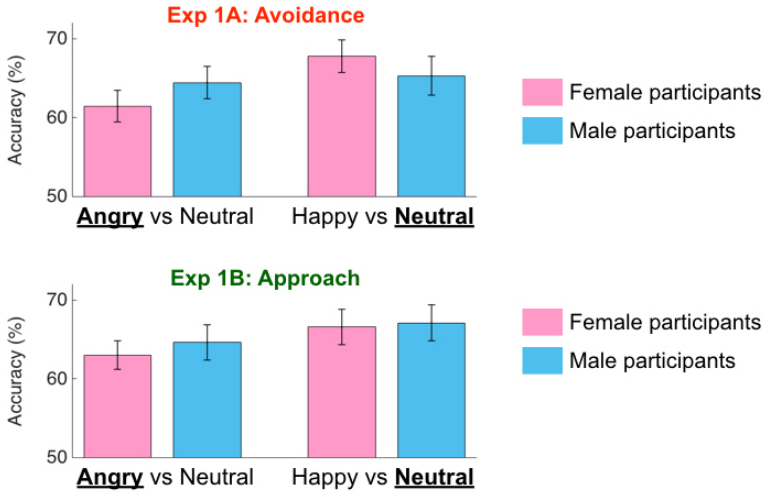


[3] For median RTs, we also found the significant main effect of the task type ($F(1,160) = 3.681$, $p < 0.05$), although the main effect of the emotional distance ($F(3,160) = 1.132$, $p = 0.268$) and the interaction ($F(3,160) = 0.986$, $p = 0.289$) were not significant.

[4] The results remained the same for median RTs as well: The main effect of type of cue (clear vs. ambiguous) was significant (avoidance task: $F(1,20) = 16.690$, $p < 0.001$, $\eta_p^2 = 0.455$ and approach task: $F(1,20) = 11.958$, $p < 0.01$, $\eta_p^2 = 0.399$), although the main effect of visual field (avoidance task: $F(1,20) = 0.730$, $p = 0.403$, $\eta_p^2 = 0.035$ and approach task: $F(1,20) = 0.177$, $p = 0.679$, $\eta_p^2 = 0.010$) and the interaction (avoidance task: $F(1,20) = 0.446$, $p = 0.512$, $\eta_p^2 = 0.022$ and approach task: $F(1,20) = 1.435$, $p = 0.246$, $\eta_p^2 = 0.074$) were not significant.

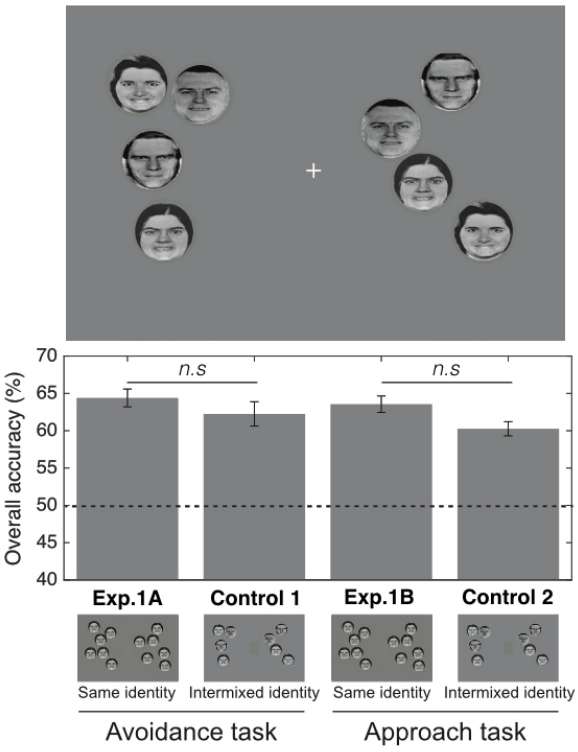
S3. No effect of the sex of participants

In Experiment 1A, 12 female and 9 male participants performed the avoidance task. We found no effect by the sex of participants on the accuracy for comparing an angry crowd vs. a neutral crowd and for comparing a neutral crowd vs. a happy crowd during the avoidance task (Fig.S3, a top panel). None of the main effects of the sex of participants ($F(1,38) = 0.014$, $p = 0.908$), the valence of the emotional crowd stimuli ($F(1,38) = 2.695$, $p = 0.109$), or the interaction ($F(1,38) = 1.552$, $p = 0.221$) was significant. In Experiment 1B, 11 female and 10 male participants performed the approach task. As shown in Fig.S3 (bottom panel), we found the same pattern in Experiment 1B: No significant main effect of the sex of participants ($F(1,38) = 0.105$, $p = 0.748$) or of the emotional valence of the crowd stimuli ($F(1,38) = 0.967$, $p = 0.332$), and no significant interaction ($F(1,38) = 0.721$, $p = 0.401$). Thus, the sex of participants did not affect the processing of crowd emotions, despite that the sex of face images influenced crowd emotion perception (as described in the results section).



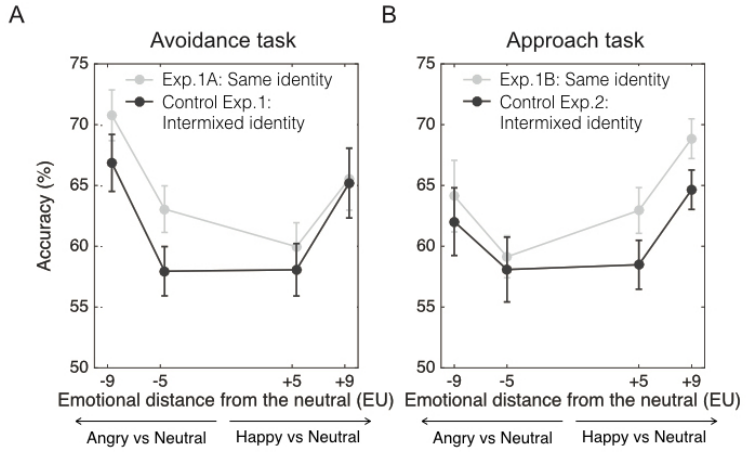
S4. Replication and Extension Experiments 1 and 2: The effect of the facial identity

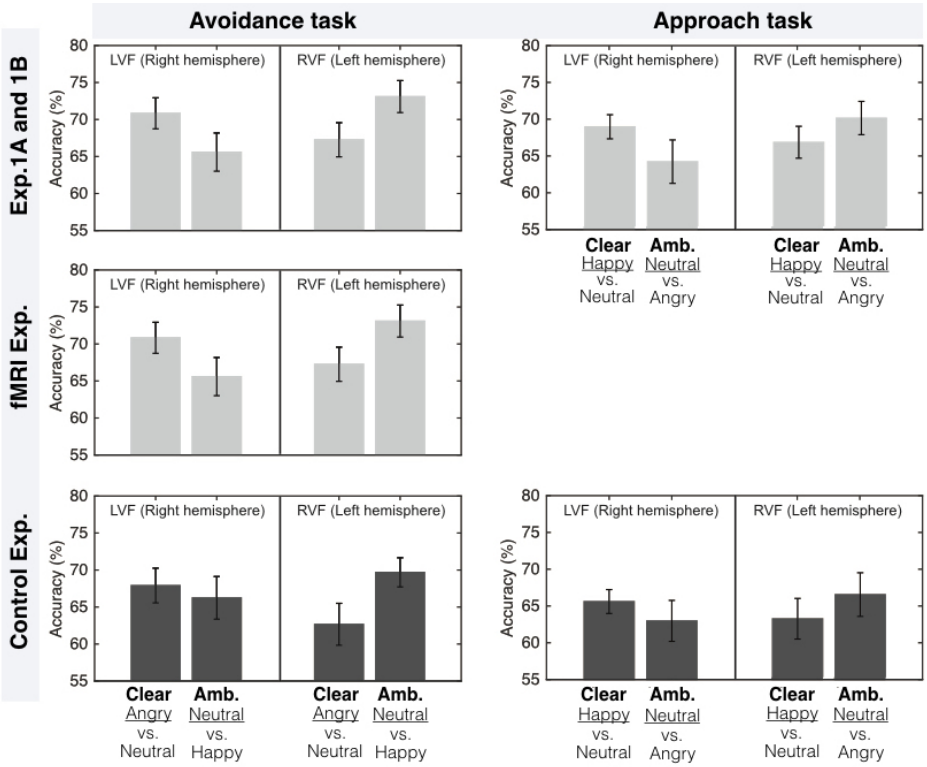
Although most of the previous studies of average crowd emotion constructed each picture using faces of the same person that varied in emotionality, we never view one person's various emotional expressions simultaneously, except under laboratory conditions. Rather, we encounter groups of individuals who differ not only in their emotional expression, but also identity, age, and gender cues. To test whether facial identity cues interfere with processing of crowd emotion, we ran two control experiments that employed visual stimuli containing intermixed, different identities and directly compared participants' accuracy to Experiments 1A and 1B where the emotional expressions of the same identity were contained. The procedures of the control experiments were identical to Experiments 1A and 1B, except for the presentation of visual stimuli containing intermixed identities. The top panel in Fig.S4 demonstrates the display that was used for the control experiments. A new group of 21 participants performed the avoidance task (Control Experiment 1) and another group of 19 participants performed the approach task (Control Experiment 2). The participants' accuracy for Experiments 1A, 1B and the two control experiments are plotted together in a bar graph (Fig.S4 bottom panel) for each comparison. Although we found that participants were slightly less accurate when the display contained intermixed facial identities than when the display contained the emotional expressions of the same identity, the difference was not significant both for the avoidance task ($t(40) = 1.230$, $p = 0.291$, Cohen's d = 0.310) and for the approach task ($t(38) = 1.396$, $p = 0.243$, Cohen's d = 0.385). These results suggest that facial identity cues do not interfere with processing crowd emotion and that participants were still able to extract accurately the average emotion from emotional faces of different identities.



S5. Results of Control Experiments: The effect of emotional distance between two facial crowds to be compared

From the two control experiments where participants were presented with visual stimuli containing intermixed identities, we could replicate the effects of emotional distance between two facial crowds to be compared with a different set of stimuli and a new participant cohort. We found that the accuracy was systematically affected by the emotional distance between two crowds being compared, with increased accuracy for greater emotional distance. Fig.S5 shows the results from Control Experiment 1 (A: avoidance task) and Control Experiment 2 (B: approach task). In Fig.S5, we also overlaid the results from our main Experiments 1A and 1B for easy comparison. Just as in Experiments 1A and 1B, we also observed facilitation of the task-goal congruent crowd emotion (e.g., -9 being higher than any others for the avoidance task; +9 being higher than any others for the approach task), suggesting that it is a robust, generalizable effect.



S6. Replication of the task-goal dependent hemispheric lateralization during crowd emotion processing

In all the behavioral tests we ran and reported here, we could replicate the pattern of the task-goal dependent hemispheric lateralization for crowd emotion processing. The ability to replicate the same pattern in different experiments with different settings (e.g., behavioral testing room or fMRI scanner), with different stimuli (e.g., containing same identity or intermixed identities), and with different cohorts of participants allows us to conclude that this effect is robust and reproducible. Fig.S6 shows the task-goal dependent lateralization replicated in five experimental sessions altogether, to highlight our main finding that that valence of emotional processing can be biased differently in RH and LH, depending on the task goal and viewers' intent.