Title

Holistic ensemble perception

Permalink

https://escholarship.org/uc/item/8wf397wz

Journal

Attention, Perception, & Psychophysics, 83(3)

ISSN

1943-3921

Authors

Han, Linfeng
Yamanashi Leib, Allison
Chen, Zhimin
et al.

Publication Date

2021-04-01

DOI

10.3758/s13414-020-02173-1

Peer reviewed

Published in final edited form as:
Atten Percept Psychophys. 2021 April ; 83(3): 998–1013. doi:10.3758/s13414-020-02173-1.

Holistic Ensemble Perception

Linfeng Han[1,*], **Allison Yamanashi Leib**[2,*], **Zhimin Chen**[2], **David Whitney**[2]

[1]Tsinghua University, China 100084

[2]University of California, Berkeley, USA 94720

Abstract

In a glance, observers can evaluate gist characteristics from crowds of faces, such as the average emotional tenor or the average family resemblance. Prior research suggests that high-level ensemble percepts rely on holistic and viewpoint-invariant information. However, it is also possible that feature-based analysis was sufficient to yield successful ensemble percepts in many situations. To confirm that ensemble percepts can be extracted holistically, we asked observers to report the average emotional valence in crowds of Mooney faces. Mooney faces are two-tone, shadow-defined images that cannot be recognized in a part-based manner. To recognize features in a Mooney face, one must first recognize the image as a face by processing it holistically. Across experiments, we demonstrated that observers successfully extracted the average emotional valence from crowds that were spatially distributed or viewed in a rapid temporal sequence. In a subsequent set of experiments, we maximized holistic processing by including only those Mooney faces that were difficult to recognize when inverted. Under these conditions, participants remained highly sensitive to the average emotional valence of Mooney face crowds. Taken together, these experiments provide evidence that ensemble perception can operate selectively on holistic representations of human faces, even when feature-based information is not readily available.

Introduction

Ensemble perception is a heuristic that provides critical information about the visual environment in a brief glance (Alvarez, 2011; Hochstein, Pavlovskaya, Bonneh, & Soroker, 2015; Whitney & Yamanashi Leib, 2018). For example, when someone is approaching a social gathering, ensemble perception makes it possible to rapidly perceive the average

Correspondence should be addressed to: Linfeng Han, hanlf@sas.upenn.edu, Department of Psychology, Tsinghua University, Beijing, China 100084.

[*]L. H. and A. Y. L. contributed equally to this work

Open Practices Statement
Stimuli, data, and codes are available at the Open Science Framework repository: https://osf.io/drmez/7view_only=87e9ad90c922410b9eaf2604762dfdf4.
None of the experiments was preregistered.

gregariousness of the crowd (Haberman & Whitney, 2007, 2009). In contrast, visually searching through the crowd to analyze the individual expressions of specific faces would be relatively slow. Ensemble perception therefore provides a rapid estimate of the summary facial expressions of groups of people. Visual ensembles also permit efficient summary statistical percepts of other socially-relevant crowd characteristics such as average family resemblance (de Fockert & Wolfenstein, 2009; Neumann, Schweinberger, & Burton, 2013; Yamanashi Leib et al., 2014) and the average gender-ratio of a crowd (Haberman & Whitney, 2007; Alt, Goodale, Lick, & Johnson, 2019; Phillips, Slepian, & Hughes, 2018). Despite the crucial influence of ensembles on conscious visual perception, the contribution of low- vs. high-level information to ensemble representations remains ambiguous.

Ensemble perception has been studied across numerous visual domains, starting with low-level stimuli including the direction and speed of moving dots (Dakin & Watt, 1997; Watamaniuk & Duchon, 1992), the average orientation of Gabor patches (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001) and the average size of briefly displayed circles (Ariely, 2001; Chong & Treisman, 2003, 2005). In addition to consolidating low-level cues, ensemble perception also incorporates high-level holistic cues. For example, ensemble perception of faces, biological motion, and animacy was negatively impacted by scrambling, a technique that impairs holistic processing (Haberman & Whitney, 2007; Sweeny, Haroz, & Whitney, 2013; Yamanashi Leib, Kosovicheva, & Whitney, 2016). Moreover, successful ensemble perception occurs for viewpoint-invariant representations of faces, necessarily indicating that some high-level cues are assimilated into gist percepts of crowds (Neumann et al., 2013; Yamanashi Leib et al., 2014). These studies are compelling, but they employ photographs of faces that include both low-level features and high-level attributes such as holistic cues, which introduces some limitations. For example, Sekuler and colleagues (2004) argued that the inversion of face photographs does not lead to qualitative change in processing style. It may only delay the retrieval of holistic representations (Richler, Mack, Palmeri, & Gauthier, 2011), as opposed to traditional views that inversion blocks access to holistic information in face photographs (e.g., Hole, 1994; Rossion & Boremanse, 2008). Also, ensemble coding of inverted crowds can still occur under several circumstances (e.g., Elias, Dyer, & Sweeny, 2017; Sweeny & Whitney, 2014). Therefore, it remains unclear based on prior literature using face photographs whether there is ensemble coding for holistic representations of faces in the absence of part-based cues (such as shading, identifiable features, and surface texture characteristics).

A holistic representation is a unitary coherent structure that cannot be parsed or segmented into components without first knowing what the image is. An ideal example of this is a Mooney face, a two-tone shadow defined image that appears to be a face (Fig. 1A). Mooney faces do not have identifiable parts or features until they are recognized as faces; for example, one cannot localize or identify the eye, nose, or emotion of a Mooney face without first recognizing the image as a face. To address whether summary statistical information can be extracted for holistic faces, we investigated whether observers could perceive the ensemble emotion of Mooney faces.

Mooney faces were initially created to investigate visual closure development in children (Mooney, 1957), and can be extremely challenging to recognize, relative to natural photographs of faces (Farzin, Rivera, & Whitney, 2009). Past literature suggests that the recognition of some Mooney faces can take up to several seconds (Ramachandran, Armel, Foster, & Stoddard, 1998). However, once the Mooney face is recognized as a face—it is difficult to be “unseen” as a face (Ludmer, Dudai, & Rubin, 2011). Not only are Mooney stimuli recognized as faces in behavioral experiments, but also Mooney stimuli activate selective face-processing regions, such as the FFA (Andrews & Schluppeck, 2004; George, Jemel, Fiori, Chaby, & Renault, 2005; Kanwisher, Tong, & Nakayama, 1998; Latinus & Taylor, 2005, 2006; McKeeff, Remus, & Tong, 2007; Moscovitch, Winocur, & Behrmann, 1997). Importantly, in order to successfully discriminate any particular feature within a Mooney face, it is necessary to first process the image as a Gestalt (Cavanagh, 1991; Kemelmacher-Shlizerman, Basri, & Nadler, 2008; Moore & Cavanagh, 1998). Indeed, seminal face-processing experiments rely on Mooney faces to fully isolate holistic processing (McKone, 2004). In the following set of experiments, we probe the ability of observers to assess the emotional tenor of Mooney face crowds across both spatial and sequential experimental displays.

Experiment 1

Standardization of the Mooney Face Stimulus Set

The general goal across all experiments was to measure sensitivity to ensemble emotion within Mooney face crowds. Experiment 1 was conducted to obtain a standardized emotional valence value for each individual Mooney face.

Methods

Observers.—Nineteen observers on Amazon Mechanical Turk rated the emotional valence of each Mooney face. We asked participants to proceed with the experiments only if they had no neurological history and normal or corrected-to-normal vision. Participants did not report their gender or age. All consent and experimental procedures were approved by the Institutional Review Board at the University of California, Berkeley.

Stimuli.—Ke and his colleagues created a large Mooney face stimulus set using a Generative Adversarial Network (GAN). Comprehensive details of the Mooney face stimulus set can be accessed in their publication (Ke, Yu, & Whitney, 2017). We pseudo-randomly chose 136 of these Mooney face images using the following constraints: 1) the stimulus set should not contain repeated or visually identical images; 2) the facial expressions in the selected images should be visually recognizable, and 3) the stimulus set should exhibit a broad range of emotional valence (i.e., from positive to negative). Three independent raters cross-validated the images based on these constraints. Some example stimuli are shown in Fig. 1A.

Procedure.—Amazon Mechanical Turk participants were asked to place their personal computer monitor in a central position in front of them and were asked to maintain a clear, unobstructed view of the screen. They were also asked to sit an arm’s length away from the computer screen. They were instructed that displays would be brief and were encouraged to maintain vigilant attention throughout the experiment. On each trial, one Mooney face was presented for 1 second on the center of the screen. After the face disappeared, participants were required to rate the emotional valence of the face using a Likert scale, ranging from 1-9, with 1 representing the most negative emotional valence and 9 representing the most positive emotional valence. Participants were not given a time limit to make their responses. After the participant entered their response, they were allowed to proceed to the next trial. The experimental procedure is shown in Fig. 1B.

Results

We averaged across the ratings of the 19 observers to obtain a single rating for each image in the 136-image set. This yielded a range of single Mooney emotional valence ratings from 3.84 to 7.74 ($M = 5.97$, $SD = 0.97$). To confirm that the ratings of emotional valence were reliable, we used a split-half measure of inter-rater reliability. The split-half measure yielded an average bootstrapped Fisher's $z = 1.39$, $SD = 0.11$, We then transformed the average Fisher's z score back to Pearson's r, using the method described in Corey, Dunlap, & Burke, 1998, r_z, = .88, $p < .001$, indicating a high level of agreement among participants. We used these average single Mooney face ratings as a baseline measure in Experiments 2–4.

Experiment 2

Ensemble Coding Mooney Faces: Spatially Distributed Display

The purpose of Experiment 2 was to examine whether participants were able to extract ensemble percepts from crowds of simultaneously presented Mooney faces.

Methods

Observers.—Ten observers participated in Experiment 2 (seven females, three males, $M_{\text{age}} = 22.70$, $SD = 4.95$). Our sample size was based on prior ensemble perception studies that reported robust findings using similar sample sizes (e.g., Haberman & Whitney, 2011; Sweeny, Wurnitsch, Gopnik, & Whitney, 2015). Observers were undergraduate students from the University of California, Berkeley, and were given course credit for their participation. The experiment was conducted in accordance with the guidelines of the Institutional Review Board at the University of California, Berkeley.

Stimuli.—We used the same 136 Mooney face stimuli that were previously tested in Experiment 1. All stimuli were viewed on an LCD monitor, with a 1920 × 1080 resolution and refresh rate of 60-Hz. On each trial, a maximum of 6 Mooney faces was displayed (each subtending 5.4° by 5.4° of viewing angle). The stimuli were displayed on a 2 × 3 grid centrally positioned on the screen, subtending 12.9° by 18.6° of viewing angle. Stimuli were presented using Matlab Version R2017a and Psychophysics Toolbox Version 3 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

Procedure.—On each trial, participants viewed a fixation cross for 600 ms. After the fixation cross disappeared, participants viewed a crowd of six Mooney faces (Fig. 2A). The six Mooney faces were pseudo-randomly drawn from the entire stimulus set of 136 faces. Across the experiment, the mean emotional valence of crowds was uniformly distributed, ranging from 4.39 to 7.35. The emotional variability of crowds (i.e., standard deviation of the emotional valence) on each trial ranged from 0.16 to 1.60, with an average of 0.75 (the distributions of mean and variability were similarly consistent in Experiments 3a, 3b, and 4 because our randomization approach was identical). The crowd of 6 Mooney faces was displayed for one second. This exposure time was similar or less than that used with naturalistic photographs of faces in prior ensemble perception paradigms (Haberman & Whitney, 2007; de Fockert & Wolfenstein, 2009; Neumann et al., 2013; Wolfe, Kosovicheva, Yamanashi Leib, Wood, & Whitney, 2015; Sweeny et al., 2013; Sweeny & Whitney, 2014). After the crowd of Mooney faces disappeared, a continuous scale appeared. Participants clicked on any point between 1 and 9, with 1 representing the most negative emotional valence and 9 representing the most positive emotional valence (Fig. 2A). No time limit was imposed during the rating phase.

Results

To measure the ensemble sensitivity to the crowd of Mooney faces, we correlated the mean of individual Mooney face ratings (generated in Experiment 1) with the ensemble ratings in Experiment 2 (Fig. 2B illustrates one representative observer's data). We analyzed the data in this manner for each individual subject. The average ensemble sensitivity across the group of participants was high (Fisher's $z = .96$; transformed r_z, = .75, $p < .001$), $SD = .19$; Fig. 3C, triangular data point). This suggested that participants were sensitive to the ensemble emotion of the Mooney face crowds, and they were able to extract ensemble percepts even when feature-based information was minimal.

Although we explicitly asked subjects to report the average of *all* displayed faces, it is possible that they did not actually integrate the emotional valence of multiple faces (a necessary requirement of ensemble perception), but rather randomly reported the emotional valence of one face from the crowd. To control for this possibility, we incorporated subset conditions into our experimental design. The subset conditions allow us to measure how many faces were integrated from the crowd; this measure is commonly used as a control in ensemble perception paradigms (Chong, Joo, Emmmanouil, & Treisman, 2008; Piazza, Sweeny, Wessel, Silver, & Whitney, 2013; Sweeny et al., 2013; Sweeny & Whitney, 2014; Sweeny et al., 2015; Wolfe et al., 2015; Yamanashi Leib et al., 2014; Yamanashi Leib et al., 2016). In 50 trials, we showed the whole crowd (six random faces) to subjects. In the remaining 250 trials, only subsets of the whole crowd (i.e., 1, 2, 3, 4, or 5 faces) were presented. There were 50 trials in each subset condition. In these "subset" conditions, we randomly drew 6 Mooney faces from the stimulus set but displayed only k faces (k = 1~5), with (6-k) faces being visibly hidden. However, we recorded the mean emotional valence of the 6-face crowd, which we referred to as the whole set mean. We measured integration by correlating subjects' ratings of the subsets with the *whole set mean.* In this way, we empirically simulated observers' performance in the whole set condition as if they were using only k faces to make their judgement. By comparing the subjects' performance in these simulated conditions in which (6-k) faces were obscured with their actual performance in the whole set condition, we could learn whether subjects were visually integrating all six faces. For example, if subjects based their report on one randomly sampled Mooney face from the crowd, their Fisher's z score would be consistently low across the range of subset sizes, even when more faces were revealed in the display (Fig. 3A). In contrast, if all six faces were integrated (an ideal version of ensemble coding), their Fisher's *z* score would systematically increase as the number of displayed faces increased (Fig. 3B).

We analyzed ensemble sensitivity (measured as Fisher's *z* in each subset size (Fig. 3C) and observed a clear monotonic trend; Fisher's *z* systematically increased as the number of faces displayed increased. An one-way repeated measures ANOVA (rmANOVA) confirmed a significant main effect across subset sizes, $F(5,45) = 22.169$, $p < .001$, $\eta^2 = .711$. Although the data in Fig. 3C exhibit some degree of compressive nonlinearity, we applied a linear regression model to quantitatively measure this monotonic trend. The linear slope of this trend was significantly different from the slope of a permuted null distribution (based on shuffling subset size labels, $p < 0.001$). These results confirm that participants integrated multiple faces into their ensemble percepts instead of subsampling a single or a small number of faces. Taken together, the results of Experiment 2 clearly show that ensemble perception effectively operates across crowds of Mooney faces.

Experiment 3a

Ensemble Coding Mooney Faces: Sequential Display

Ensemble perception functions in both spatial and temporal domains (see Fig. 4 in Whitney & Yamanashi Leib, 2018, for a summary). Besides ensemble coding of simultaneously presented information, robust ensemble sensitivity during statistical averaging over time has been well documented (e.g., Florey et al., 2017; Haberman, Harp, & Whitney, 2009; Hubert-Wallander & Boynton, 2015; Oriet & Corbett, 2008; Piazza et al., 2013). Such rapid sequential averaging is regarded as an essential form of ensemble perception. Therefore, we designed Experiment 3a to examine whether ensemble perception is sensitive to the emotional valence of Mooney faces displayed in a sequence. This type of display mimics crowds walking past on the street or eye scanning patterns across a visual scene.

Methods

Observers.—Ten observers participated in Experiment 3a (six females, three males, one did not report gender, $M_{\text{age}} = 24.40$, $SD = 9.79$). Observers were undergraduate students from the University of California, Berkeley, and were given course credit for their participation. The experiment was conducted in accordance with the guidelines of the Institutional Review Board at the University of California, Berkeley.

Stimuli and Procedure.—The stimulus set and response phase were identical to Experiment 2. However, in the display phase, faces were shown sequentially rather than simultaneously. Each individual face subtended 10.2° by 10.2° of viewing angle and was displayed with random spatial jitter, up to 2.7° in both horizontal and vertical directions. In the whole set, six faces were shown at the center of the screen (with spatial jitter) for 50 ms each, with a 50 ms inter-stimulus-interval (ISI) (Fig. 4A). In the subset conditions, the face exposure time varied depending on the size of the subset – 1 face: 300 ms, 2 faces: 150 ms, 3 faces: 100 ms, 4 faces: 67 ms, 5 faces: 50 ms (all above refer to display durations per each face). Across the subset sizes, the total face exposure duration and ISI were equivalent or nearly equivalent (due to a 60-Hz refresh rate constraint). Note that lengthening the exposure time for each face in the smaller subsets allows individuals to more carefully scrutinize/view faces. However, if multiple faces were truly integrated into the ensemble percept, we would still observe the predicted subset effect (increasing error with smaller subsets), despite longer viewing durations in the smaller subsets. After the sequential display of faces disappeared, participants reported the average emotional valence of the crowd using a continuous scale (Fig. 4A).

Results

We analyzed ensemble sensitivity by correlating ratings of Mooney face crowds in Experiment 3a to the mean of the individual Mooney faces, rated by independent observers in Experiment 1. We observed robust ensemble sensitivity to the whole set of Mooney faces, Average Fisher's $z = 1.04$ (transformed r_z, = .78, $p < .001$), $SD = .22$ (Fig. 4B, triangular data-point). This is strong evidence that participants were able to extract an ensemble percept from temporally displayed faces when part-based information was difficult to access. A one-way rmANOVA also confirmed that ensemble sensitivity (Fisher's *z* scores) systematically increased as the number of faces revealed increased. $F(5,45) = 28.684$, $p < .001$, $\eta^2 = .761$ (Fig. 4B). Taken together, this is convincing evidence for an ensemble representation of Mooney face emotion.

Experiment 3b

Ensemble Coding Mooney Faces: Sequential Display with Shorter Durations

In Experiment 3b, we repeated Experiment 3a, except that we utilized a briefer duration to display the individual faces. We reasoned that rapid displays would make it harder to process individual faces and local features within them (Ramachandran et al., 1998), but holistic ensemble representations should persist given that summary statistics remain intact even when attentional resources are insufficient to process local information (Corbett & Oriet, 2011). If ensemble perception operated on holistic representations of Mooney faces, observers would exhibit strong ensemble sensitivity in the current experiment.

Methods

Observers.—Ten observers participated in Experiment 3b (eight females, two males, $M_{age} = 19.40$, $SD = .97$). Observers were undergraduate students from the University of California, Berkeley, and were given course credit for their participation. The experiment was conducted in accordance with the guidelines of the Institutional Review Board at the University of California, Berkeley.

Stimuli and Procedure.—The stimuli and procedure were identical to Experiment 3a except that we employed briefer display durations. In the whole set condition, we displayed six faces for 33 ms with a 33 ms 1ST In the subset conditions, we displayed 1-5 faces for 200, 100, 67, 50, 33 ms (per each face) respectively. The total exposure duration remained nearly constant across the set sizes; the ISI also remained constant at 33 ms. Each set size was tested for 50 trials, tested for 50 trials.

Results

Using analyses identical to the prior experiments, we found robust evidence of ensemble sensitivity when Mooney faces were shown for briefer exposure durations. The average Fisher's z in the whole set condition across observers = .84 (transformed r_z, = .68, $p <$ .001), SD = .30, (Fig. 4C, triangular data point). We found evidence that multiple faces were integrated into the ensemble percept during this rapid stream of Mooney faces (Fig. 4C, triangular data-point). An one-way rmANOVA confirmed that ensemble sensitivity increased as the set size increased, $F(5,45) = 17.242$, $p < .001$, $\eta^2 = .657$. Taking these analyses together, we demonstrated that participants incorporated holistic information into their ensemble percept when faces were presented as briefly as 33 ms each.

Experiment 4

Ensemble Perception of Mooney Faces During Upright vs. Inverted Viewing Conditions

Mooney faces are generally agreed to require holistic processing (Basri, & Nadler, 2008; Canas-Bajo & Whitney, 2020; Cavanagh, 1991; Kemelmacher-Shlizerman, McKone, 2004; Moore & Cavanagh, 1998). Despite this general agreement, specific Mooney faces require holistic processing more than others (Canas-Bajo & Whitney, 2020). The purpose of Experiment 4 was to confirm that the Mooney faces utilized in our experiments specifically induced holistic processing rather than feature-based analysis. We used a standard inversion paradigm (Maurer, Le Grand, & Mondloch, 2002; McKone, 2004; Valentine, 1988) to determine whether this specific Mooney face stimuli set required holistic processing. At the very least, face inversion should delay access to holistic representations (Richler et al., 2011; Sekuler et al., 2004). Moreover, unlike gray-scale faces, inverting Mooney faces often prevents their recognition, and the inversion effect in Mooney faces has therefore been thought of as a gold-standard for isolating holistic face processing (Canas-Bajo & Whitney, 2020). If the stimulus set requires holistic ensemble perception, we expected to observe substantially reduced ensemble sensitivity caused by face inversion.

Methods

Observers.—Seventeen observers participated in Experiment 4 (thirteen females, four males, $M_{\text{age}} = 19.70$, $SD = 1.16$). Observers were undergraduate students from the University of California, Berkeley, and were given course credit for their participation. The experiment was conducted in accordance with the guidelines of the Institutional Review Board at the University of California, Berkeley.

Stimuli and Procedure.—We repeated Experiment 2 using both upright and inverted blocked conditions. The order of these blocks was counterbalanced across subjects. In Experiment 4, each Mooney face subtended 10.2° by 10.2° of viewing angle. The stimuli were displayed on a 2 × 3 grid centrally positioned on the screen, subtending 21.4° by 34.7° of viewing angle. All other aspects of the experimental paradigm were identical to Experiment 2 (Fig. 5A).

Results

We analyzed performance in the upright and inverted conditions separately, using the same analysis described in prior experiments. We measured ensemble sensitivity by correlating the ratings of upright Mooney face crowds in Experiment 4 with the mean of individual Mooney faces, rated by independent observers from Experiment 1. Ensemble sensitivity was high in the upright ensemble viewing condition, average Fisher's $z = .97$ (transformed r_Z, = .75, $p < .001$), $SD = .22$ and lower in the inverted viewing condition, Average Fisher's $z = .62$ (transformed r_Z, = .55, $p < .001$), $SD = 0.20$ (Fig. 5B, black and gray triangular data points, respectively). This difference in sensitivity was confirmed with a paired-sample t-test, $t(16) = 7.198$, $p < .001$. We also confirmed that ensemble sensitivity increased as the set size increased for both upright and inverted displays using rmANOVA (upright: $F(5,80) = 59.730$, $p < .001$, $\eta^2 = .789$; inverted: $F(5,80) = 18.477$, $p < .001$, $\eta^2 = .536$). Importantly, the slope of the linear integration pattern during the inverted viewing condition was significantly shallower compared to the upright viewing condition ($t(16) = 4.417$, $p < 0.001$), which indicates that integration was less efficient for inverted faces.

Despite the difference in ensemble sensitivity between the two viewing conditions, there are two caveats. First, observers did (albeit to a significantly lesser degree) integrate inverted face information into an ensemble percept. This may have occurred because the current stimulus set retained some non-holistic cues or low-level features that were available even when inverted (Canas-Bajo & Whitney, 2020). Second, we incorporated only 136 Mooney face images in our stimulus set. It was thus likely that the observers gained familiarity with these stimuli over the course of the experiment. For example, exposure to upright faces in the first block may introduce some information that impacts subsequent viewing of inverted faces. To address this issue, we tested whether observers' performance varied as a function of the counterbalancing order of upright and inverted blocks. It turned out that the counterbalancing order did *not* affect ensemble performance in both viewing orientations (ensemble sensitivity difference for upright 6-face crowds: $t(15) = .612$, $p = .550$; inverted: $t(15) = .06$ $p = .949$). Nonetheless, as a precaution against future familiarity confounds, we used a larger stimulus set and further minimized feature-based information in subsequent experiments.

Experiment 5

Establishing a Maximally Holistic Set of Mooney Faces

Although Mooney face stimuli are thought to require holistic processing (Latinus and Taylor, 2005; Moscovitch, Winocur, & Behrmann, 1997), some variations have been documented (Canas-Bajo & Whitney, 2020; McKone, Brewer, MacPherson, Rhodes, & Hayward, 2007). Our particular Mooney face stimulus set, generated by an adversarial network (Ke, Yu, & Whitney, 2017), has not been widely tested within the face literature and may retain some feature-based cues (Canas-Bajo & Whitney, 2020). To ensure that our results were not artifacts of the particular Mooney face set used, we employed another set of Mooney faces to confirm generalizability and designed Experiment 5 to ensure that holistic processing was optimally engaged.

Methods

Observers.—Eleven observers participated in Experiment 5 (ten females, one male, $M_{\text{age}} = 20.64$, $SD = 1.12$). Observers were undergraduate students from the University of California, Berkeley, and were given course credit for their participation. The experiment was conducted in accordance with the guidelines of the Institutional Review Board at the University of California, Berkeley.

Stimuli and Procedure.—The primary stimulus set that we selected was created by Schwiedrzik, Melloni, and Schurger (2018), and was generated by taking profile or off-set views of faces under extreme lighting conditions, and further manipulating the faces in photoshop (see Schwiedrzik, Melloni, and Schurger (2018) for details). The stimulus set contained 500 different facial identities, with each identity having both upright and inverted Mooney face images (see Fig. 6A for example stimuli). Our goal was to identify the Mooney faces that strongly rely on holistic processing, and we employed a technique similar to that used by McKone (2004) to do so. We displayed individual faces from the 500-face set centrally on the computer screen for 1 second each, in either upright or inverted orientation. Upright and inverted faces were randomly interleaved. Each stimulus subtended 9.2° by 6.4° of viewing angle. After the individual Mooney face disappeared, subjects were asked to rate whether the image looked “face-like” using a continuous scale of 1-9, with 1 representing “definitely not a face” and 9 representing “definitely a face” (Fig. 6B).

Results

We averaged the ratings of the eleven subjects to obtain a group-level rating for each image in the 1000-image set (500 upright and 500 inverted images). Because subjects viewed each image in both orientations, all images received two ratings. To confirm that the ratings of face-likeness were reliable, we used a split-half measure of inter-rater reliability. The split-half measure of all faces (including both upright and inverted orientations) across subjects yielded an average bootstrapped Fisher’s $z = 1.06$ (transformed r_{z}, = .79, $p < .001$, $SD = 0.03$), indicating a high-level agreement among subjects. For each Mooney face image, we subtracted the rating for the inverted face from the rating of the upright face to calculate the inversion effect score (indicated as blue in Fig. 7). This inversion effect score reveals inter-stimulus differences in how recognizably face-like different Mooney faces are, and confirms previous work (Canas-Bajo & Whitney, 2020). Those that have strong inversion effects are operationally defined as being more holistic, following the logic of prior work (McKone, 2004). Half of the images that had the largest inversion effect were used as stimuli in the subsequent ensemble perception experiments.

Experiment 6

Emotional Valance Ratings of Single Mooney Faces from the Maximally Holistic Set

The purpose of Experiment 6 was to establish a baseline emotional valence rating for each Mooney face that yielded a strong inversion effect in Experiment 5.

Methods

Observers.—Thirteen observers participated in Experiment 6. Out of these 13 observers, we excluded one person because this participant failed to use the whole rating scale (primarily reporting only the extreme ends of the scale (1, 9). The remaining twelve observers were: seven females, five males, $M_{\text{age}} = 21.58$, $SD = 6.10$. Observers were undergraduate students from the University of California, Berkeley, and were given course credit or $15 remuneration for their participation. The experiment was conducted in accordance with the guidelines of the Institutional Review Board at the University of California, Berkeley.

Stimuli and Procedure.—Observers viewed the 250 Mooney faces selected from Experiment 5 in the upright orientation for one second each. After the face disappeared, they rated the emotional valence of the single Mooney face using a continuous scale. The procedure was identical to Experiment 1.

Results

We averaged across the ratings of the twelve observers to obtain a single rating for each of the images in the upright and inverted conditions. The emotional valence ratings ranged from 3.10 to 7.89 ($M = 5.24$, $SD = 1.03$). To confirm that the ratings of emotional valence were reliable, we used a split-half measure of inter-rater reliability. The split-half measure yielded an average bootstrapped Fisher's $z = 1.47$ (transformed r_Z, = .90, $p < .001$), $SD =$ 0.28, indicating a high level of agreement among observers. We used the emotional valence ratings for the new stimulus set as a baseline in subsequent ensemble perception experiments (Experiments 7a and 7b).

Experiment 7a

Ensemble Coding Maximally Holistic Mooney Faces: Spatially Distributed Display

The goal of Experiment 7a was to ensure that the ensemble perception of Mooney faces found in the previous experiments was not an artifact of the particular Mooney faces used, and to investigate if ensemble perception successfully functions on faces with stronger holistic information. In Experiment 7a, we presented crowds of maximally holistic Mooney faces in upright and inverted orientations to test ensemble sensitivity. The Mooney face crowds were presented in spatial arrays, which was identical to Experiment 4 but with the new set of Mooney faces identified in Experiments 5 and 6.

Methods

Observers.—Eleven observers participated in Experiment 7a, (six females, five males, $M_{\text{age}} = 21.18$, $SD = 3.31$). Observers were undergraduate students from the University of California, Berkeley, and were given course credit or $15 remuneration for their participation. The experiment was conducted in accordance with the guidelines of the Institutional Review Board at the University of California, Berkeley.

Stimuli and Procedure.—The Mooney face stimuli used in Experiment 7a were identical to Experiment 6. The experimental procedure was identical to Experiment 4, except that upright and inverted viewing trials were interleaved within the whole experiment (while we adopted a block design in Experiment 4). In the whole set condition, six Mooney faces were simultaneously displayed for 1 second. Each Mooney face subtended 9.2° by 6.4° of viewing angle. The stimuli were displayed on a 2 × 3 grid centrally positioned on the screen, subtending 19.3° by 21.4° of viewing angle. In the subset conditions, 1~5 Mooney faces were presented on random locations of the grid. For each viewing orientation, there were 50 trials in the whole set condition and each subset condition. Across the whole set trials, the mean emotional valence of crowds was uniformly distributed in the range of 3.75 to 7.34. The emotional variability of crowds (i.e., standard deviation of the emotional valence) ranged from 0.15 to 1.79, with an average of 0.90 (these distributions remained similarly consistent in Experiment 7b).

Results

Upright and inverted conditions were separately analyzed using the same procedures described in prior experiments. In the upright condition, we correlated ratings of ensemble Mooney crowds with the mean of individual Mooney faces, rated in from Experiment 6, Average Fisher's $z = 0.83$ (Fig. 8A, black triangular data-point, transformed r_z, = .68, $p < .001$), $SD = 0.29$. The robust Fisher's z score indicates high ensemble sensitivity. Ensemble sensitivity was lower in the inverted condition, Average Fisher's $z = 0.29$ (Fig. 8A, gray triangular data-point, transformed r_z, = .28, $p < .001$), $SD = 0.21$. compared to the upright condition; this difference was confirmed by a paired-sample t-test, $t(10) = 5.35$, $p < .001$. Ensemble sensitivity increased as the set size increased, $F(5,50) = 17.300$, $p < .001$, $\eta^2 = .634$; this indicates that participants integrated multiple upright faces into their ensemble representation. However, the main effect of set size was *not* statistically significant for inverted faces, $F(5,50) = 2.262$, $p = .062$, $\eta^2 = .184$. We observed robust ensemble perception during the upright viewing condition, but ensemble integration was severely impaired during the inverted viewing condition when holistic information was reduced.

Experiment 7b

Ensemble Coding Maximally Holistic Mooney Faces: Sequential Display

The purpose of Experiment 7b was to test observers' ensemble sensitivity with sequentially presented maximally holistic Mooney faces identified in Experiments 5 and 6. The display fashion of the Mooney face crowds were similar to Experiments 3a and 3b, but we also included the inverted viewing orientation. If ensemble perception can rely on holistic representations of faces, observers' ensemble performance should be robust when viewing upright Mooney face crowds but not inverted ones.

Methods

Observers.—Thirteen observers participated in Experiment 7a, (nine females, four males, M_{age}= 21.92, $SD = 3.70$). Observers were undergraduate students from the University of California, Berkeley, and were given course credit or $15 remuneration for their participation. The experiment was conducted in accordance with the guidelines of the Institutional Review Board at the University of California, Berkeley.

Stimuli and Procedure.—The Mooney face stimuli used in Experiment 7b were identical to Experiment 6. In the whole set condition, six Mooney faces were sequentially displayed for 100 ms each, with a 100 ms ISI. In the subset conditions, 1-5 faces were sequentially displayed for 600, 300, 200, 150, 117 ms, respectively. Each face was also displayed with random spatial jitter, up to 2.7° horizontally and vertically. The ISI remained constant at 100 ms across all subset conditions. After the temporal crowd disappeared, participants used a continuous scale (1-9) to rate the average emotional valence of the crowd.

Results

We analyzed the upright and inverted conditions separately, using the same analysis described in prior experiments. In the upright condition, where holistic information was mostly isolated, we observed strong evidence for ensemble sensitivity, average Fisher's z = 1.05 (Fig. 8B, black triangular data-point, transformed r_z, = .78, $p < .001$), $SD = 0.17$. Ensemble sensitivity was significantly lower in the inverted condition, average Fisher's z = 0.44 (Fig. 8B, gray triangular data-point, transformed r_z, = .43, $p < .001$), $SD = 0.30$. The decrement in ensemble performance for inverted compared to upright faces was confirmed by a paired-sample t-test, $t(12) = 6.599$, $p < .001$. We also found strong evidence that participants integrated multiple upright faces into their ensemble report, with ensemble sensitivity increasing as the set size increased using rmANOVA: $F(5,60) = 41.425$, $p < .001$, $\eta^2 = .775$. A statistically significant but much shallower integration effect was observed for inverted face crowds, $F(5,60) = 4.671$, $p = .001$, $\eta^2 = .280$. We measured the slope of the subset effect in both upright and inverted conditions. The slope was significantly steeper in the upright condition, $t(12) = 4.836$, $p < .001$. This difference suggests that integration efficiency was reduced for inverted faces.

General Discussion

Across multiple experiments, we documented robust ensemble sensitivity to Mooney face crowds displayed in multiple spatial and sequential configurations. We also included additional control measures to ensure that piece-meal based processing strategies were minimized across the Mooney face experiments. Importantly, our analyses confirmed the integration of multiple Mooney faces into the ensemble representations, a necessary criterion for ensemble perception (Whitney & Yamanashi Leib, 2018). Taken together, these experiments provide strong evidence that ensemble perception functions when holistic information is maximized, and feature-based information is minimized.

Mooney faces require extensive holistic processing and contain minimal part-based cues (Cavanagh, 1991; Kemelmacher-Shlizerman, Basri, & Nadler, 2008; Moore & Cavanagh, 1998). The two-toned faces cannot be interpreted when utilizing bottom-up processes alone (Cavanagh, 1991). The stimuli are sufficiently sparse that computer vision, despite substantial recent advances, still cannot identify Mooney faces (Ke, Yu, & Whitney, 2017). Only *after* human observers process the image as a whole, can they discriminate specific facial features (Cavanagh, 1991; Kemelmacher-Shlizerman, Basri, & Nadler, 2008; Moore & Cavanagh, 1998), and recognize the identity, gender, and emotional expression of Mooney faces (Maurer, Le Grand, & Mondloch, 2002; McKone, 2004; Mooney, 1957). Because of their unique and evocative properties, previous researchers employed Mooney faces to isolate holistic processing in face recognition tasks (McKone, 2004; McKone, Martini, & Nakayama, 2003), probe holistic processing in visual crowding tasks (Farzin, Rivera, & Whitney, 2009), and investigate holistic processing deficits in neurological patients (Busigny, Joubert, Felician, Ceccaldi, & Rossion, 2010; Wasserstein, Barr, Zappulla, & Rock, 2004). Here, we used Mooney faces to demonstrate that ensemble perception continues to operate when part-based cues, such as segmented features and surface texture, are strikingly difficult to access.

In the current research, we included several controls to ensure that participants engaged ensemble coding, a visual mechanism marked by relatively rapid perception (Haberman & Whitney, 2009; Li et al., 2016; Yamanashi Leib et al., 2016) and integration of multiple items (see Whitney & Yamanashi Leib, 2018, for a review). Note that our goal was not to demonstrate *parallel* processing of multiple Mooney faces in either simultaneous- or sequential-display experiments, as that is not a prerequisite for ensemble perception. With regards to rapid perception, although some experiments here included relatively long exposure durations (one second), we also included briefer exposure durations that minimized the possibility of slower processes such as cognitive deliberation or visual search. For example, in Experiments 3a, 3b, and 7a, we confirmed that holistic ensemble statistics are efficiently extracted during rapid serial visual presentation (RSVP) streams, even with stimuli as brief as 33 ms each. This extends previous research on ensemble perception in RSVP streams (Gorea, Belkoura, & Solomon, 2014; Haberman & Whitney, 2009; Hubert-Wallander & Boynton, 2015; Khayat & Hochstein, 2018, Oriet & Corbett, 2008). Participants successfully integrated multiple Mooney faces from the sequentially presented crowd, including temporally-adjacent faces, whose individual representations may even be masked by the attentional blink (Raymond, Shapiro, & Arnell, 1992; Shapiro, Raymond, & Arnell, 1994). As for integration of multiple Mooney faces, we confirmed it by including subset measures of integration commonly used in ensemble perception paradigms (Chong et al., 2008; Piazza et al., 2013; Sweeny et al., 2013; Sweeny & Whitney, 2014; Sweeny et al., 2015; Wolfe et al., 2015; Yamanashi Leib et al., 2014; Yamanashi Leib et al., 2016). Finally, in Experiments 7a and 7b, we confirmed our original findings with maximally holistic Mooney faces. Taken together, these controls establish that participants' ensemble sensitivity was not based on subsampling a single face or on feature-based processing. Rather, the observed performance in our studies reflects holistic ensemble perception.

Our findings contribute to the field with a systematic investigation of holistic ensemble perception using Mooney faces. In the domain of ensemble face perception, many prior studies demonstrate that the visual system rapidly aggregates characteristics from crowds of faces including: crowd threat (Alt et al., 2017), crowd emotion (Haberman & Whitney, 2007, 2009), gaze direction (Mareschal, Otsuka, Clifford, & Mareschal, 2016; Sweeny & Whitney, 2014), and head rotation direction (Florey, Clifford, Dakin, & Mareschal, 2016; Florey, Dakin, & Mareschal, 2017). Yet, observers may rely on feature-based visual analysis to extract summary statistics of multiple faces. Our findings extend past research by revealing that participants are strikingly sensitive to ensembles under conditions when part-based analysis becomes particularly difficult. Furthermore, in addition to prior studies that tap into high-level ensemble perception using scrambled images (e.g., Sweeny, Haroz, & Whitney, 2012, 2013), our study aims to systematically demonstrate holistic ensemble perception and accomplishes this goal *uniquely* with Mooney faces.

Ensemble perception is a beneficial visual mechanism that rapidly provides information about the visual environment, even when other visual processes are bottlenecked (Alvarez, 2011; Haberman & Whitney, 2012; Hochstein et al., 2015; Whitney, Haberman, & Sweeny, 2014; Whitney & Yamanashi Leib., 2018). It flexibly operates across a diversity of low-level (Dakin & Watt, 1997; Miller & Sheldon, 1969; Parkes et al., 2001; Watamaniuk & McKee, 1998; Watamaniuk, Sekuler, & Williams, 1989) and high-level stimuli (de Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007, 2009; Neumann et al., 2013; Yamanashi Leib et al., 2012), and our current findings extend ensemble perception to high-level holistic information. Despite the complexity of Mooney face stimuli, and the challenges that computer vision faces with identifying them, we show that humans can perceive ensemble characteristics of Mooney faces efficiently when viewing multiple Mooney faces simultaneously, or at display speeds of at least 10 and 15 Hz. The fact that ensemble perception successfully analyzes these impoverished stimuli, which lack shading, texture, and identifiable features, may indicate that ensemble perception is a broadly useful tactic to quickly summarize virtually any level of visual representation.

Acknowledgements

We thank Shuchen Liu, Danielle Budish, Azin Mirzaagha, and Julie Liu for assistance with data collection and analysis. This work is supported by Tsinghua University Initiative Scientific Research Program.

References

Alt NP, Goodale B, Lick DJ, & Johnson KL (2019). Threat in the company of men: ensemble perception and threat evaluations of groups varying in sex ratio. Social Psychological and Personality Science, 10(2), 152–159.

Alvarez GA (2011). Representing multiple objects as an ensemble enhances visual cognition. Trends in Cognitive Sciences, 15(3), 122–131. [PubMed: 21292539]

Andrews TJ, & Schluppeck D (2004). Neural responses to Mooney images reveal a modular representation of faces in human visual cortex. Neuroimage, 21(1), 91–98. [PubMed: 14741646]

Ariely D (2001). Seeing sets: Representation by statistical properties. Psychological Science, 12(2), 157–162. [PubMed: 11340926]

Brainard DH (1997). The psychophysics toolbox. Spatial Vision, 10, 433–436. [PubMed: 9176952]

Busigny T, Joubert S, Felician O, Ceccaldi M, & Rossion B (2010). Holistic perception of the individual face is specific and necessary: evidence from an extensive case study of acquired prosopagnosia. Neuropsychologia, 48(14), 4057–4092. [PubMed: 20875437]

Canas-Bajo T, & Whitney D (in press). Stimulus-specific individual differences in holistic perception of Mooney faces.

Cavanagh P (1991). What's up in top-down processing. Representations of vision: Trends and tacit assumptions in vision research, 295–304.

Chong SC, Joo SJ, Emmmanouil TA, & Treisman A (2008). Statistical processing: Not so implausible after all. Perception & Psychophysics, 70(7), 1327–1334. [PubMed: 18927015]

Chong SC, & Treisman A (2003). Representation of statistical properties. Vision Research, 43(4), 393–404. [PubMed: 12535996]

Chong SC, & Treisman A (2005). Attentional spread in the statistical processing of visual displays. Perception & Psychophysics, 67(1), 1–13. [PubMed: 15912869]

Corbett JE, & Oriet C (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. Acta Psychologica, 138(2), 289–301. [PubMed: 21903186]

Corey DM, Dunlap WP, & Burke MJ (1998). Averaging correlations: Expected values and bias in combined Pearson’s r and Fisher’s z transformations. The Journal of General Psychology, 125(3), 245–261.

Dakin SC, & Watt RJ (1997). The computation of orientation statistics from visual texture. Vision Research, 37(22), 3181–3192. [PubMed: 9463699]

de Fockert J, & Wolfenstein C (2009). Rapid extraction of mean identity from sets of faces. The Quarterly Journal of Experimental Psychology, 62(9), 1716–1722. [PubMed: 19382009]

Elias E, Dyer M, & Sweeny TD (2017). Ensemble perception of dynamic emotional groups. Psychological Science, 28(2), 193–203. [PubMed: 28036236]

Farzin F, Rivera SM, & Whitney D (2009). Holistic crowding of Mooney faces. Journal of Vision, 9(6), 18–18.

Florey J, Clifford CW, Dakin S, & Mareschal I (2016). Spatial limitations in averaging social cues. Scientific Reports, 6, 32210. [PubMed: 27573589]

Florey J, Dakin SC, & Mareschal I (2017). Comparing averaging limits for social cues over space and time. Journal of Vision, 17(9), 17.

George N, Jemel B, Fiori N, Chaby L, & Renault B (2005). Electrophysiological correlates of facial decision: insights from upright and upside-down Mooney-face perception. Cognitive Brain Research, 24(3), 663–673. [PubMed: 15890502]

Gorea A, Belkoura S, & Solomon JA (2014). Summary statistics for size over space and time. Journal of Vision, 14(9), 22.

Haberman J, Harp T, & Whitney D (2009). Averaging facial expression over time. Journal of Vision, 9(11), 1–1.

Haberman J, & Whitney D (2007). Rapid extraction of mean emotion and gender from sets of faces. Current Biology, 17(17), R751–R753. [PubMed: 17803921]

Haberman J, & Whitney D (2009). Seeing the mean: ensemble coding for sets of faces. Journal of Experimental Psychology: Human Perception and Performance, 35(3), 718. [PubMed: 19485687]

Haberman J, & Whitney D (2011). Efficient summary statistical representation when change localization fails. Psychonomic Bulletin & Review, 18(5), 855. [PubMed: 21748419]

Haberman J, & Whitney D (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. From perception to consciousness: Searching with Anne Treisman, 339–349.

Hochstein S, Pavlovskaya M, Bonneh YS, & Soroker N (2015). Global statistics are not neglected. Journal of Vision, 75(4), 7.

Hole GJ (1994). Configurational factors in the perception of unfamiliar faces. Perception, 23(1), 65–74. [PubMed: 7936977]

Hubert-Wallander B, & Boynton GM (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. Journal of Vision, 75(4), 5.

Kanwisher N, Tong F, & Nakayama K (1998). The effect of face inversion on the human fusiform face area. Cognition, 68(1), B1–B11. [PubMed: 9775518]

Ke TW, Stella XY, & Whitney D (2017, September). Mooney face classification and prediction by learning across tone. In Image Processing (ICIP), 2017 IEEE International Conference on (pp. 2025–2029). IEEE.

Kemelmacher-Shlizerman I, Basri R, & Nadler B (2008, June). 3D shape reconstruction of Mooney faces. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on (pp. 1–8). IEEE.

Khayat N, & Hochstein S (2018). Perceiving set mean and range: Automaticity and precision. Journal of Vision, 18(9), 23.

Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, & Broussard C (2007). What’s new in Psychtoolbox-3. Perception, 36(14), 1.

Latinus M, & Taylor MJ (2005). Holistic processing of faces: learning effects with Mooney faces. Journal of Cognitive Neuroscience, 77(8), 1316–1327.

Latinus M, & Taylor MJ (2006). Face processing stages: impact of difficulty and the separation of effects. Brain Research, 1123(1), 179–187. [PubMed: 17054923]

Li H, Ji L, Tong K, Ren N, Chen W, Liu CH, & Fu X (2016). Processing of individual items during ensemble coding of facial expressions. Frontiers in Psychology, 7, 1332. [PubMed: 27656154]

Ludmer R, Dudai Y, & Rubin N (2011). Uncovering camouflage: amygdala activation predicts long-term memory of induced perceptual insight. Neuron, 69(5), 1002–1014. [PubMed: 21382558]

Mareschal I, Otsuka Y, Clifford CW, & Mareschal D (2016). “Are you looking at me?” How children’s gaze judgments improve with age. Developmental Psychology, 52(5), 695. [PubMed: 26866727]

Maurer D, Le Grand R, & Mondloch CJ (2002). The many faces of configural processing. Trends in Cognitive Sciences, 6(6), 255–260. [PubMed: 12039607]

McKeeff TJ, Remus DA, & Tong F (2007). Temporal limitations in object processing across the human ventral visual pathway. Journal of Neurophysiology, 98(1), 382–393. [PubMed: 17493920]

McKone E (2004). Isolating the special component of face recognition: peripheral identification and a Mooney face. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(1), 181.

McKone E, Brewer JL, MacPherson S, Rhodes G, & Hayward WG (2007). Familiar other-race faces show normal holistic processing and are robust to perceptual stress. Perception, 36(2), 224–248. [PubMed: 17402665]

McKone E, Martini P, & Nakayama K (2003). Isolating holistic processing in faces (and perhaps objects). Perception of faces, objects, and scenes: Analytic and holistic processes, 92–119.

Miller AL, & Sheldon R (1969). Magnitude estimation of average length and average inclination. Journal of Experimental Psychology, 81(1), 16. [PubMed: 5812170]

Mooney CM (1957). Age in the development of closure ability in children. Canadian Journal of Psychology/Revue canadienne de psychologie, 11(4), 219.

Moore C, & Cavanagh P (1998). Recovery of 3D volume from 2-tone images of novel objects. Cognition, 67(1-2), 45–71. [PubMed: 9735536]

Moscovitch M, Winocur G, & Behrmann M (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. Journal of Cognitive Neuroscience, 9(5), 555–604. [PubMed: 23965118]

Neumann MF, Schweinberger SR, & Burton AM (2013). Viewers extract mean and individual identity from sets of famous faces. Cognition, 128(1), 56–63. [PubMed: 23587844]

Oriet C, & Corbett J (2008). Evidence for rapid extraction of average size in RSVP displays of circles. Journal of Vision, 8(6), 13.

Parkes L, Lund J, Angelucci A, Solomon JA, & Morgan M (2001). Compulsory averaging of crowded orientation signals in human vision. Nature Neuroscience, 4(7), 739. [PubMed: 11426231]

Pelli DG (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10(4), 437–442. [PubMed: 9176953]

Phillips LT, Slepian ML, & Hughes BL (2018). Perceiving groups: The people perception of diversity and hierarchy. Journal of Personality and Social Psychology, 114(5), 766–785. [PubMed: 29337582]

Piazza EA, Sweeny TD, Wessel D, Silver MA, & Whitney D (2013). Humans use summary statistics to perceive auditory sequences. Psychological Science, 24(8), 1389–1397. [PubMed: 23761928]

Ramachandran VS, Armel C, Foster C, & Stoddard R (1998). Object recognition can drive motion perception. Nature, 395(6705), 852. [PubMed: 9804417]

Raymond JE, Shapiro KL, & Arnell KM (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink?. Journal of Experimental Psychology: Human Perception and Performance, 18(3), 849. [PubMed: 1500880]

Rossion B, & Boremanse A (2008). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: Evidence from the face composite illusion. Journal of Vision, 8(4), 3.

Schwiedrzik CM, Melloni L, & Schurger A (2018). Mooney face stimuli for visual perception research. PloS One, 13(7), e0200106. [PubMed: 29979727]

Sekuler AB, Gaspar CM, Gold JM, & Bennett PJ (2004). Inversion leads to quantitative, not qualitative, changes in face processing. Current Biology, 14(5), 391–396. [PubMed: 15028214]

Shapiro KL, Raymond JE, & Arnell KM (1994). Attention to visual pattern information produces the attentional blink in rapid serial visual presentation. Journal of Experimental Psychology: Human Perception and Performance, 20(2), 357. [PubMed: 8189198]

Sweeny TD, Haroz S, & Whitney D (2012). Reference repulsion in the categorical perception of biological motion. Vision Research, 64, 26–34. [PubMed: 22634421]

Sweeny TD, Haroz S, & Whitney D (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. Journal of Experimental Psychology: Human Perception and Performance, 39(2), 329. [PubMed: 22708744]

Sweeny TD, & Whitney D (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. Psychological Science, 25(10), 1903–1913. [PubMed: 25125428]

Sweeny TD, Wurnitsch N, Gopnik A, & Whitney D (2015). Ensemble perception of size in 4–5-year-old children. Developmental Science, 75(4), 556–568.

Valentine T (1988). Upside-down faces: A review of the effect of inversion upon face recognition. British Journal of Psychology, 79(4), 471–491. [PubMed: 3061544]

Wasserstein J, Barr WB, Zappulla R, & Rock D (2004). Facial closure: interrelationship with facial discrimination, other closure tests, and subjective contour illusions. Nenropsychologia, 42(2), 158–163.

Watamaniuk SN, & Duchon A (1992). The human visual system averages speed information. Vision Research, 32(5), 931–941. [PubMed: 1604862]

Watamaniuk SN, & McKee SP (1998). Simultaneous encoding of direction at a local and global scale. Perception & Psychophysics, 60(2), 191–200. [PubMed: 9529903]

Watamaniuk SN, Sekuler R, & Williams DW (1989). Direction perception in complex dynamic displays: the integration of direction information. Vision Research, 29(1), 47–59. [PubMed: 2773336]

Whitney D, Haberman J, & Sweeny TD (2014). From textures to crowds: multiple levels of summary statistical perception. The New Visual Neurosciences, 695–710.

Whitney D, & Yamanashi Leib A (2018). Ensemble perception. Annual Review of Psychology, 69, 105–129.

Wolfe BA, Kosovicheva AA, Yamanashi Leib A, Wood K, & Whitney D (2015). Foveal input is not required for perception of crowd facial expression. Journal of Vision, 15(4), 11.

Yamanashi Leib A, Fischer J, Liu Y, Qiu S, Robertson L, & Whitney D (2014). Ensemble crowd perception: A viewpoint-invariant mechanism to represent average crowd identity. Journal of Vision, 14(8), 26.

Yamanashi Leib A, Kosovicheva A, & Whitney D (2016). Fast ensemble representations for abstract visual impressions. Nature Communications, 7, 13186.

Yamanashi Leib A, Landau AN, Baek Y, Chong SC, & Robertson L (2012). Extracting the mean size across the visual field in patients with mild, chronic unilateral neglect. Frontiers in Human Neuroscience, 6, 267. [PubMed: 23060771]

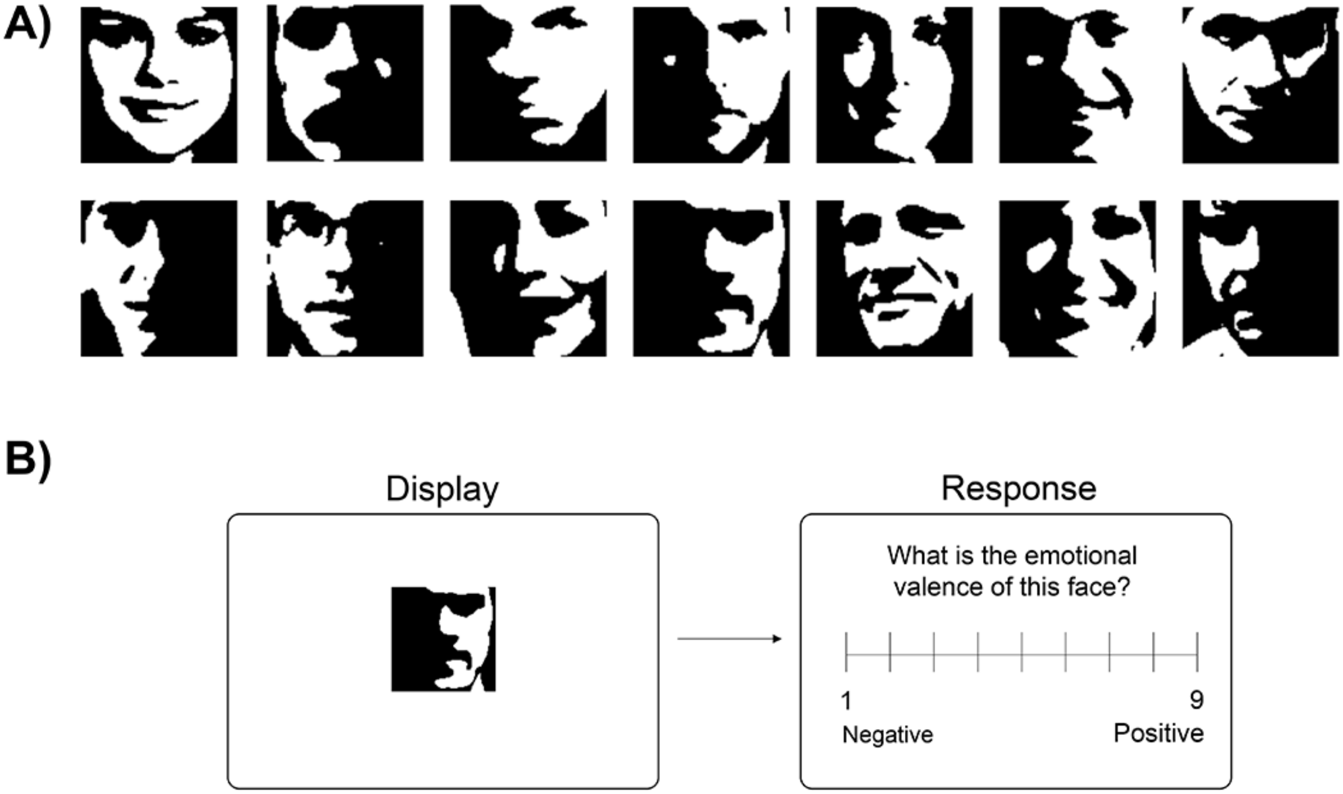

Figure 1. Example stimuli and Experiment 1 procedure.

A) Example stimuli. Fourteen example stimuli from the stimulus set used in Experiments 1, 2, 3a, 3b, and 4. The stimulus set contained 136 Mooney faces generated using a GAN network on celebrity pictures taken from different viewpoints and under different lighting conditions. **B)** Experiment 1 **procedure.** In Experiment 1, Amazon Mechanical Turk participants viewed a single Mooney face for 1 second. After the image disappeared, participants rated the emotional valence of the Mooney face using a 9-point Likert scale, ranging from negative to positive.

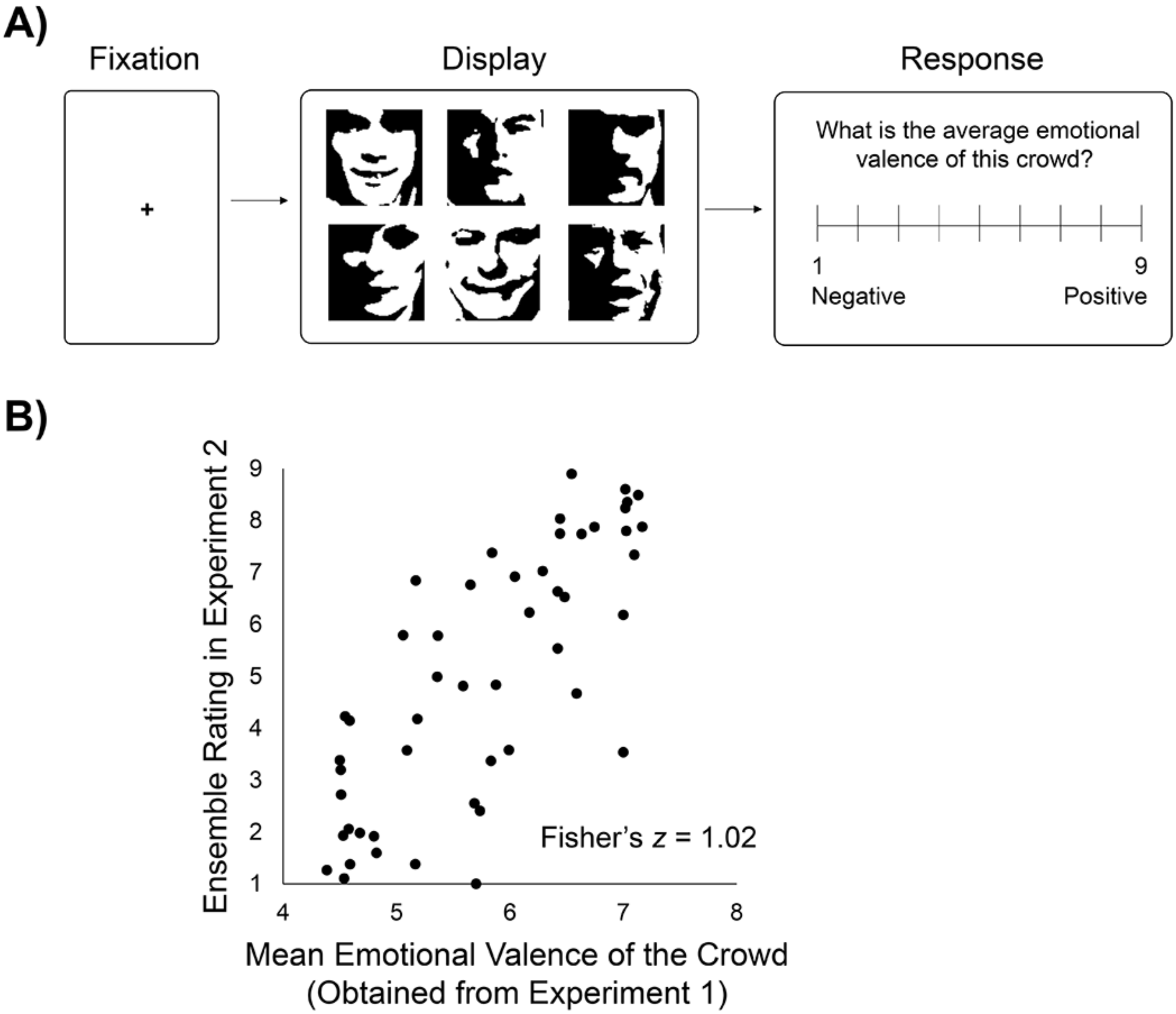

Figure 2. Experiment 2 procedure & a representative subject's data.

A) Experiment 2 **procedure.** A set of six Mooney faces was viewed for one second, followed by a continuous rating scale. No time limit was imposed on the rating phase. **B) Representative data for a single subject.** We computed the mean emotional valence of the crowds based on the single face ratings obtained from Experiment 1 and plotted these values on the x-axis. We plotted this subject's ratings of the Mooney crowds on the y-axis. The correlation between the two axes, as measured by Fisher's *z*, indicates ensemble sensitivity.

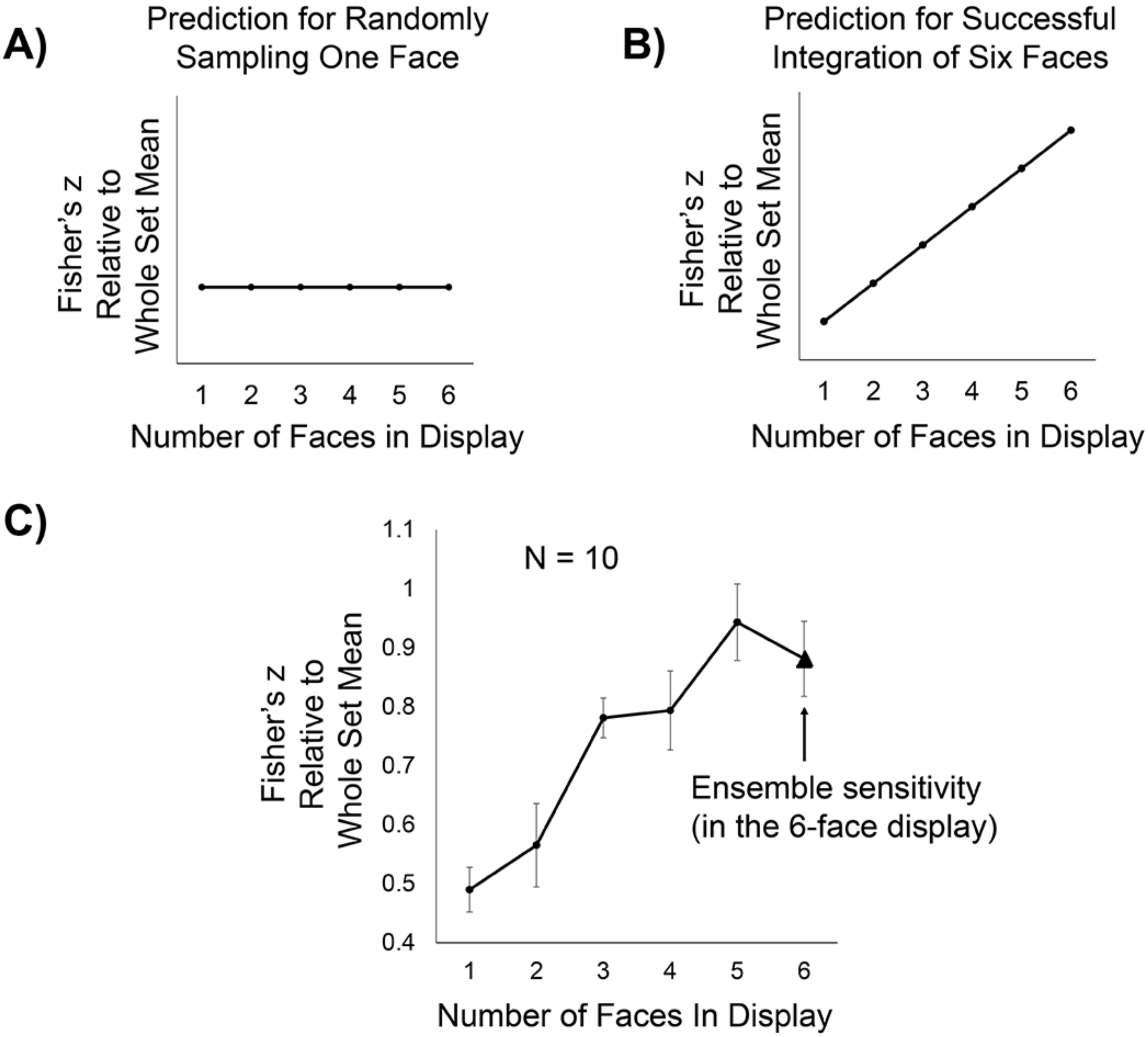

Figure 3. Experiment 2 predictions & results.

A) Prediction for sampling one face. Predicted pattern of performance for randomly sampling one face out of the crowd. Ensemble sensitivity (as measured by Fisher's *z* on the y-axis) would not increase, even as more faces were revealed. **B) Prediction for integrating six faces.** Predicted pattern of performance if all six faces were integrated into the ensemble percept. Ensemble sensitivity (y-axis) would systematically increase as the number of faces revealed increased. **C)** Experiment 2 **results.** Group results of the subset analysis in Experiment 2. Ensemble sensitivity (y-axis) monotonically increased as the number of faces revealed (x-axis) increased. This is evidence that subjects were integrating 3-6 faces into their ensemble percept.

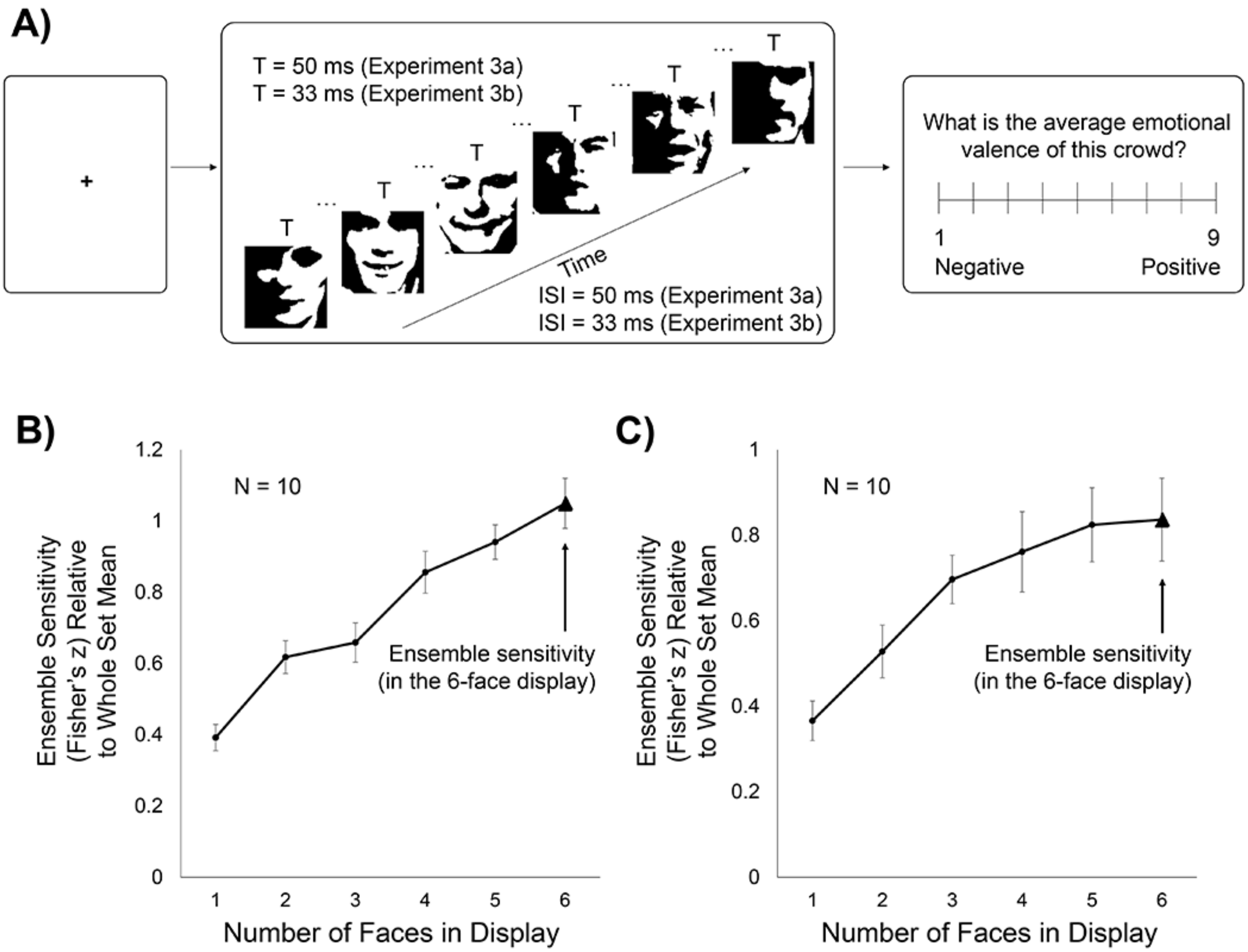

Figure 4. Experiments 3a & 3b procedure & results.

A) Experiments 3a & 3b **procedure.** Display phase of Experiments 3a & 3b: In the whole set condition, six faces were sequentially displayed for 50 ms (Experiment 3a) or 33 ms (Experiment 3b) each with a 50 ms (Experiment 3a) or 33 ms (Experiment 3b) inter-stimulus-interval (ISI). Response phase: A continuous scale was presented until the response was made in both experiments. **B & C)** Experiments 3a & 3b **results:** High ensemble sensitivity was observed for Experiments 3a (panel B) and 3b (panel C). The average Fisher's *z* across subjects in each experiment were 1.05 and 0.84, respectively. The subset analysis conducted for both experiments separately demonstrated that ensemble sensitivity increased as more faces were revealed in the display. This monotonic pattern suggests that multiple faces were integrated into the ensemble percept in both experiments.

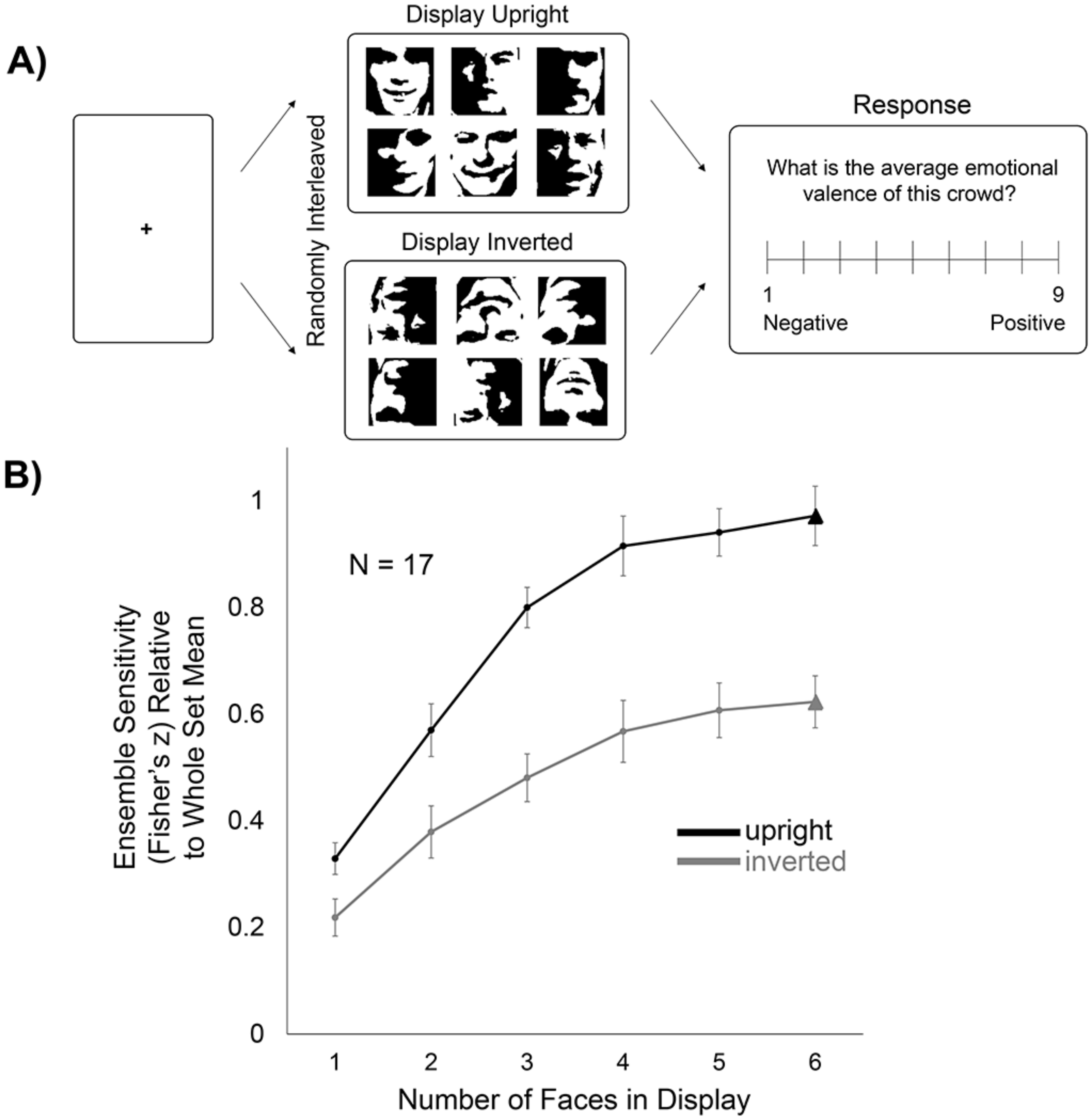

Figure 5. Experiment 4 procedure and results.

A) Experiment 4 **procedure.** Six Mooney faces were displayed for 1 second in upright or inverted orientations (interleaved blocks), followed by a continuous scale presented until response. **B)** Experiment 4 **results.** The subset analysis demonstrates that multiple faces were integrated into the ensemble percept in both viewing conditions. An analysis of the slope of the data indicated that participants integrated fewer faces in the inverted viewing condition.

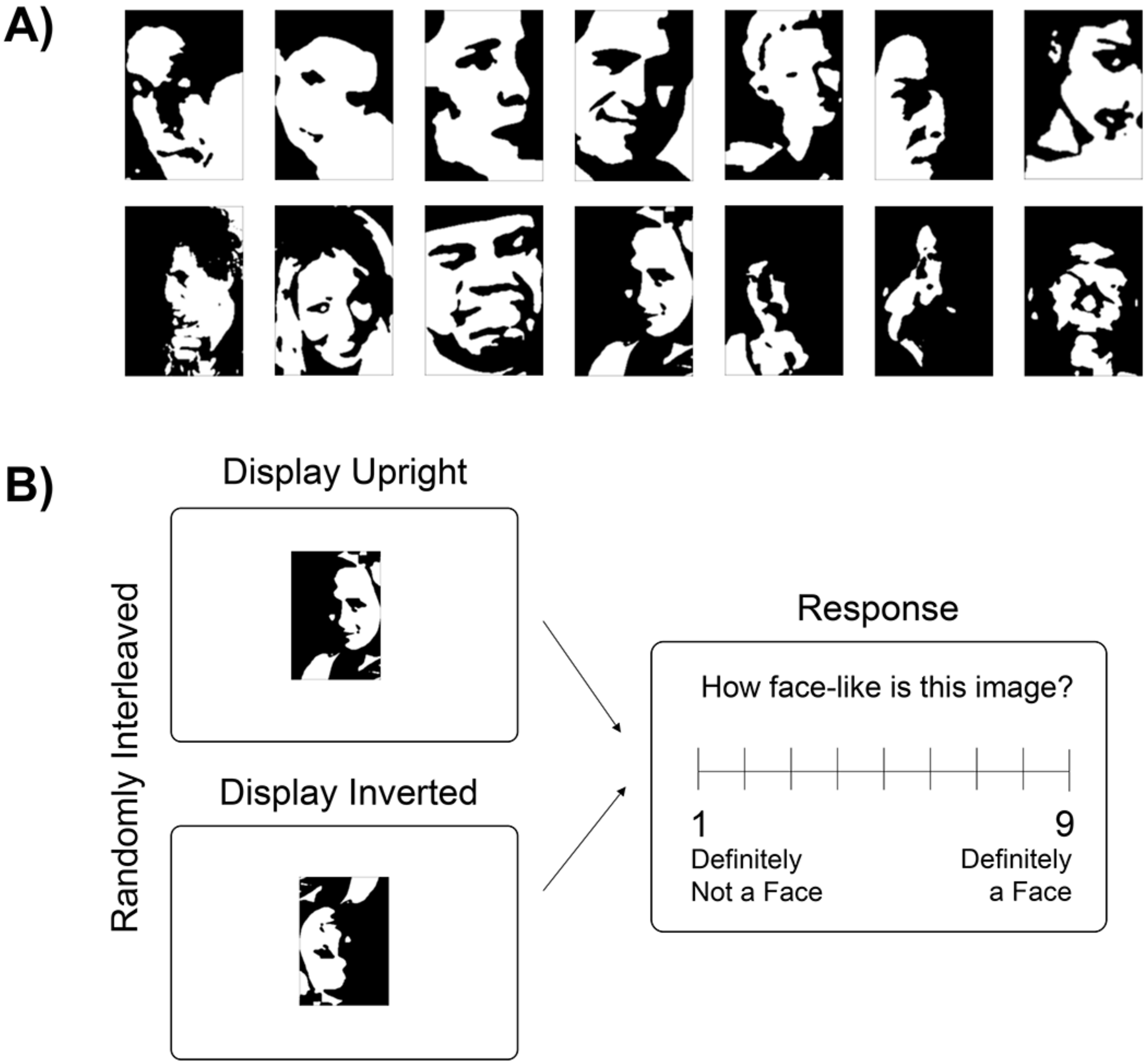

Figure 6. New Mooney face stimulus set & Experiment 5 procedure.

A) Example stimuli. Examples of Fourteen Mooney faces from the new stimulus set used in Experiments 5, 6, 7a, & 7b. **B)** Experiment 5 **procedure.** A single Mooney face was displayed for one second in upright or inverted viewing orientations (randomly interleaved trials). Subsequently, participants reported how "face-like" the image was using a continuous scale.

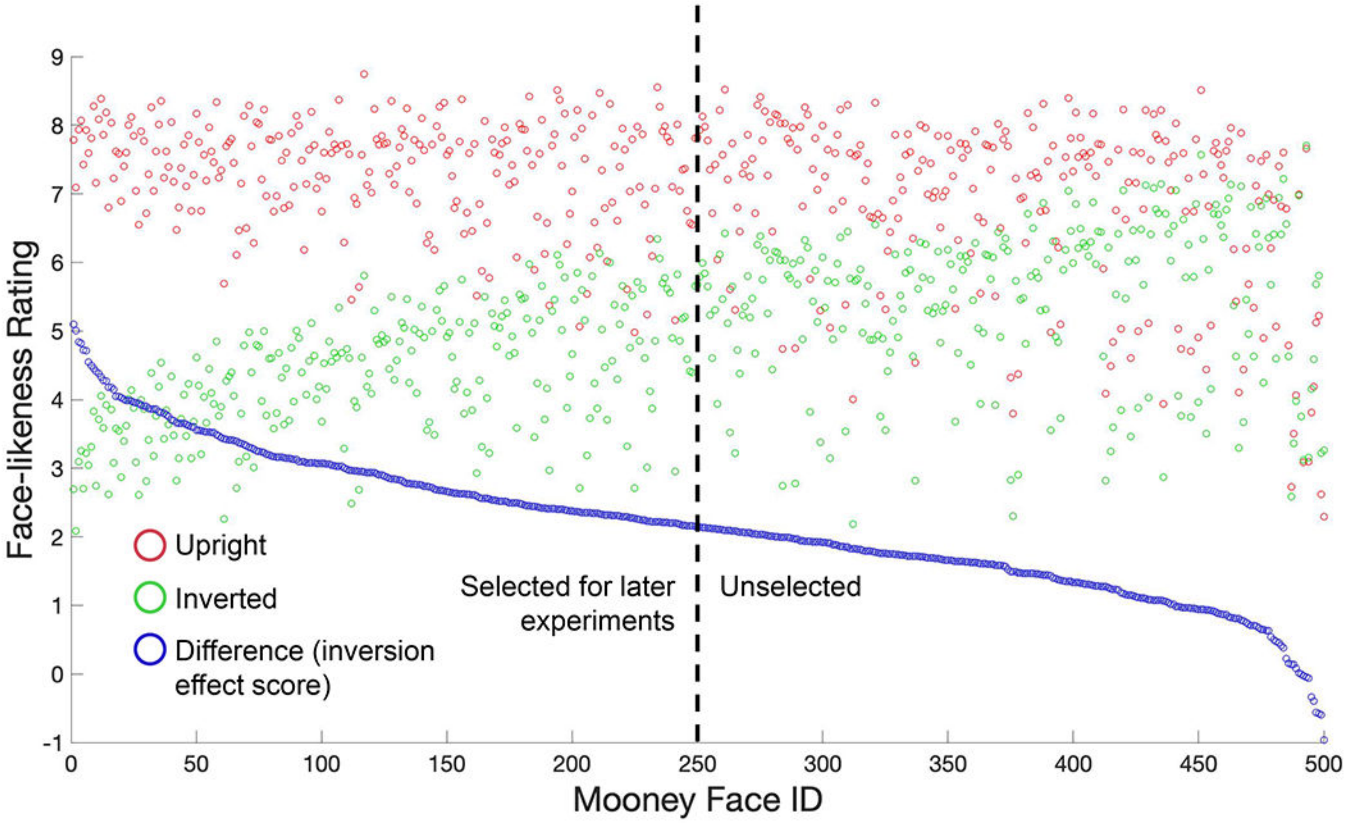

Figure 7. An illustration of the rationale for selecting maximally holistic Mooney faces.

Eleven observers rated the face-likeness of 500 Mooney faces. Each face was displayed in both upright and inverted viewing orientations. We computed the inversion effect scores by subtracting the inverted ratings from the upright ratings. Half (250) of the Mooney faces with higher inversion effect scores were identified as “more holistic,” being selected for subsequent experiments.

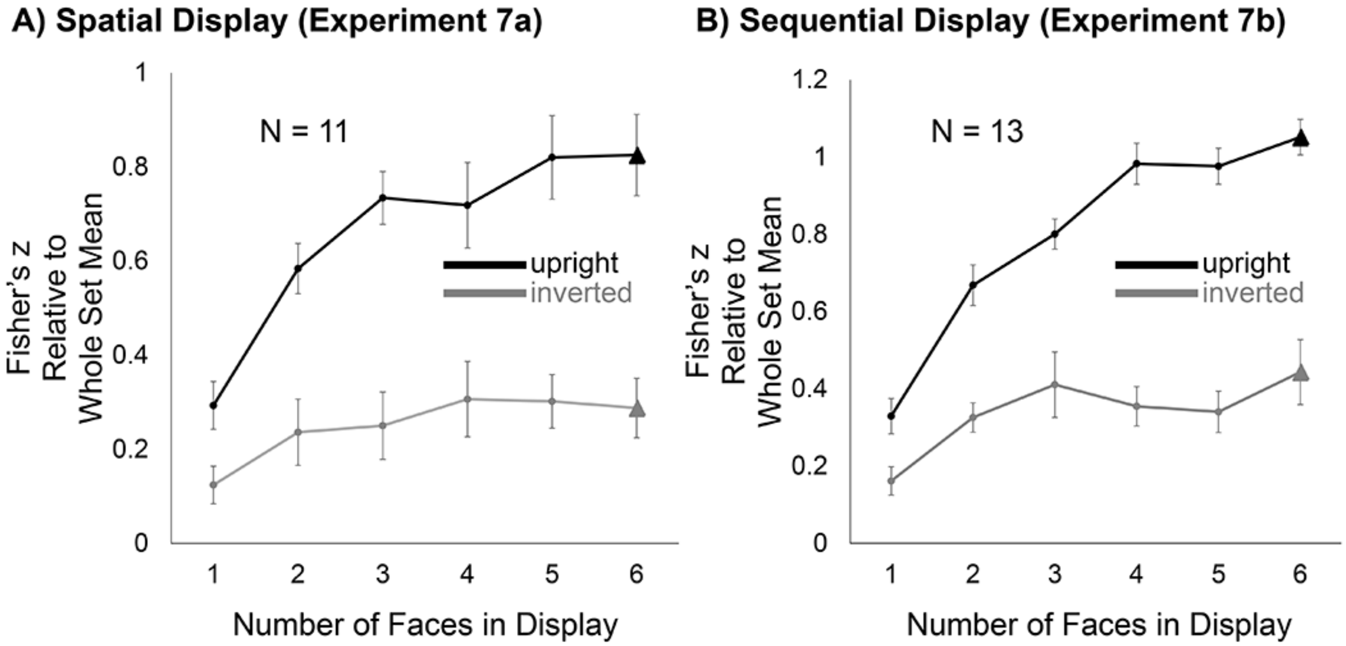

Figure 8. Experiments 7a & 7b results.

Experiments 7a & 7b utilized stimuli that required maximal holistic processing. **A)** Experiment 7a **results.** In Experiment 7a, we displayed 6 Mooney faces simultaneously in a spatial array. We observed higher ensemble sensitivity in the upright condition. The subset analysis of Experiment 7a showed that participants integrated multiple faces in the upright viewing condition (black line), but the integration effect was poor in the inverted viewing condition (gray line). **B)** Experiment 7b **results.** In Experiment 7b, we observed high ensemble sensitivity to upright Mooney face crowds displayed in a rapid temporal sequence (black line). Ensemble sensitivity increased across subset sizes. We found lower ensemble sensitivity to inverted Mooney face crowds (gray line). The subset analysis demonstrated that participants integrated multiple faces in both upright and inverted RSVP conditions; however, slope analysis indicated that participants integrated fewer faces in the inverted condition.