

# Judging Emotion in Natural Images of Crowds

Susan Hao<sup>1</sup>, David Whitney<sup>1, 2, 3</sup>, and Sonia J. Bishop<sup>1, 2, 3, 4, 5</sup>

<sup>1</sup> Department of Psychology, University of California, Berkeley

<sup>2</sup> Helen Wills Neuroscience Institute, University of California, Berkeley

<sup>3</sup> Vision Science Program, University of California, Berkeley

<sup>4</sup> School of Psychology, Trinity College Dublin, University of Dublin

<sup>5</sup> Trinity College Institute of Neuroscience, Trinity College Dublin, University of Dublin

It has been suggested that humans use summary statistics such as the average of the emotion of individual faces when they rapidly judge group emotion. Previous studies have mainly used faces of actors posing basic emotions, and morphed versions of these faces, against a plain background. In the present study, photographs taken in real-world settings were used to investigate the influence of mean facial emotion, maximal facial emotion, and background context on judgments of group emotion, assessed using dimensional ratings of valence, arousal, and dominance. Background context explained a significant amount of unique variance in group ratings for each dimension. Mean emotion explained additional unique variance for valence ratings, whereas maximal emotion explained additional unique variance for arousal, with dominance showing more mixed results. Removing background context and disrupting the contextual and spatial relationship between faces by randomly replacing faces with ones from other images within the stimulus set increased reliance on mean emotion. However, under all conditions, the maximally arousing face continued to exert an influence on ratings of group arousal, in line with theoretical accounts arguing for a unique bottom-up effect of emotional arousal on attentional competition and postattentive perceptual processing. Together these findings suggest that individuals' reliance on average emotion when judging crowd scenes differs as a function of the dimension of affect. In addition, the presence of background context both directly impacts judgments of crowd emotion and modulates the relative influence of maximal versus mean emotion on these judgments.

Keywords: ensemble, emotion, context, facial expression, natural image

Supplemental materials: https://doi.org/10.1037/emo0001358.supp

In the visual world, there is an overabundance of information in the environment around us. Our visual system is incapable of processing a large amount of this information, yet our percept of the world is rich and undisrupted. Previous studies have suggested that this rich, gist-level percept relies, in part, on the use of ensemble information. Ensemble perception, that is, the perception of groups of similar objects using summary statistics, is argued to happen at every level of visual processing (Whitney & Yamanashi Leib, 2018). In the case of facial expressions, findings have suggested that when adult human participants judge the emotion of a crowd, they take the mean across individual facial expressions (Haberman & Whitney, 2007, 2009; Haberman et al., 2009; Im et al., 2017; Li et al., 2016; Sun & Chong, 2020).

Most of the studies to date that have explored the potential use of summary statistics in the perception of facial expression have asked participants to make judgments regarding basic emotions (e.g., happiness, sadness, anger). There have been far fewer studies using dimensional measures of affect. Russell and colleagues conducted seminal work on dimensional models of affect. The two dimensions proposed by Russell (1980), now typically referred to as valence

This article was published Online First August 29, 2024. Timothy Sweeny served as action editor.

Susan Hao D https://orcid.org/0000-0002-6805-959X

David Whitney (b) https://orcid.org/0000-0002-4531-5570

Sonia J. Bishop D https://orcid.org/0000-0001-7833-3030

Data is available via the open science framework (https://www.osf.io; https://osf.io/vk8nr/?view\_only=8363d0025d544f729d91f41f924d4fbc). This work was not preregistered. Portions of these findings were presented at the 2020 Vision Sciences Society online conference as a poster.

The authors have no known conflicts of interest to disclose. Data collection was supported by the National Institute of Mental Health (Grant R01MH112541) awarded to Sonia J. Bishop. The authors thank Alison Yamanashi Leib, and Zhimin Chen for their assistance with analyses.

This work is licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International License (CC BY-NC-ND 4.0; https://creativecommons.org/licenses/by-nc-nd/4.0). This license permits copying and redistributing the work in any medium or format for noncommercial use provided the original authors and source are credited and a link to the license is included in attribution. No derivative works are permitted under this license.

Susan Hao played a lead role in data curation, formal analysis, investigation, methodology, visualization, writing–original draft, and writing–review and editing. David Whitney played a supporting role in conceptualization, supervision, and writing–review and editing. Sonia J. Bishop played a lead role in funding acquisition and supervision, a supporting role in formal analysis, investigation, methodology, resources, visualization, writing–original draft, and writing–review and editing, and an equal role in conceptualization.

Correspondence concerning this article should be addressed to Sonia J. Bishop, Trinity College Institute of Neuroscience, Trinity College Dublin, University of Dublin, Lloyd Building, Dublin 2, Ireland. Email: bishops@tcd.ie (positive to negative) and arousal (low to high excitability or intensity) are widely used in the emotion literature. Russell and Mehrabian (1977) initially proposed a third dimension of dominance-submissiveness related to potency. While there has been debate regarding this dimension, recent work in the face literature using natural images has indicated that dominance is a key dimension on which natural faces are evaluated (Oosterhof & Todorov, 2008; Sutherland et al., 2013). A relationship has also been shown between basic emotions and ratings of dominance, with the negative, high arousal emotions of fear and anger being differentiated by perceived dominance-ratings being higher for faces showing anger than for those showing fear (Hareli et al., 2009). Within the summary statistic literature, Han et al. (2021) demonstrated that participants were able to correctly judge the average valence of Mooney face crowds. Phillips et al. (2018) also demonstrated that participants were influenced by mean dominance when rating the dominance of groups of computer-generated faces. Here, differences in perceived dominance were achieved by varying facial structure rather than facial expression.

To our knowledge, there has been no work addressing whether judgments of the emotional arousal of groups is consistent with the use of summary statistics. Evidence from the attention literature suggests that highly arousing stimuli may be preferentially processed. In their arousal-biased competition model, Mather and Sutherland (2011) argued that arousal modulates the strength of competing mental representations in a manner that favors highly arousing stimuli. This is held to explain phenomena such as "pop out" which is observed when a single high arousal stimulus is rapidly detected, regardless of the number of surrounding neutral stimuli (Lundqvist et al., 2015; Pinkham et al., 2010). According to Mather and Sutherland's model (2011), arousal influences selective attention regardless of stimulus valence. In line with this, studies examining attentional orienting and disengagement have found that arousal modulates selective attention to a greater extent than valence (Ossenfort & Isaacowitz, 2021; Vogt et al., 2008). However, it remains an open question as to whether the relative salience of individual faces in a group setting will also impact the perception of group affect. One way this might be observed is in terms of a greater reliance on "max" than "mean" emotion for judgments of facial arousal than for face valence. In other words, participants might be disproportionately influenced by the highest arousal face when judging the arousal of groups of individuals, but this might not be observed to an equivalent extent for facial valence or other dimensions of affect.

Another important question pertains to whether reliance on summary statistics might differ for natural faces versus posed or computer-generated faces presented without background context. To date, most ensemble perception studies that have examined facial expression have only used a limited number of emotional expressions with faces in isolation without context. Further, while there are studies which have shown that background context affects face emotion perception (Alwis & Haberman, 2020; Barrett & Kensinger, 2010; Righart & de Gelder, 2008a, 2008b; Sasson et al., 2016), many of these have artificially inserted faces into a scene, ignoring the background context's orienting role in the perception of facial emotion. In an initial study using naturalistic stimuli to understand the background context's influence on participants' perception of individual facial expressions, Chen and Whitney (2019) investigated the extent to which affective ratings of background context—derived by taking a scene and blurring out a central character—predicted affective ratings of the complete scene. A trackball was used by participants to rate affect in a two-dimensional valence by arousal space. Across both movies and naturalistic footage, context explained nearly as much unique variance in the ratings of scene affect as that explained by affect ratings of the central character with the background blurred out. To the extent that other characters were present in the scene, they were included within the background context, and ensemble perception processes were not investigated. Here, we build on this prior work by examining the extent to which background context influences the processing of the facial expression of groups, asking whether background context alters reliance on summary statistics, and whether the impact of background context varies as a function of the dimension of affect in question.

With these questions in mind, our aims were as follows: (a) to investigate participants' relative reliance on mean versus maximal emotion when judging crowd emotion; (b) to determine if this varies across different dimensions of affect (namely valence, arousal, and dominance); (3) to investigate the influence of background context on judgments of crowd emotion—in particular to examine if background context has an additive effect in predicting crowd emotion ratings and/or alters reliance on mean versus maximal emotion; and (d) to understand whether the extent to which statistical dependencies in visual scenes are intact or disrupted impacts the use of summary statistics.

# **Experiment 1**

# Method

# **Participants**

Participants were recruited via Amazon's Mechanical Turk platform. Experimental procedures including recruitment were approved by the University of California, Berkeley Committee for the Protection of Human Subjects. All participants provided informed consent (online tick box). Participants were required to reside in the United States and could only complete one experimental condition; data from any participant who completed more than one condition or had an internet protocol address that was outside of the United States were excluded. Data from participants who had greater than 50 trials in which the stimuli did not load because of internet connection issues were also excluded. Data from 325 participants (192 male, 133 female) between the ages of 18 and 40 were retained (see Supplemental Table S1 for additional demographic details).

# Stimuli

Stimuli were obtained from Google Images search queries. The stimulus set was formed using natural scenes of crowds photographed in real-world settings containing individuals of different genders, races, and ages showing varying expressions. Stimuli used as *Full Images* (n = 168) comprised crowd scenes containing 3 to 5 emotional faces, with each image being  $500 \times 500$  pixels in size (Figure 1, Column 1). Stimuli used in the *Context Condition* (n = 168) comprised the same images with the individual faces cropped out to leave only the background context of the scene (Figure 1, Column 2). We opted to remove all information about the face by cropping out individual faces instead of using methods such as blurring which may still leave small, but difficult to measure, amounts of visual





*Note.* In the left-hand column, we show three example *Full Images* stimuli. *Full Images* comprised 500 × 500 pixel images of naturalistic scenes with three to five emotional faces. The second column shows the corresponding stimuli presented in the *Context Condition*. Here, the individual faces were cropped out of each full image leaving just the background context. As mean luminance and color varied across images, we set the cropped-out areas to white. The remaining columns show the corresponding stimuli presented in the *Individual Face Condition*. Here, each face in the full image has been cropped out and presented separately. Example images included in this figure are in the public domain (https://commons.wikimedia.org) or licensed for use in supporting educational texts (Alamy) as described under Alamy editorial licensing policy (https://www.alamy.com/). See the online article for the color version of this figure.

information (Bombari et al., 2013; Hayward et al., 2008; Karbasi et al., 2018). We note that in some images, both individuals' bodies and faces were present in the image. In these cases, only the faces were cropped out and the bodies were included in the background. Stimuli used in the Individual Face Condition comprised the individual faces cropped out from the original Full Images and were presented separately (Figure 1, right hand columns). These face images were divided into three sets (set 1: n = 214, set 2: n = 222, set 3: n = 200). All faces from the same original image were kept within the same set. The order of faces within a given set was randomized. Sixty-two participants rated the Full Image stimuli, 62 participants rated the Context stimuli, and 201 participants rated the Individual Face stimuli (n = 62 for set 1, n = 73 for set 2, n = 66 for set 3). Following Yamanashi Leib et al. (2020), we used a two-way random effects intraclass correlation model to measure consistency in ratings across raters. This revealed very high inter-rater consistencies (Supplemental Table S2).

#### **Experimental Procedure**

Ratings of stimuli were conducted online using Qualtrics. At the beginning of the experimental session, participants were asked to give consent online. Participants were then asked to maximize the window in which the stimuli were presented and to sit an arm's length distance away from the display. Participants viewed a fixation square for 1 s followed by a stimulus image for 1 s. Valence, arousal, and dominance were rated on a 9-point Likert scale. The dimensions

were defined as follows: (a) "By valence, we mean how negative or positive the emotion of the scene is;" (b) "By arousal, we mean how calm (low arousal) or exciting (high arousal) the emotion of the scene is"; (c) "High dominance is associated with the actual assumed or projected possession of power, low dominance is associated with submissiveness."

Participants were given unlimited time to make their ratings. All participants were given instructions defining each emotional dimension along with examples at the beginning of the experimental session. The full instructions can be found in the Supplemental Materials. There were four practice trials prior to the start of the task.

Participants rating the *Full Image* were asked to rate the average valence, arousal, and dominance of the faces in the crowd (see Figure 2). In the *Context Condition*, participants were asked to rate the average valence, arousal, and dominance of the scene. In the *Individual Face Condition*, participants were asked to rate the valence, arousal, and dominance of the face.

## Data Analysis

Descriptive statistics of the average raw scores for each rating condition can be found in Supplemental Table S3. Each participant's ratings were *z*-scored across images. Next, for each *Full Image* stimulus, scores on each dimension were averaged across participants to obtain a single score for that image on the given dimension. Following the approach used in prior ensemble studies (e.g., Leib et al., 2016), the rating of crowd emotion for each *Full Image* was



Figure 2 Rating Procedure for Full Images

*Note.* A fixation square was shown for 1 s followed by the stimulus, an image of a naturalistic scene of a crowd, for 1 s. Participants were then asked to rate the average valence, arousal, and dominance of the faces in the scene. Response time was not constrained. Image used in this Figure is in the public domain (https://commons.wikimedia.org). See the online article for the color version of this figure.

used as the dependent variable in each of the regressions described below. Here, this was conducted separately for each of the three dimensions of affect under consideration: valence, arousal, and dominance. Subsequently, for each individual face in each image, ratings from participants who completed the Individual Faces Condition were averaged across participants to obtain a single score for that face on each dimension of affect. These scores were used to create the following regressors: (a) The MEAN facial emotion regressor for each dimension of affect was calculated using the mean of the scores for all individual faces present in a given image on the dimension in question; (b) For the dimensions of arousal, valence, and dominance, the MAX facial emotion regressors used the score for the most arousing face, the most positive face, and the most dominant face in each image, respectively; (c) The MIN facial emotion regressors used the score for the least arousing face, the least positive (most negative) face, and the least dominant face in each image; and (d) The rating on each dimension of each image with the faces cropped out (as obtained from participants who completed the Context condition) was averaged across participants. This was used to create the CONTEXT regressor for each dimension of affect.

# **Regression Analyses**

Regression analyses were conducted across images with ratings averaged across participants; as described above, separate groups of

participants provided ratings for predictor and dependent variables. Power analyses indicated that for a regression model with four regressors (the most used), 108 images would give an  $\alpha$  of p < .05for a medium effect size of  $f^2 = .15$ , and power level of .9; with 145 images giving an  $\alpha$  of p < .01. We used 168 images. For each of the three dimensions of affect, we first conducted a forward stepwise regression with ratings of crowd emotion (Full Image ratings) as the dependent variable and MEAN, MAX, MIN facial emotion regressors and the CONTEXT regressor entered as potential predictors. A stopping threshold of p < .05 was used (i.e., the predictor that explained most variance was entered at each step, with only features that significantly improved model fit at p < .05 being allowed to enter each model). Forward stepwise regression gives insight into which variable out of several candidates is the strongest predictor of the dependent variable of interest and also addresses the extent to which other variables explain additional significant variance. However, each predictor variable entered may contain shared variance with other variables, and if two predictors are highly correlated, this does not allow for understanding of the unique variance explained by each. Hence, to understand the unique contribution of each predictor variable, we also conducted a drop column feature importance analysis for each dimension of affect. Here, we included MAX facial emotion, MEAN facial emotion, and CONTEXT regressors as features of interest. We did not include MIN facial emotion as this was not a significant predictor in any of the three stepwise regressions (see Results and Table 1). In the drop column

Table 1Stepwise Regression

Feature	β	t	Significance
Valence			
Intercept	-0.08	-2.23	p = .027
MEAN	0.74	9.10	p < .001
CONTEXT	0.35	7.38	p < .001
MAX	0.19	2.73	p = .007
Arousal			I
Intercept	-0.36	-14.58	p < .001
MAX	0.68	19.38	p < .001
CONTEXT	0.32	7.24	p < .001
Dominance			P
Intercept	-0.14	-5.52	p < .001
MAX	0.41	8.27	p < .001
CONTEXT	0.45	8.27	p < .001

Note. Three forward stepwise regression analyses were conducted with ratings of *Full Image* crowd affect as the dependent variable. The features entered into the final model for each dimension of affect (valence, arousal, and dominance) are presented here together with feature  $\beta$  weights, *t*-values for the estimated weights, and associated *p*-values. The variables considered included the maximal facial emotion (MAX) regressor, the mean facial emotion (MEAN) regressor (MIN), and the CONTEXT regressor for the dimension in question. The MIN regressor did not enter the regression model for any of the three dimensions of affect.

feature importance analyses conducted, all features of interest were entered into a regression model predicting *Full Image* ratings on a given dimension of affect. Each feature, in turn, was dropped from the regression, and the model retrained with all features except the one dropped. The importance of that feature was then estimated as the difference between the model performance (adjusted  $R^2$ ) with all features as predictors and model performance with all features except the one dropped. Permutation tests were performed to obtain the null distribution of feature importance scores to determine the significance of the unique contribution of each feature (significance was assessed using p < .05).

# Data Sharing

Data are shared online via the open science framework (https://www.osf.io/vk8nr).

# Results

#### Stepwise Regression

For valence, the features that entered the final stepwise regression model ( $R_{adj}^2 = .87$ ) (i.e., that predicted ratings of crowd valence for each *Full Image*) were the MEAN facial emotion regressor (p < .001), the CONTEXT regressor (p < .001), and the MAX facial emotion regressor (p = .007), see Table 1. For arousal, the features included in the final stepwise regression model ( $R_{adj}^2 = .80$ ) were the MAX facial emotion regressor (p < .001), and the CONTEXT regressor (p < .001) and the CONTEXT regressor (p < .001) and the CONTEXT regressor (p < .001), see Table 1. Similarly, for dominance, the MAX facial emotion regressor (p < .001) and the CONTEXT regressor (p < .001) were included in the final stepwise regression model ( $R_{adj}^2 = .61$ ), see Table 1.

# Drop Column Feature Importance

For each dimension of affect, we conducted a drop column feature importance analysis to examine the unique variance explained by the MAX facial emotion, MEAN facial emotion, and CONTEXT regressors (see Figure 3). Permutation tests (10,000 shuffles) were performed to test whether each feature's unique  $R_{adj}^2$  was significant. For valence, we found that both the MEAN facial emotion regressor (unique  $R_{adj}^2 = .06$ , p < .001) and the CONTEXT regressor (unique  $R_{adj}^2 = .04$ , p = .003) made significant unique contributions to the prediction of crowd valence ratings.

For arousal, the MAX facial emotion regressor (unique  $R_{adj}^2 = .15$ , p < .001) and the CONTEXT regressor (unique  $R_{adj}^2 = .06$ , p < .001) both made significant unique contributions to the prediction of crowd arousal ratings. For dominance, the CONTEXT regressor (unique  $R_{adj}^2 = .15$ , p < .001) and the MAX facial emotion regressor (unique  $R_{adj}^2 = .04$ , p = .005) both made significant unique contributions to the prediction of crowd dominance ratings. For total variance explained by all three regressors for each dimension of affect, including shared variance, see Supplemental Figure S1. As expected, this was close to the  $R_{adj}^2$  for the winning stepwise models as described above.

These results suggest that CONTEXT influences judgment of crowd emotion across all three dimensions of affect. The influence of CONTEXT was especially notable for dominance ratings. In addition to CONTEXT, MEAN facial emotion explained significant unique variance in the judgments of crowd valence while the maximum (MAX) facial emotion regressor explained significant unique variance in judgments of crowd arousal and dominance.

### **Control Analyses**

As reported above, we observed an influence of MAX facial emotion as opposed to MEAN facial emotion for two of our three dimensions of affect. This might reflect reduced reliance on summary statistics when judgments of crowd affect are made in relation to natural scenes. However, using natural images raises the potential for confounds that are not present for artificially generated experimental stimuli. In particular, the faces within our images often varied in size. To ensure that the apparent effect of MAX arousal and MAX dominance were not simply driven by a confounding relationship with face size, we took MAX size arousal and MAX size dominance ratings (i.e., ratings of arousal and dominance for the largest face in the image) and used them to predict affect ratings for the full image, retaining residual scores in each case. We then repeated the drop column analysis described above using the residual full image ratings as the dependent variable. In this manner we controlled for the influence of the largest face on the ratings of the full image. These analyses revealed that the maximally arousing and maximally dominant face in each image contributed significantly to the prediction of full image ratings even when the variance explained by the arousal or dominance ratings of the largest face had been residualized out,  $(R_{adj}^2 = .10, p < .001; R_{adj}^2 = .03, p = .018,$  respectively, Supplemental Figure S2).

Another important question is whether the influence of the maximally dominant face on judgments of crowd dominance might be secondary to the findings for arousal, or indeed vice versa. If the most dominant face is also often the most arousing, then any attentional orienting toward the most arousing face, leading to



Figure 3 Drop Column Feature Importance Analysis for Experiment 1

*Note.* A drop column feature importance analysis was performed for each dimension of affect using the max facial emotion (MAX), mean facial emotion (MEAN), and context regressors as features of interest. The results for each dimension of affect are illustrated in the three panels presented here. In each panel, each bar represents the unique variance explained by a given feature; this is calculated as the difference in adjusted  $R^2$  for the model with every feature versus that for the model with all features except the feature of interest. Scattered dots overlaying each bar show the permuted null distribution of the unique adjusted  $R^2$  values. Left panel: mean facial valence and context contributed significantly to ratings of *Full Image* crowd arousal. Right panel: the most dominant face (max facial dominance) and context contributed significantly to ratings of *Full Image* crowd dominance. See the online article for the color version of this figure. \*\* p < .01.

preferential processing, might also influence dominance ratings. The reverse could also apply. To address this, we reconducted the drop column feature importance analysis for dominance on the residual scores obtained after using dominance ratings for the maximally arousing face to predict full image dominance ratings. We additionally conducted a drop-column analysis for arousal on the residual scores obtained after using arousal ratings for the most dominant face to predict full image arousal ratings. These analyses revealed that the maximally arousing face and the maximally dominant face still contributed significantly to the prediction of arousal and dominance ratings, respectively, for the full image  $(R_{adj}^2 = .12, p < .001, R_{adj}^2 = .02, p = .036$ , respectively, Supplemental Figure S3).

In the case of valence, the scale's ends are defined as negative (rating of 1) to positive (rating of 9). In the analyses reported above, we used the face with the highest score (i.e., most positive expression) as the MAX face, and the face with the lowest score (i.e., most negative face) as the MIN face. MIN face ratings did not enter into the model identified by our initial stepwise regression. A supplementary drop feature analysis replacing the valence MAX with the valence MIN regressor confirmed that entering the MIN (most negative) face did not account for a significant amount of unique variance in the valence regressions (Supplemental Figure S4).

Another possible issue arising from the use of natural images is that the variance in the emotion of faces in a group might potentially differ as a function of the dimension of affect under consideration. To control for this, we conducted an additional drop-column analysis where we randomly re-sampled 120 images for each dimension until the standard deviation of faces within each image did not differ significantly across the given dimension of affect ( $t_{val\_ars} = 0.08$ ,  $p_{val\_ars} = .936$ ,  $t_{dom\_ars} = 1.42$ ,  $p_{dom\_ars} = .156$ ,  $t_{dom\_val} = 1.33$ ,  $p_{dom\_val} = .185$ ). Despite the slightly reduced power due to the smaller number of images, we continued to observe a significant contribution of mean, but not maximal, facial emotion in the prediction of full image valence ratings and of maximal, but not

mean, emotion in the prediction of full image arousal and dominance ratings (Supplemental Figure S5).

# **Experiment 2**

Most prior summary statistic experiments have not used background context. An interesting question is whether the presence of background context influences participants' reliance on the maximally emotional face versus mean facial emotion when rating group affect. This might arise as a result of information in the background context or in the relative positioning and pose of faces leading to differential reliance on individual faces within the image. We sought to address this through two follow up experiments. We hypothesized that removing background context (Experiment 2a) and disrupting the relationship between faces in each image (Experiment 2b) would result in greater reliance on mean facial emotion and reduced reliance on max facial emotion, but that for arousal, there would still be a unique contribution of the most arousing face no matter where in the image it was located. This hypothesis builds on findings from the attention literature indicating that high arousal stimuli capture spatial attention, with this resulting in enhanced perceptual processing (Mather & Sutherland, 2011; Ossenfort & Isaacowitz, 2021; Vogt et al., 2008).

## Method

#### **Participants**

Additional groups of participants were recruited from Amazon Mechanical Turk for Experiments 2a and 2b. Quality assurance and exclusion criteria were identical to those for Experiment 1. Two participants were excluded from Experiment 2a and one from Experiment 2b due to having more than 50 trials failing to load because of internet issues. After these exclusions, 71 participants (52 males, 19 females) aged between 18 and 40 years took part in Experiment 2a, and 63 participants (33 males, 30 females) aged between 18 and 40 years took part in Experiment 2b. See Supplemental Tables S3 and S4 for additional demographic details.

## Stimuli

In Experiment 2a, the Full Image stimuli used in Experiment 1 had the background context removed such that only the faces that were in the image were visible (see Figure 4, top panel)-we named these stimuli Faces without Context. In Experiment 2b, faces from the Individual Face Condition were randomly drawn and placed in the positions held by faces in the Experiment 2a stimuli (see Figure 4, bottom panel). This was to examine the effect of disrupting the dependencies between faces in natural images. We called the Experiment 2b stimuli Scrambled Faces.

## **Experimental Procedure**

Participants were asked to judge the average valence, arousal, and dominance of the group of faces for the Faces without Context stimuli in Experiment 2a and the Scrambled Faces stimuli in Experiment 2b. Rating scales were as described in Experiment 1. Participants were given unlimited time to make their ratings, and all participants were given instructions on each dimension and examples together with four practice trials (as for Experiment 1).

## Data Analysis

Descriptive statistics of the average raw scores across participants for each rating condition are reported in Supplemental Table S3. In Experiment 2a, ratings for all the Faces Without Context stimuli on a given dimension of affect were z-scored within participants. Following this, ratings for each stimulus for a given dimension of affect were averaged across participants. In Experiment 2b, a similar procedure was followed for the Scrambled Faces stimuli. For Experiment 2a, drop column feature importance analyses were performed with scores for the Faces Without Context stimuli on each dimension of affect as the dependent variable. MAX and MEAN facial emotion scores for each image were entered as predictor variables. These scores were calculated using the Individual Face Condition ratings from Experiment 1. For Experiment 2a, stimulus MAX and MEAN facial emotion scores were identical to those in Experiment 1.

For Experiment 2b, drop column feature importance analyses were similarly performed with scores for the Scrambled Faces stimuli on each dimension of affect as the dependent variable and MAX and MEAN facial emotion scores for each image entered as predictor variables. These scores were calculated using the Individual Face Condition ratings from Experiment 1. Because the faces within each image for Experiment 2b were no longer the same as in Experiment 1 (as a result of the scrambling procedure), these scores had to be re-estimated.

#### Figure 4

Example Stimuli for Experiment 2



Note. The top panel illustrates Experiment 2a stimuli: Faces without Context and the bottom panel illustrates corresponding Experiment 2b stimuli: Scrambled Faces. In the Scrambled Faces stimuli, individual faces were randomly drawn from the stimulus set and were used to replace the original faces in each image while keeping the relative positioning of the faces intact. For illustration purposes, where the individual faces actually used did not have suitable licenses for publication, they have been replaced with other face images from our set which are licensed for use in educational text as described under Alamy editorial licensing policy (https://www.alamy.com/). See the online article for the color version of this figure.

# Results

In Experiment 2a, the MEAN facial emotion regressor explained unique variance in valence and dominance ratings for the Faces Without Context stimuli (valence: unique  $R_{adj}^2 = .11$ , p < .001; dominance: unique  $R_{adj}^2 = .07, p < .001$ , Figure 5). The MAX facial emotion regressor did not make a significant unique contribution to explained variance in these ratings (valence: unique  $R_{adj}^2 < .01$ , p = .217; dominance: unique  $R_{adj}^2 = .01$ , p = .079). In contrast, for arousal ratings, both the MAX facial emotion regressor and the MEAN emotion regressor explained significant unique variance in ratings of the *Faces Without Context* stimuli (MAX: unique  $R_{adj}^2 = .10$ , p < .001; MEAN: unique  $R_{adj}^2 = .06$ , p < .001, Figure 5). These findings are in line with our hypotheses and contrast with the results for Experiment 1 where mean facial emotion ratings did not significantly contribute to arousal or dominance ratings for the full image and a larger proportion of variance was explained by the maximally arousing and maximally dominant faces. Despite this increased reliance on mean emotion, the maximally arousing face in each image continued to significantly influence arousal ratings for groups of natural faces with the background context removed.

In Experiment 2b, in addition to removing the background, we randomly replaced the faces in each image to eliminate any cues from one face to other faces (i.e., perhaps less arousing faces are turned toward the maximally arousing face) and to disrupt other possible spatial relationships (i.e., perhaps the maximally arousing face tends to be in the center). Drop column feature importance analyses revealed that, as in Experiments 1 and 2a, the MEAN facial emotion regressor explained unique variance in valence ratings for the *Scrambled Faces* stimuli ( $R_{adj}^2 = .26$ , p < .001, Figure 6). In contrast, both the MEAN facial emotion regressor explained unique variance in arousal ratings (MEAN: unique  $R_{adj}^2 = .07$ , p < .001; MAX: unique  $R_{adj}^2 = .06$ , p < .001; MAX: unique  $R_{adj}^2 = .08$ , p < .001. Together, these findings

Drop Column Feature Importance Analysis for Experiment 2a

suggest that when background context is removed, participants show greater reliance on mean emotion when judging the emotion of a group of faces. This is also observed when the relationship between faces in a given image is disrupted. However, in both of these cases, the maximally arousing face continues to significantly influence judgments of group arousal. The results are more equivocal for dominance, with a significance effect of the most dominant face on group dominance ratings in Experiment 2b (background context removed and faces from the whole set randomly allocated to each face position in a given image), but not Experiment 2a (background context removed, original faces retained in original positions).

# **Control Analyses**

As in Experiment 1, we conducted a control analysis to ensure that face size was not influencing our observed findings. We used the ratings on each dimension of affect for the largest face (MAX size) in each image to predict judgments of group emotion for both the Faces without Context stimuli (Experiment 2a) and the Scrambled Faces stimuli (Experiment 2b). The residuals were then used as dependent variables in an additional set of drop column analyses (see Supplemental Figure S6). The maximally arousing face continued to contribute significantly to the prediction of group arousal ratings for both the Faces without Context stimuli (Experiment 2a) and the Scrambled Faces stimuli (Experiment 2b) when the variance explained by the arousal ratings of the largest face were residualized out of these group ratings ( $R_{adj}^2 = .04$ , p = .005;  $R_{adj}^2 = .03$ , p = .015, respectively). The maximally dominant face also contributed significantly to the prediction of group dominance ratings for both the Faces without Context stimuli (Experiment 2a) and the Scrambled Faces stimuli (Experiment 2b) when the variance explained by the dominance ratings of the largest face were residualized out of the group ratings. For the Faces without Context stimuli (Experiment 2a), removal of variance explained by the largest face led to the contribution



## **Experiment 2a Feature Importance Analysis**

*Note.* A drop column feature importance analysis was performed for each dimension of affect using max facial emotion (MAX) and mean facial emotion (MEAN) for each image as features and ratings of crowd affect for the *Faces without Context* stimuli as the dependent variable. Bars represent the unique variance explained by each regressor or "feature"; this is calculated as the difference in adjusted  $R^2$  for the model with both features versus that for the model with the feature in question removed. Scattered dots overlaying each bar show the permuted null distribution of the unique adjusted  $R^2$  values. Mean facial emotion explained unique variance in *Faces without Context* ratings for all three dimensions (left panel: valence, center panel: arousal, right panel: dominance). The maximally arousing face also explained unique variance in *Faces without Context* arousal ratings. See the online article for the color version of this figure.

 $^{***}p < .001.$ 

Figure 5





**Experiment 2b Feature Importance Analysis** 

*Note.* A drop column feature importance analysis was performed for each dimension of affect using max facial emotion (MAX) and mean facial emotion (MEAN) for each image as features and ratings of crowd affect for the *Scrambled Faces* stimuli as the dependent variable. Bars represent the unique variance explained by each regressor or "feature"; this is calculated as the difference in adjusted  $R^2$  for the model with both features versus that for the model with the feature in question removed. Scattered dots overlaying each bar show the permuted null distribution of the unique adjusted  $R^2$  values. Mean facial emotion explained unique variance in *Scrambled Faces* ratings for all three dimensions (left panel: valence, center panel: arousal, right panel: dominance). The maximally arousing and the maximally dominant face also explained unique variance in the *Scrambled Faces* arousal and dominance ratings, respectively. See the online article for the color version of this figure.

of the maximally dominant face becoming more apparent ( $R_{adj}^2 = .02$ , p = .026). For the *Scrambled Faces* stimuli (Experiment 2b), the maximally dominant face also explained a significant amount of unique variance in group dominance ratings after the variance explained by the dominance ratings of largest face was regressed out ( $R_{adj}^2 = .08$ , p < .001).

As in Experiment 1, we also conducted a supplementary drop column analysis replacing the valence MAX with the valence MIN regressor. As was observed for Experiment 1, this confirmed that entering the MIN (most negative) face did not account for a significant amount of unique variance in the valence regressions (Supplemental Figure S7).

We also conducted a further control analysis to ensure that differences in findings between Experiments 2a and 2b were not driven by differences in the range of affect shown by faces in a given image across the two experiments. For this, we randomly sampled 120 images from the set used in Experiment 2a (*Faces without Context*) and from the set used in Experiment 2b (*Scrambled Faces*) so that the mean standard deviation for ratings on each dimension of affect across faces in each image did not differ significantly between sets (valence: t = 0.29, p = .772; arousal: t = -0.19, p = .849; dominance: t = 0.08, p = .940). We then reconducted the drop column feature importance analyses. The results obtained closely replicated those from the main drop column feature importance analyses for Experiments 2a and 2b, see Supplemental Figure S8.

In Experiments 1 and 2, participants were asked to rate the full image stimuli on each of the three dimensions of affect in the following order: valence, arousal, and then dominance. As a final control, we conducted a supplementary experiment in which separate groups of participants rated the full image stimuli on a single dimension of affect. This avoids any possibility of rating order effects. The methods and results are presented in the online Supplements: see Experiment S1, Supplemental Table S6 and Supplemental Figure S9. The findings

from this supplementary experiment largely paralleled those reported in Experiment 1. Context explained significant unique variance in full image ratings across all three dimensions of affect. As in Experiment 1, mean facial emotion, but not maximal facial emotion, explained significant variance in full image valence ratings while the reverse held for arousal. For dominance, in contrast to Experiment 1, but consistent with Experiment 2a, the mean rather than the maximally dominant face contributed significantly to variance in full image ratings.

#### Discussion

Results from our first experiment indicated that when participants viewed natural static photographs of crowds, the perceived emotion of the crowd was predicted significantly both by the facial expressions of the individuals in the scene and by the background context. The extent to which a MEAN or MAX rule was used for judgments of crowd emotion appeared to depend upon the specific emotional dimension being judged. In the case of valence, mean valence (as calculated across all faces in the image) explained unique variance in the ratings of crowd facial valence alongside background context. In contrast, for both arousal and dominance, the most arousing and most dominant face explained unique variance in ratings of crowd emotion, whereas mean arousal and mean dominance regressors did not have a unique effect. These effects survived controlling for the affective ratings of the largest face in each image, suggesting that it was not simply that photographers have focused on the most arousing or dominant face in each group.

In our second experiment, we conducted additional manipulations, both removing background context (Experiment 2a) and additionally removing the faces from a given image and replacing them with faces randomly drawn from other images (Experiment 2b). The latter manipulation ensured that the maximally emotional face, for any given dimension of affect, would not be consistently in a given position across images. In addition, it disrupted the relationship between faces in a given image, for example, ensuring that there was no information from other faces such as gaze or head orientation that could cue the maximally emotional face. Under both these manipulations, group valence ratings continued to be predicted by the mean of the valence of faces within each image. For arousal and dominance ratings, a shift toward partial reliance on both MEAN and MAX facial emotion occurred. For arousal ratings, both the mean arousal of faces in a given image and the maximally arousing face in a given image predicted ratings of group arousal. This was observed regardless of whether ratings of the largest face in the image were regressed out before performing the drop column analysis or not. In the case of dominance, a stronger shift toward reliance on mean dominance was observed, with the maximally dominant face explaining unique variance in group ratings for Experiment 2b (background context removed and face relationships disrupted) but only in Experiment 2a (background context removed) after controlling for variance explained by the dominance ratings of the largest face in the image.

The findings from Experiment 1 suggest that across affective dimensions (valence, arousal, dominance), background context explains unique variance in ratings of crowd emotion when considered alongside MEAN and MAX facial emotion. Meanwhile, the findings from Experiments 1 and 2 taken together suggest that the extent to which participants use the average affect shown by individuals in a group to inform their judgments of group affect depends both on the dimension of affect under consideration and upon the presence of background context.

Ratings of crowd emotional valence (highly negative to highly positive) were strongly influenced by the average valence of faces in the image, and there was no unique contribution of the most positive (MAX) face, or indeed of the most negative (MIN) face. This held across Experiment 1, Experiment 2a, and Experiment 2b. This finding falls in line with prior summary statistic studies, which have shown that mean emotion is a strong predictor of implicit or explicit evaluation of group emotion for valence (Han et al., 2021) and simple basic emotions such as "happy" or "sad" (Haberman & Whitney, 2007, 2009).

In contrast, across Experiments 1, 2a, and 2b, we observed a significant unique influence of the maximally arousing face upon judgments of arousal for groups of faces. This held for natural full images (Experiment 1), faces taken from these images with the background context removed (Experiment 2a), and similarly positioned groups of faces comprising faces randomly selected from across our image set (Experiment 2b). This unique influence of the maximally arousing face was greatest in Experiment 1 where no unique influence of mean arousal across faces was observed, and less, though still significant, in Experiments 2a and 2b. In these latter experiments, mean arousal also contributed to predictions of ratings of group facial emotion.

The attentional literature may potentially help explain this difference in findings for arousal versus valence. Specifically, findings from the attention literature suggest that highly arousing stimuli are preferentially processed leading to earlier breakthrough in binocular rivalry experiments, automatic orienting, slowed disengagement in dot-probe and anti-saccade studies, and increased pop-out in visual search paradigms (Bradley et al., 2012; Lundqvist et al., 2015; Ossenfort & Isaacowitz, 2021; Sheth & Pham, 2008; Vogt et al., 2008). This had led to the theory that emotionally arousing stimuli exert stronger bottom-up salience effects upon attentional competition (Mather & Sutherland, 2011). There is also some similar, though less extensive, evidence that this might also hold for stimuli indicative of dominance (Foulsham et al., 2010; Ohlsen et al., 2013; Ratcliff et al., 2011). In contrast, it has been argued that stimulus valence has a far lesser effect than stimulus arousal upon attentional competition (Mather & Sutherland, 2011) with findings from studies examining the relative effects of stimulus valence and arousal on visual search and attentional disengagement supporting this contention (Lundqvist et al., 2015; Ossenfort & Isaacowitz, 2021; Vogt et al., 2008). Heightened selective attention to the most arousing face in each image could result in enhanced perceptual processing and increased influence of this face on ratings of group arousal, potentially explaining our findings of a significant unique influence of the maximally arousing face upon judgments of arousal across all three experiments.

Our findings also suggest a potentially heightened reliance on the maximally arousing face and reduced reliance on mean arousal when background context is present (Experiment 1), relative to when it is absent (Experiments 2a and 2b). While we need to be cautious in drawing conclusions across experiments, one possible explanation is that, in Experiment 1, background contextual cues as to which part of the image was most salient further focused selective attention on the most arousing face. In line with this, many studies have found that when there is a plethora of visual information, background context can cue participants into attending to the most salient part of the scene (Brockmole & Henderson, 2006; Chun, 2000). There is also evidence that background context's orienting role is especially apparent in high arousal scenes (Bradley et al., 2012; Calbi et al., 2017).

Findings from Experiment 1 also indicated that background contextual information exerted a strong influence on judgments of crowd dominance. Interestingly, facial dominance was the dimension where reliance on MAX versus MEAN emotion was most variable across experiments. In Experiments 1 and 2b, the maximally dominant face contributed to ratings of group emotion, whereas in Experiments 2a and S1, only mean facial dominance had a significant effect. Arguably, dominance might be thought of as more of a relative judgment than arousal or valence, where someone is more or less dominant than someone else. In addition, dominance might, to some extent, be contextually bound (dominance in a board room setting might be predicted by very different cues than in an athletics setting). This could potentially explain the strong influence of background context on judgments of group dominance in Experiment 1. The difference in findings between Experiments 1 and S1 for the dimension of dominance alone, meanwhile, suggest that requiring participants to initially rate other aspects of group facial emotion may also impact dominance ratings. Alternatively, ratings of group facial dominance might simply be less consistent than ratings of other dimensions of group facial emotion; future work will hopefully enable us to further distinguish between these possibilities.

In natural images, there are many different nonface sources of information that might impact the processing of crowd facial emotion. In our present study, background context was defined as everything in the full image except for the cropped-out faces. Hence, in some cases this included the bodies of the individuals whose faces were removed. Work with artificially manipulated stimuli suggests that body pose and gesture can influence perception of the emotion of single faces (Aviezer et al., 2011), and also that background context influences perception of full body, that is, body and face, emotion (Kret et al., 2013.) There is less work on the use of summary statistics to integrate information from multiple bodies (Sweeny et al., 2013). We cannot rule out the possibility that the presence versus absence of bodies might contribute to the impact of context on image ratings and the observed change in reliance on facial maximal emotion versus facial mean emotion. In future work, this issue could be further explored empirically by the selection of a balanced set of images split evenly into ones only including individuals' faces, and ones including individuals' faces and bodies, and use of experimental conditions in which bodies as well as faces are removed, but other aspects of background context are retained.

Future work incorporating eye-tracking would enable us to determine if background context-including or excluding bodies-alone can orient participants to the position of the most arousing face and if, in the absence of background context, participants also preferentially orient to the most arousing face, regardless of its position in the image. Here, it would also be of interest to vary viewing duration. In the experimental work reported here, each image was presented for 1 s. This is sufficient for more than one fixation. We used this duration of presentation to match that in some of the prior work on ensemble perception (Haberman & Whitney, 2011; Im et al., 2017) and out of a desire to simulate comfortable natural viewing. We did not control fixation, given that the variability of the position of faces in natural images does not allow for easy selection of a single fixation point and out of concern for ecological validity. At shorter presentation times, participants might rely to a greater extent on summary statistics. However, it is also possible that early orienting to maximally arousing stimuli, and potentially maximally dominant stimuli, would exert an even stronger influence. This might also vary depending on the presence or absence of background context.

Repeating the work conducted here with controlled fixation and shorter presentation times might also valuably add to the literature on the role of foveal versus peripheral processing in summary statistics, in general, and judgments of group facial affect, in specific. Prior findings have indicated that ensemble processing is not dependent on foveal information (e.g., Wolfe et al., 2015). In addition, studies of facial emotion from outside of the ensemble literature have shown that overt attentional allocation is not necessary for processing highly arousing facial stimuli (e.g., Phelps et al., 2006). However, recent work within the ensemble processing field has demonstrated that when foveal information is present, it can disproportionately influence judgments (Jung et al., 2017; Tiurina et al., 2023). In the context of facial affect, this work has been conducted with artificially generated or posed emotional faces presented against a grey background (Dandan et al., 2023; Ueda, 2022), hence it would be of value to extend this work using natural images of emotional crowds.

Our findings may be of pertinence not only for the field of psychology, but also for the computer vision and machine learning literature. Machine learning and computer vision are being integrated into our everyday lives through social media and security-oriented face recognition systems. Many models have only been trained to categorize either singleton faces or groups of faces in isolation and not in background context. Those models that categorize groups of faces in context (Dhall et al., 2016, 2017; Kosti et al., 2020; Mou et al., 2015; Rassadin et al., 2017) fail to simulate human ratings of crowd emotion and perform worse than models that categorize faces in isolation. The findings reported here may indicate how models can be refined to better mimic human perception of crowd emotion in uncontrolled, natural conditions.

In conclusion, our results suggest that in natural scenes, background context both directly and indirectly influences judgments of crowd emotion. It has been well noted that context contributes to the perception of single facial expressions (Aviezer et al., 2017; Chen & Whitney, 2019, 2022; Righart & de Gelder, 2008a, 2008b; Sasson et al., 2016) and that context directs attention to salient parts to the scene (Bradley et al., 2012; Brockmole & Henderson, 2006; Calbi et al., 2017; Chun, 2000). Here, we show that, when judging a group of faces, background context significantly contributes to the perception of group facial emotion across all three dimensions of affect considered, namely valence, arousal, and dominance. Further, the presence of background context also increases the influence of the maximally arousing face upon ratings of group facial arousal. In the absence of background context, there is more evident reliance on summary statistics across all dimensions of facial affect. However, it is of note that the maximally arousing face continues to significantly influence judgments of crowd arousal under all conditions, both with and without background context, and when the relationship between faces in an image is disrupted (Experiment 2b). This is in line with emotional arousal influencing the perception of group facial affect by biasing of selective attention and subsequent perceptual processing in favor of the maximally arousing face.

## References

- Alwis, Y., & Haberman, J. M. (2020). Emotional judgments of scenes are influenced by unintentional averaging. *Cognitive Research: Principles and Implications*, 5(1), Article 28. https://doi.org/10.1186/s41235-020-00228-3
- Aviezer, H., Bentin, S., Dudarev, V., & Hassin, R. R. (2011). The automaticity of emotional face-context integration. *Emotion*, 11(6), 1406–1414. https:// doi.org/10.1037/a0023578
- Aviezer, H., Ensenberg, N., & Hassin, R. R. (2017). The inherently contextualized nature of facial emotion perception. *Current Opinion in Psychology*, 17, 47–54. https://doi.org/10.1016/j.copsyc.2017.06.006
- Barrett, L. F., & Kensinger, E. A. (2010). Context is routinely encoded during emotion perception. *Psychological Science*, 21(4), 595–599. https://doi.org/ 10.1177/0956797610363547
- Bombari, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., & Lobmaier, J. S. (2013). Emotion recognition: The role of featural and configural face information. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 66(12), 2426–2442. https://doi.org/10 .1080/17470218.2013.789065
- Bradley, M. M., Keil, A., & Lang, P. J. (2012). Orienting and emotional perception: Facilitation, attenuation, and interference. *Frontiers in Psychology*, 3, Article 493. https://doi.org/10.3389/fpsyg.2012.00493
- Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, 13(1), 99–108. https://doi.org/ 10.1080/13506280500165188
- Calbi, M., Heimann, K., Barratt, D., Siri, F., Umiltà, M. A., & Gallese, V. (2017). How context influences our perception of emotional faces: A behavioral study on the Kuleshov effect. *Frontiers in Psychology*, 8, Article 1684. https://doi.org/10.3389/fpsyg.2017.01684
- Chen, Z., & Whitney, D. (2019). Tracking the affective state of unseen persons. Proceedings of the National Academy of Sciences of the United States of America, 116(15), 7559–7564. https://doi.org/10.1073/pnas.1812250116
- Chen, Z., & Whitney, D. (2022). Inferential emotion tracking (IET) reveals the critical role of context in emotion recognition. *Emotion*, 22(6), 1185–1192. https://doi.org/10.1037/emo0000934
- Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4(5), 170–178. https://doi.org/10.1016/S1364-6613(00)01476-5
- Dandan, Y. R., Ji, L., Song, Y., & Sayim, B. (2023). Foveal vision determines the perceived emotion of face ensembles. *Attention, Perception*

& Psychophysics, 85(1), 209–221. https://doi.org/10.3758/s13414-022-02614-z

- Dhall, A., Goecke, R., Ghosh, S., Joshi, J., Hoey, J., & Gedeon, T. (2017). From individual to group-level emotion recognition: EmotiW 5.0 [Conference session]. Proceedings of the 19th ACM International Conference on Multimodal Interaction, New York, NY, United States. https://doi.org/10.1145/3136755.3143004
- Dhall, A., Goecke, R., Joshi, J., Hoey, J., & Gedeon, T. (2016). *EmotiW 2016: Video and group-level emotion recognition challenges* [Conference session]. Proceedings of the 18th ACM International Conference on Multimodal Interaction, New York, NY, United States. https://doi.org/10.1145/2993148 .2997638
- Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, 117(3), 319–331. https://doi.org/10.1016/j.cognition .2010.09.003
- Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision*, 9(11), Article 1. https://doi.org/10.1167/9.11.1
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, 17(17), R751–R753. https:// doi.org/10.1016/j.cub.2007.06.039
- Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 718–734. https://doi.org/10.1037/a0013899
- Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review*, 18(5), 855–859. https://doi.org/10.3758/s13423-011-0125-6
- Han, L., Yamanashi Leib, A., Chen, Z., & Whitney, D. (2021). Holistic ensemble perception. Attention, Perception & Psychophysics, 83(3), 998–1013. https://doi.org/10.3758/s13414-020-02173-1
- Hareli, S., Shomrat, N., & Hess, U. (2009). Emotional versus neutral expressions and perceptions of social dominance and submissiveness. *Emotion*, 9(3), 378–384. https://doi.org/10.1037/a0015958
- Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition*, 106(2), 1017–1027. https://doi.org/10.1016/j.cognition.2007 .04.002
- Im, H. Y., Chong, S. C., Sun, J., Steiner, T. G., Albohn, D. N., Adams, R. B., Jr., & Kveraga, K. (2017). Cross-cultural and hemispheric laterality effects on the ensemble coding of emotion in facial crowds. *Culture and Brain*, 5(2), 125–152. https://doi.org/10.1007/s40167-017-0054-y
- Jung, W., Bülthoff, I., & Armann, R. G. M. (2017). The contribution of foveal and peripheral visual information to ensemble representation of face race. *Journal of Vision*, 17(13), Article 11. https://doi.org/10.1167/17 .13.11
- Karbasi, V., Tehrani-Doost, M., & Ghassemi, F. (2018). Investigating the effect of image blurring on facial emotion recognition. Advances in Cognitive Science, 20(3), 1–14.
- Kosti, R., Alvarez, J. M., Recasens, A., & Lapedriza, A. (2020). Context based emotion recognition using EMOTIC dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11), 2755–2766. https:// doi.org/10.1109/TPAMI.2019.2916866
- Kret, M. E., Roelofs, K., Stekelenburg, J. J., & de Gelder, B. (2013). Emotional signals from faces, bodies and scenes influence observers' face expressions, fixations and pupil-size. *Frontiers in Human Neuroscience*, 7, Article 810. https://doi.org/10.3389/fnhum.2013.00810
- Leib, A. Y., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, 7(1), Article 13186. https://doi.org/10.1038/ncomms13186
- Li, H., Ji, L., Tong, K., Ren, N., Chen, W., Liu, C. H., & Fu, X. (2016). Processing of individual items during ensemble coding of facial expressions. *Frontiers in Psychology*, 7, Article 1332. https://doi.org/10.3389/fpsyg .2016.01332

- Lundqvist, D., Bruce, N., & Öhman, A. (2015). Finding an emotional face in a crowd: Emotional and perceptual stimulus factors influence visual search efficiency. *Cognition and Emotion*, 29(4), 621–633. https://doi.org/10 .1080/02699931.2014.927352
- Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science: A Journal* of the Association for Psychological Science, 6(2), 114–133. https://doi.org/ 10.1177/1745691611400234
- Mou, W., Celiktutan, O., & Gunes, H. (2015). Group-level arousal and valence recognition in static images: Face, body and context [Conference session]. 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG), Ljubljana, Slovenia. https:// doi.org/10.1109/FG.2015.7284862
- Ohlsen, G., van Zoest, W., & van Vugt, M. (2013). Gender and facial dominance in gaze cuing: Emotional context matters in the eyes that we follow. *PLOS ONE*, 8(4), Article e59471. https://doi.org/10.1371/journal .pone.0059471
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. Proceedings of the National Academy of Sciences of the United States of America, 105(32), 11087–11092. https://doi.org/10.1073/pnas .0805664105
- Ossenfort, K. L., & Isaacowitz, D. M. (2021). Spatial attention to arousing emotional stimuli in younger and older adults. *Motivation and Emotion*, 45(6), 790–797. https://doi.org/10.1007/s11031-021-09899-x
- Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychological Science*, 17(4), 292–299. https://doi.org/10.1111/j.1467-9280.2006.01701.x
- Phillips, L. T., Slepian, M. L., & Hughes, B. L. (2018). Perceiving groups: The people perception of diversity and hierarchy. *Journal of Personality* and Social Psychology, 114(5), 766–785. https://doi.org/10.1037/pspi 0000120
- Pinkham, A. E., Griffin, M., Baron, R., Sasson, N. J., & Gur, R. C. (2010). The face in the crowd effect: Anger superiority when using real faces and multiple identities. *Emotion*, 10(1), 141–146. https://doi.org/10.1037/ a0017387
- Rassadin, A., Gruzdev, A., & Savchenko, A. (2017). Group-level emotion recognition using transfer learning from face identification [Conference session]. Proceedings of the 19th ACM International Conference on Multimodal Interaction, New York, NY, United States. https://doi.org/10 .1145/3136755.3143007
- Ratcliff, N. J., Hugenberg, K., Shriver, E. R., & Bernstein, M. J. (2011). The allure of status: High-status targets are privileged in face processing and memory. *Personality and Social Psychology Bulletin*, 37(8), 1003–1015. https://doi.org/10.1177/0146167211407210
- Righart, R., & de Gelder, B. (2008a). Rapid influence of emotional scenes on encoding of facial expressions: An ERP study. *Social Cognitive and Affective Neuroscience*, 3(3), 270–278. https://doi.org/10.1093/scan/nsn021
- Righart, R., & de Gelder, B. (2008b). Recognition of facial expressions is influenced by emotional scene gist. *Cognitive, Affective & Behavioral Neuroscience*, 8(3), 264–272. https://doi.org/10.3758/CABN.8.3.264
- Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178. https://doi.org/10.1037/ h0077714
- Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3), 273–294. https:// doi.org/10.1016/0092-6566(77)90037-X
- Sasson, N. J., Pinkham, A. E., Weittenhiller, L. P., Faso, D. J., & Simpson, C. (2016). Context effects on facial affect recognition in schizophrenia and autism: Behavioral and eye-tracking evidence. *Schizophrenia Bulletin*, 42(3), 675–683. https://doi.org/10.1093/schbul/sbv176
- Sheth, B. R., & Pham, T. (2008). How emotional arousal and valence influence access to awareness. *Vision Research*, 48(23–24), 2415–2424. https://doi.org/10.1016/j.visres.2008.07.013

- Sun, J., & Chong, S. C. (2020). Power of averaging: Noise reduction by ensemble coding of multiple faces. *Journal of Experimental Psychology: General*, 149(3), 550–563. https://doi.org/10.1037/xge0000667
- Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, 127(1), 105–118. https://doi.org/10.1016/j.cognition.2012.12.001
- Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology. Human Perception and Performance*, 39(2), 329–337. https://doi.org/10.1037/a0028712
- Tiurina, N., Markov, Y., Whitney, D., & Pascucci, D. (2023). The functional role of spatial anisotropies in ensemble perception. https://doi.org/10 .31234/osf.io/yd3n8
- Ueda, Y. (2022). Understanding mood of the crowd with facial expressions: Majority judgment for evaluation of statistical summary perception. *Attention, Perception & Psychophysics*, 84(3), 843–860. https://doi.org/10 .3758/s13414-022-02449-8
- Vogt, J., De Houwer, J., Koster, E. H. W., Van Damme, S., & Crombez, G. (2008). Allocation of spatial attention to emotional stimuli depends upon

arousal and not valence. *Emotion*, 8(6), 880-885. https://doi.org/10.1037/a0013981

- Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. Annual Review of Psychology, 69(1), 105–129. https://doi.org/10.1146/annurevpsych-010416-044232
- Wolfe, B. A., Kosovicheva, A. A., Leib, A. Y., Wood, K., & Whitney, D. (2015). Foveal input is not required for perception of crowd facial expression. *Journal of Vision*, 15(4), Article 11. https://doi.org/10.1167/ 15.4.11
- Yamanashi Leib, A., Chang, K., Xia, Y., Peng, A., & Whitney, D. (2020). Fleeting impressions of economic value via summary statistical representations. *Journal of Experimental Psychology: General*, 149(10), 1811–1822. https://doi.org/10.1037/xge0000745

Received September 12, 2022

Revision received December 5, 2023

Accepted January 22, 2024