



OPEN The effect of masks on the emotion perception of a facial crowd

Jieun Cho¹, Hee Yeon Im², Young Jun Yoon³, Sung Jun Joo³ & Sang Chul Chong^{1,4}✉

The present study investigated the effect of facial masks on people's ability to perceive emotions in crowds. We presented faces with the bottom halves occluded by masks or full faces without occlusion. In two sequentially presented crowds, we varied the number of faces, emotional valence, and intensity of facial expressions, examining the impact of masks on the perception of crowd emotion. Participants reported which of the two crowds they would avoid based on the crowds' average emotions. The participants' ability to judge the average emotion of a crowd, especially a crowd expressing happiness, was impaired when the crowd wore masks. For faces covered by masks, crowd emotion judgments were more negatively biased than those without masks. However, participants could still distinguish the emotional intensities of a crowd wearing masks above chance. Additionally, participants responded more quickly to a crowd with more people without compromising accuracy, despite the perceptual challenges imposed by facial masks. Our results suggest that under ambiguous social situations in which individuals' emotions are partially hidden by masks, a large group may provide stronger social cues than a small group, thereby promoting communication and regulating social behaviors.

Due to the COVID-19 pandemic, mask-wearing has become a requirement worldwide. As a result, many social interactions have been allowed only when people wear masks. However, wearing masks conceals the visual information conveyed by the lower half of a face, especially the region containing the mouth and nose. Like COVID-19 itself, mask-wearing will likely not be a permanent reality; however, it may not be eliminated since people have learned this new habit as a potential response to future outbreaks (<https://www.nature.com/articles/d41586-021-01394-0>). Although the impacts of mask-wearing on public health are under extensive investigation as an effective nonpharmacologic intervention to reduce the spread of infection, the potential impacts of continued mask-wearing on social cognition and behaviors related to everyday activities have not been evaluated.

Recent studies have reported that perceiving faces obstructed by masks hampered adults' recognition of facial identity¹ and emotion^{2,3}. Similar outcomes were found in school-aged children⁴. Specifically, Freud et al.¹ suggested that perceiving faces with masks that cover the lower half of a face engaged a local, feature-based processing and significantly reduced holistic processing, which is considered a hallmark of face perception⁵⁻⁷. Regarding emotion perception, both adults and children showed reduced accuracy, although they could still categorize the emotion expressed by a masked face better than chance²⁻⁴. In these previous studies, observers identified the emotion of a face wearing a mask among discrete emotional labels (e.g., happy, angry, neutral) by engaging in categorical emotion perception⁸.

Labeling an emotion with a specific category modulates how people perceive facial expressions⁹⁻¹². For example, two faces expressing emotions from distinct categories (e.g., happiness and sadness) tended to be distinguished more easily than two faces expressing emotions within the same category, even when the emotional intensity and physical differences were matched for both pairs of faces^{9,10,12,13}. When some deterministic facial features of a face wearing a mask become ambiguous or unavailable, the perceived emotion of the face can be strengthened by verbal labels due to the categorical perception. For example, a masked face can be perceived as happier more easily once labeled as such. However, in many day-to-day interactions, labeling a facial emotion using a discrete category (e.g., "This person looks happy, and this person looks angry") is not always practical given the high-dimensional taxonomy¹⁴ and continuous nature (mixed emotions¹⁵) of emotions and emotional experiences.

People in any social setting tend to share a consistent emotional state rather than expressing categorically different emotions, although the intensity and magnitude of the emotion may vary (e.g., joy at a rock concert, sadness at a funeral, fear of terrorism, or anger at a political or social demonstration). The overall mood of a crowd creates a meaningful context, providing a social-emotional signal that rapidly informs observers of

¹Graduate Program in Cognitive Science, Yonsei University, Seoul, South Korea. ²Department of Psychology, University of British Columbia, Vancouver, Canada. ³Department of Psychology, Pusan National University, Busan, South Korea. ⁴Department of Psychology, Yonsei University, Seoul, South Korea. ✉email: scchong@yonsei.ac.kr

appropriate behaviors in a social context^{16–19}. As a member of a protest march or political demonstration, one might perceive anger as a prevalent emotion in the group. Among facial groups expressing various spectrums of anger, one might want to make a relative emotional judgment to ensure that they stay away from a group that looks infuriated to avoid any unpredictable violence. Therefore, extracting a crowd emotion is a useful social skill for comprehending and reacting appropriately to the social context of one's surroundings²⁰. Previous research has investigated whether this ability may stem from our ensemble coding ability. For example, averaging different faces' emotions can be performed rapidly and efficiently without perceiving each face serially^{21–23}. Due to noise cancellation, averaging facial expressions is also more robust against the noise produced by face inversion than processing a facial expression²⁴.

When people see others' faces covered by masks from a physical distance in public spaces, they often misinterpret the partially occluded facial expressions. Recent works on the perception of a mask-wearing face mainly examined close one-to-one interpersonal interactions and may not apply to the perception of a facial crowd. Therefore, it is still unknown how the new normal social stimuli (i.e., multiple upper halves of faces) influence the perceptual ability to extract crowd emotion and interpret the overall social ambiance. The present study investigated how covering the lower halves of faces influences people's perceptual judgments of the emotions of crowds. We compared participants' accuracy and response time (RT) according to three facial mask-wearing group conditions: when both crowds wore masks (*both* condition), when only one crowd wore masks (*either* condition), and when neither crowd wore masks (*neither* condition). We varied the number of faces to examine the effect of crowd size on participants' judgment of a crowd's emotion. More faces in crowds may increase challenges when the mask-wearing faces provide less emotion-related information. However, if partial information from the upper halves of faces can be sufficiently processed simultaneously, crowd size will not affect emotion perception in mask-wearing crowds. Instead, ensemble perception may even enhance emotion perception via noise canceling by averaging more samples, leading to improved perception of the overall emotion of the crowd.

In addition, we investigated how people perceived the overall emotion of a facial crowd wearing masks when the verbal labeling of emotion did not play a sufficient role. We did this by asking participants to choose which of two sequentially presented sets of faces they would like to avoid instead of asking participants to verbally label the emotion. Recent studies of a single face wearing a mask have shown that mask-wearing interfered with recognizing a happy face more strongly than an angry or fearful face²⁵. Therefore, we also manipulated the emotional intensity and valence of the faces to examine whether mask-wearing influences the perception of positive crowd emotions more significantly than negative emotions.

Methods

Participants. A total of 40 people participated in this experiment: 20 undergraduate and graduate students (aged 21–32; 10 women and 10 men) from Yonsei University (Seoul, South Korea) and 20 undergraduate students (aged 18–27; 13 women and seven men) from Pusan National University (Busan, South Korea). We recruited more than twice as many participants as in previous studies using similar stimuli and tasks^{22,24}. Furthermore, we analyzed our data using Bayesian methods, which are less influenced by the number of participants²⁶. All participants had a normal color vision and normal or corrected-to-normal visual acuity. The experimental protocol was approved by the Institutional Review Boards at Yonsei and Pusan National Universities, and written informed consent was obtained according to their procedures. All methods were performed in accordance with the relevant guidelines and regulations by the Institutional Review Boards.

Apparatus and stimuli. All stimuli were generated using the Psychophysics Toolbox Version 3 and its extensions (Brainard, 1997²⁷; Pelli, 1997²⁸) for MATLAB (MathWorks Inc., Natick, MA, USA). At Yonsei University, stimuli were displayed on a CRT monitor at a resolution of 1600 × 1200 with a refresh rate of 85 Hz. At Pusan National University, the same stimuli were displayed on a 27-in. full HD widescreen color monitor at a resolution of 1920 × 1080 with a refresh rate of 120 Hz. At both locations, participants were seated approximately 60 cm from the monitor with their heads on a head-chin rest inside a darkened room.

In each trial, we presented a set of faces, each of which expressed a specific emotion. Faces were morphed via FantaMorph software (Abrosoft), using face stimuli from the Yonsei Face Database²⁹. The database consists of 344 photographs of 17 amateur Korean actors (nine women and eight men) displaying six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) and a neutral facial expression.

The emotional intensity of each photograph in the original study²⁹ was evaluated by a minimum of 177 Yonsei University students (177–212) for verification purposes. The evaluators categorized the emotion of each photograph with one of the seven facial labels and assessed its emotional intensity on a linear scale from 0 (very weak) to 10 (very strong). In the current study, we selected eight actors (three women and five men) and used each actor's happy and angry faces for morphing. We only used the faces of individuals whose mouths were closed for the following reasons. Open-mouthed faces are generally perceived as more intense and extreme, resulting in higher valence and arousal levels³⁰. They also exhibit a greater attentional advantage, especially when expressing anger and happiness^{31–34}. Using faces with closed mouths only also allowed us to rule out the possibility that any difference between crowds with masked and unmasked people is explained solely by the difference in the perceptual saliency of the mouth area of the face.

Our selection criteria for the actors were as follows: (1) the face stimuli had similar intensity ratings between happy and angry; (2) the most frequent emotion judged by the observers (> 70%) was consistent with the emotional category the actor in the image intended to express (e.g., the most frequent “happy” response to a “happy” expression); (3) none of the other emotional labels were rated as high as happy (e.g., to prevent confusion with surprise or disgust) or angry (e.g., to prevent confusion with fear or sadness). Supplementary Table 1 details the

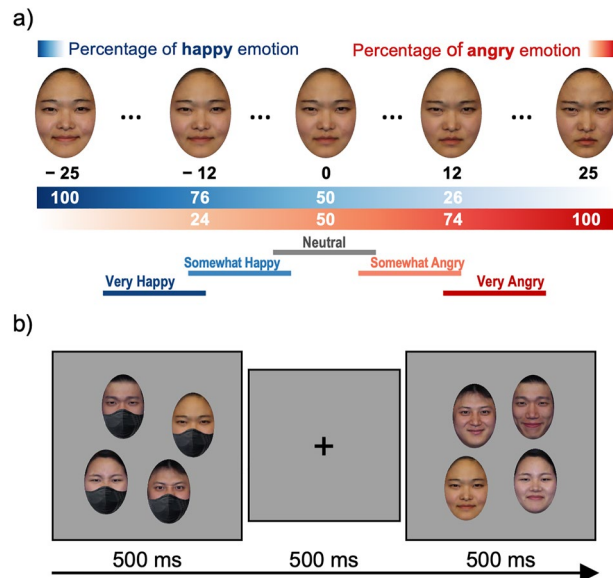


Figure 1. Stimuli and Procedure. *Note* (a) Emotional units produced by morphing happy and angry emotions (black numbers) and the morphing proportion of those emotions (white numbers in the blue and red spectrum bars). The horizontal color bars with labels ranging from Very Happy to Very Angry indicate the different emotional valence and intensity conditions used. (b) An example trial sequence. This example illustrates a trial of the *either* condition when the probe set represents happy faces with a set size of 4. Informed consent was obtained from amateur models for publishing the images.

observers' emotional judgments for the eight face stimuli with both happy and angry emotions, including the emotion judgment frequency and rated emotional intensity (mean and standard deviation).

By morphing the happy and angry face images with the same identity, we created 51 facial emotion morph levels ranging from -25 (the happiest) to $+25$ (the angriest). The morphed face images were linearly interpolated (in 2% increments) between the original happy and angry emotions. The same process was applied to the eight selected identities. The emotion level of 0 corresponded to a neutral face, which was a morph comprising 50% happy and 50% angry faces (Fig. 1a). Different faces were separated by emotional intensity units such that face 1 was one emotional unit happier than face 2. Therefore, the larger the separation between any two morphed faces in emotional units, the easier it was to discriminate them based on their emotional intensity.

Individual face images ($2.27^\circ \times 3.37^\circ$ visual angle in size at Yonsei University and $2.82^\circ \times 4.19^\circ$ visual angle in size at Pusan National University) of a crowd were randomly positioned in an invisible frame ($10^\circ \times 10^\circ$ visual angle at the center of a gray background at Yonsei University and $12.4^\circ \times 12.4^\circ$ visual angle at Pusan National University).

Design and procedure. Our experiment had a 3 set sizes (one, four, and eight) \times 4 emotions (very happy, somewhat happy, somewhat angry, and very angry) \times 3 mask conditions of face sets (neither, either, and both) within-subject design, with 28 repetitions per condition. Therefore, there were 1008 trials in total, and the sequence of the trials was randomized for each participant. Participants took a short break after every 200 trials.

In each trial, two sets of faces were presented sequentially for 500 ms each (see Fig. 1b for a sample trial). Each set contained one, four, or eight different identity faces according to the set size condition. A black fixation cross on a blank screen was presented for 500 ms between the two displays of facial stimuli. One of the two displays always contained a neutral-emotion group with an average value of 0 emotional units as a control set. The probe set either had a mean of -16 (very happy), -8 (somewhat happy), $+8$ (somewhat angry), or $+16$ (very angry) emotion units. In other words, participants consistently compared a group of faces exhibiting one of the four average emotion intensities with a group of faces displaying a neutral emotion (where the average value was 0). Whether the control or probe set was presented first was randomly determined in every trial.

The minimum and maximum emotion intensities within the same emotion condition were the same across all set sizes, spanning a 10-unit range, except for the set size 1 condition where only the average value was chosen. Therefore, the happy conditions (including both somewhat happy and very happy conditions) contained only happy individual faces with varying intensities, while the angry conditions contained only angry individual faces. We also ensured that individual faces had emotion intensities that were distinct, uniformly distributed within the 10-unit range, and symmetrical around the average emotion. For example, if the emotion units used for the very angry condition in set size 4 were^{11,15,17,21}, the corresponding emotion units for the same condition in set size 8 would be^{11,12,14,15,17,18,20,21}, with varying middle values across trials. These manipulations were implemented to minimize the possibility that one or two faces with intense emotions would evoke faster responses due to saliency. For the neutral-emotion stimuli, the extreme values in the range were identical across trials: -5 (happy) and 5 (angry). While maintaining the mean value of 0, the middle units varied across trials.

Importantly, there were three conditions, depending on whether the faces in a set (the probe or control set) wore masks. In the *both* condition, the faces wore masks in both the probe and control sets; in the *either* condition, only one set contained faces wearing masks, while the other set contained faces without masks; and in the *neither* condition, none of the sets contained faces wearing masks. Figure 1b presents a sample trial with the set size of four, when the control set (neutral; 0) was compared with the very happy crowd (−16) and when only the faces in the control set were wearing masks (i.e., the *either* condition). In this figure, the control set containing individual faces with emotional units of [−5, −4, 4, 5] is presented first, followed by a blank screen and the probe set (average = −16) containing individual faces with emotional units of [−21, −19, −13, −11].

Participants were instructed to fixate on the center of the screen to view the two successive sets of multiple faces and make a keypress as accurately and quickly as possible to indicate which of two facial crowds they would rather avoid. We explicitly informed participants that the correct answer was the facial crowd showing a more negative emotion on average. Feedback was provided only during 20 practice trials and then removed for the main experiment. Participants' responses to the first 20 practice trials were not included in the data analysis.

Statistical analysis. We used the *brms* package³⁵ for the Bayesian mixed model analysis and *bayestestR* packages³⁶ to compare models. We fit models to predict the proportion of which facial crowd was judged as negative and the RTs, separately, depending on the set size, emotion, and mask conditions. The set size and the emotion predictors were treated as continuous predictors, while the facial mask predictor was treated as a categorical predictor. The response was binary—that is, 1 indicated a “more negative” response, and 2 indicated a “less negative” response for a probe set (a more expressive crowd) compared to a control set (a neutral crowd on average). Therefore, logistic regression models were fit to predict the proportion of “more negative” responses using the Bernoulli distribution with a logit link function. The RTs were fit to the shifted lognormal distribution with an identity link function as recommended in the literature³⁷.

For each dependent variable, the full model consisted of the three predictors and a participant-level random intercept. The null model contained only the random intercept of participants. The null model was compared with other models comprising all possible combinations of the main effects or interaction terms. The model comparison results were reported as Bayes factors (BF_{10}), which indicate the extent to which the data are better explained by one model than another³⁸. For example, a BF_{10} of 3 indicates that the observed data are three times more probable under the alternative model, H_1 , than the null model, H_0 . As established in the previous literature, we define $1 < BF_{10} < 3$ as anecdotal evidence, $3 < BF_{10} < 10$ as moderate evidence, $10 < BF_{10} < 30$ as strong evidence, and $BF_{10} > 30$ as very strong evidence of an effect^{26,39}.

Our analysis was conducted in two steps. In the first step, we searched for the best model by comparing models with the null model. In the second step, we analyzed the specific effect of each predictor by calculating the inclusion BFs (BF_{incl}) across matched models, representing evidence for all models *with* a term of interest against all models *without* the term. Since we considered matched models only, the models *without* the term included only the main effect terms that constitute the interaction term⁴⁰.

To analyze the proportion of “more negative” responses for a probe set, we used a weakly informative prior, Student's *t* distribution ($v = 4$, $\mu = 0$, $s = 2.5$), for the regression coefficients of the binary response data. The models were fit using eight chains, each with 6000 iterations, including a warm-up of 1000 samples per chain. This resulted in a total of 40,000 Markov chain Monte Carlo samples. For the RT data, we used a normal prior ($Normal(0,10)$), which is also weakly informative. The models were fit using seven chains, each with 10,000 iterations, including a warm-up of 1,000 samples per chain, resulting in 63,000 Markov chain Monte Carlo samples. The convergence of Markov chain Monte Carlo chains was validated by the Rhat statistic.

Additionally, we conducted a post-hoc analysis on the *either* condition only. Trials of the *either* condition were fit to the Bernoulli distribution with a logit link function. Student's *t* distribution ($v = 4$, $\mu = 0$, $s = 2.5$) for the regression coefficients was used as a prior, and the models were fit using eight chains, each with 6000 iterations, including a warm-up of 1000 samples per chain. When constructing models, we used two population-level predictors: emotion category (happy, angry) and set with facial masks (probe, control).

We neither excluded nor trimmed specific data or trials when modeling the proportion of “more negative” response data. For the RT data, however, trials with RTs three standard deviations above or below the mean in each condition were considered outliers and subsequently excluded. The exclusion rate was 2% on average across conditions (1.95% in total).

Results

Figure 2 displays the proportion of “more negative” responses (Fig. 2a) for the probe set compared to the neutral control set and the RT (Fig. 2b). Table 1 shows the results of all the Bayesian models tested for finding the best-fitting model. Tables 2 and 3 show the summarized BFs of the effects of interest (i.e., the inclusion BFs).

As shown in Table 1, the proportion of “more negative” responses for the probe set was best predicted by the model that contained the main effect terms of emotion and facial mask and the interaction term of the two; this is evidenced by the largest BF ($\log BF_{10} = 2848.4$) when compared to the null model. The inclusion BFs shown in Table 2 provide very strong evidence for the main effect of emotion ($BF_{incl} = Inf$), indicating that the proportion of “more negative” responses increased as the intensity of the angry emotion of the probe set increased.

We also found strong evidence for the main effect of the mask condition ($BF_{incl} = 25.61$) and very strong evidence for the interaction between the emotion and mask conditions ($BF_{incl} = 6.19e+105$). The interaction pattern showed that the slopes of the proportion of “more negative” responses (as a function of the intensity of angry emotions in the probe set) varied across the mask conditions (Fig. 2a). The steepest slope was found in the *neither* condition. These results suggest that the perception of a crowd's average emotion was impaired when faces were wearing masks. Nevertheless, when conducting a simple slope analysis with the mask condition serving

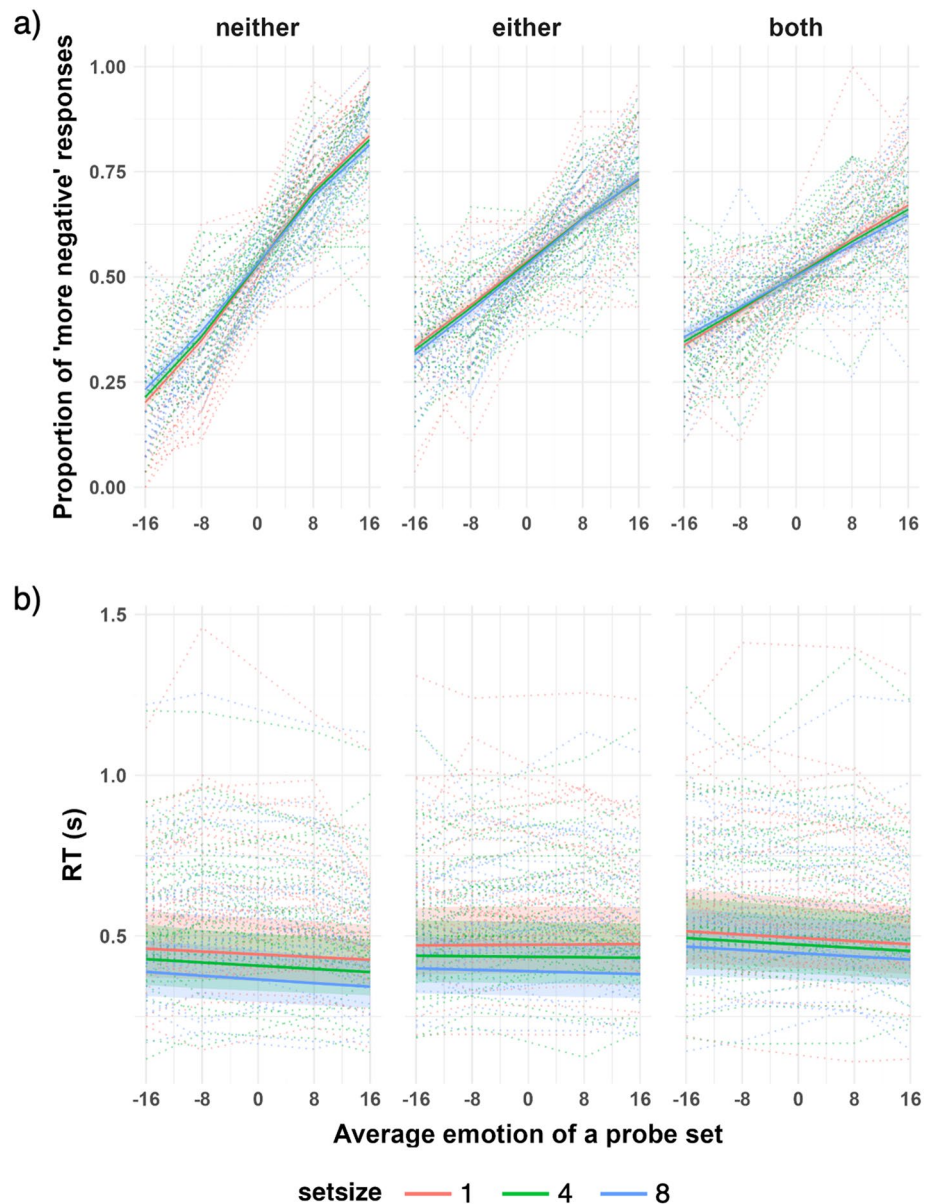


Figure 2. The results of the experiment. *Note.* (a) The predicted proportion of “more negative” responses for a probe set, as a function of the average emotion of a probe set (the left end of the x-axis indicates the happiest crowd, and the right end of the x-axis indicates the angriest crowd), using a full Bayesian logistic mixed model. Each graph shows the results of each facial mask condition, depicting how the mask condition affected response patterns. (b) The predicted RT using a full Bayesian logistic mixed model. Each graph shows the results of each facial mask condition, indicating how the mask condition affected the RT patterns. In each graph, the pink line indicates the set size 1 condition, the green line indicates the set size 4 condition, and the blue line indicates the set size 8 condition. The shaded areas of each line denote the highest and lowest credible intervals for the estimates. The dotted lines in each graph are the raw data of individuals.

as a moderator, the slope estimates at all levels were greater than zero (Table 4). This suggests that masked faces in a crowd weakened the average emotion perception but did not entirely prevent participants from correctly judging the average emotion. Other interaction terms, including a three-way interaction between the set size, emotion, and mask (all $BF_{incl} < 0.33$), did not provide any significant evidence.

In the *either* condition, either the probe set or the control set comprised faces wearing masks. We further analyzed the responses in this condition to investigate how participants compared the average emotion of a crowd wearing facial masks with that of a crowd without masks. We collapsed the data across the three levels of set size and the four levels of emotion into two categories: happy and angry. When the average emotion of the probe set was happy, there was a substantial difference in the proportion of “more negative” responses between the case in which the happy crowd wore masks (to be compared with a neutral crowd not wearing masks) and

	$\log(\text{BF}_{10})$	BF_{10}
Models	P(Negative)	RT
SetSize	-6.58	1.33e+33
Emotion	2601.57	140.97
Mask	1.41	1.19e+28
SetSize + Emotion	2595.06	5.98e+35
SetSize + Mask	-5.15	6.22e+61
Emotion + Mask	2604.81	4.94e+29
SetSize + Emotion + SetSize:Emotion	2587.80	1.61e+31
SetSize + Mask + SetSize:Mask	-16.46	4.49e+57
Emotion + Mask + Emotion:Mask	2848.40	2.28e+23
SetSize + Emotion + Mask	2598.31	2.75e+63
SetSize + Emotion + Mask + SetSize:Emotion	2591.04	7.51e+58
SetSize + Emotion + Mask + SetSize:Mask	2587.17	2.09e+59
SetSize + Emotion + Mask + Emotion:Mask	2841.83	1.32e+57
SetSize + Emotion + Mask + SetSize:Emotion + SetSize:Mask	2579.91	5.77e+54
SetSize + Emotion + Mask + SetSize:Emotion + Emotion:Mask	2834.66	3.44e+52
SetSize + Emotion + Mask + SetSize:Mask + Emotion:Mask	2830.78	9.56e+52
SetSize + Emotion + Mask + SetSize:Emotion + SetSize:Mask + Emotion:Mask	2823.50	2.70e+48
SetSize + Emotion + Mask + SetSize:Emotion + SetSize:Mask + Emotion:Mask + SetSize:Emotion:Mask	2810.49	3.36e + 39

Table 1. Bayesian analysis: model comparison. All models were compared to the null model, which included only Subject as a random factor. All models also included Subject as a random factor. The log values are reported here because the BFs were too large.

Effects	P(prior)	P(posterior)	$\text{BF}_{\text{inclusion}}$
SetSize	0.26	1.40e-03	0.001
Emotion	0.26	1.68e-106	Inf
Mask	0.26	1.61e-106	25.61
SetSize:Emotion	0.26	1.07e-06	7.66e-04
SetSize:Mask	0.26	2.21e-08	1.58e-05
Emotion:Mask	0.26	1	6.19e+105
SetSize:Emotion:Mask	0.05	3.42e-17	2.24e-06

Table 2. Analysis of effects (proportion of “more negative” responses).

Effects	P(prior)	P(posterior)	$\text{BF}_{\text{inclusion}}$
SetSize	0.26	1	5.57e+33
Emotion	0.26	0.98	44.29
Mask	0.26	1	4.61e+27
SetSize:Emotion	0.26	2.66e-05	2.72e-05
SetSize:Mask	0.26	7.58e-05	2.58e-05
Emotion:Mask	0.26	4.67e-07	4.77e-07
SetSize:Emotion:Mask	0.05	1.19e-24	1.25e-09

Table 3. Analysis of Effects (RTs).

Mask condition	Slope estimate	Est. error	CI. lower	CI. upper
Neither	0.09	0	0.086	0.091
Either	0.05	0	0.052	0.057
Both	0.04	0	0.038	0.042

Table 4. Simple slopes analysis with mask as a moderator (proportion of “more negative” response data).

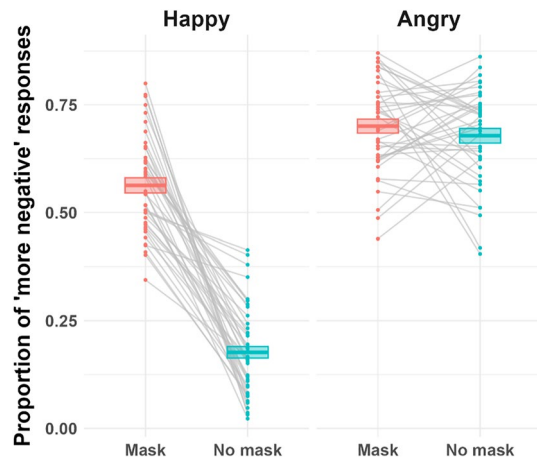


Figure 3. Direct Comparison of the Mask Crowd With the No Mask Crowd (Either Condition). *Note.* In the *either* condition, only one set—either a probe set (i.e., expressive crowd) or a control set (i.e., neutral crowd)—wore facial masks. The x-axis labels indicate the mask condition of the probe set in both the happy and angry emotion conditions. For example, in the left column labeled as “Happy,” the pink box plot indicates the proportion of “more negative” responses assigned to the happy crowd when this crowd was wearing masks and was compared with a control set containing a neutral crowd that was not wearing masks. The blue box plot indicates the proportion of “more negative” responses assigned to the happy crowd when this crowd was not wearing masks and was compared with a control set containing a neutral crowd that was wearing masks.

that in which the happy crowd did not wear masks (to be compared with a neutral crowd wearing masks) as shown in the left panel of Fig. 3.

In approximately 53% of the trials, participants judged the happy crowd (probe set) wearing masks as “more negative” than the neutral crowd (control set) not wearing masks. Meanwhile, participants perceived the happy crowd as “more negative” only in approximately 17% of the trials when faces in the happy crowd were not wearing masks and compared with a neutral crowd wearing masks. When the average emotion of the probe set was angry, the proportion of “more negative” responses for the probe set was the same, regardless of whether the crowd in the probe set or control set wore facial masks (right panel of Fig. 3).

The differential effects of facial masks on the perception of happy and angry average emotions suggest that facial masks considerably disrupt the perception of happy expressions to the extent that observers misperceive a happy crowd wearing masks as an angry crowd. These results were confirmed by the inclusion BFs, which indicated very strong evidence for the main effects of the emotion category ($BF_{incl} = Inf$) and the set with facial masks ($BF_{incl} = 1.30e+135$), as well as the interaction between the two factors ($BF_{incl} = 5.43e+103$; Table 5).

The RT data showed very strong evidence for the effect of set size, which is supported by the inclusion BFs ($BF_{incl} = 5.57e+33$; Table 3). The faster RTs with the larger set sizes (Fig. 2b) suggested that having more faces in a crowd enabled participants to make judgments about average emotions more quickly without affecting the response pattern. The RTs decreased as a function of the average intensity of angry emotion in a crowd, indicating that the more intense the angry emotion was on average, the faster the participants judged a crowd as more negative than a neutral crowd. This observation was supported by very strong evidence for the effect of emotion ($BF_{incl} = 44.29$). Lastly, there was very strong evidence for the effect of the mask condition ($BF_{incl} = 4.61e+27$): the RTs increased across the mask conditions in the order of *neither*, *either*, and *both*. The results showed no other supporting evidence for interaction terms (all $BF_{incl} < 0.33$).

Similar to our analysis of the proportion of “more negative” responses (Fig. 3), we analyzed the RTs in the *either* condition. We found that participants’ RTs were faster for the crowds without masks ($BF_{incl} = 2.34e+04$; Table 6). However, there was no significant difference between the emotion categories, nor was there interaction between the factors.

Effects	P(prior)	P(posterior)	BF _{inclusion}
EmotionCategory	0.40	1.84e−104	Inf
Mask presence of the emotional crowd	0.40	1.84e−104	1.30e+135
EmotionCategory:Mask presence of the emotional crowd	0.20	1	5.43e+103

Table 5. Either condition: analysis of effects (proportion of “more negative” responses).

Effects	P(prior)	P(posterior)	BF _{inclusion}
EmotionCategory	0.40	1.82e-03	0.002
Mask presence of the emotional crowd	0.40	1	2.34e+04
EmotionCategory:Mask presence of the emotional crowd	0.20	5.73e-04	0.296

Table 6. Either condition: analysis of effects (RTs).

General discussion

Reading emotions only from the upper half of faces is more relevant to people's daily experiences today than ever before. This situation provides important scientific implications on how partial facial information affects the evaluation of different groups of multiple faces in terms of emotional valences and intensities. Everyday items, such as sunglasses, scarves, veils, caps, hats, helmets, and medical masks, lead to the partial obstruction of faces, hampering people's ability to recognize others' facial identities and emotions.

However, it has not been tested whether people can still quickly make efficient and reliable judgments about the overall moods of crowds wearing masks, relying only on the visible upper halves of faces. In the present study, we investigated how people's ability to perceive crowds' emotions changed when the faces in the crowds varied in terms of emotional valence, intensity, and crowd size. We found that wearing masks impaired people's ability to judge the average emotions of crowds, especially happy crowds: Happy crowds that wore masks were likely to be judged as negative. Facial masks reduced the sensitivity and speed with which participants perceived the overall mood of a facial crowd. Nevertheless, people could correctly distinguish the emotional intensities of facial crowds wearing masks better than chance. As the crowd size increased, participants could judge the crowd's overall mood faster with similar accuracy.

When the bottom halves of people's faces are hidden by facial masks, one's overall mood judgments of facial crowds can be systematically biased to be more negative than when people do not wear masks. Social communication could be improved if individuals who wear masks were to be over-expressive and convey their emotions to compensate for wearing a mask. However, facial features contribute to emotional perceptions in different ways depending on which facial expression is expressed. Beaudry et al.⁴¹ showed that hiding the mouth area disrupted the categorization of happy expressions more than hiding the eye/brow area; the opposite trend was observed for angry expressions. Furthermore, hiding the eye/brow region did not affect happy emotion categorization, whereas it disrupted angry emotion categorization. Although East Asians tend to distribute their fixations around eye areas to judge facial expressions more than Westerners⁴², the effect of hiding the mouth or eye/brow areas on categorizing happy and angry faces was similar between Canadian⁴¹ and Korean observers⁴³. Our results are consistent with these previous findings, as we found that hiding the mouth region with a mask impaired the discrimination of the intensity of the angry and (especially) the happy emotions of a crowd.

Our results suggest that concealing the mouth area hampers not only the emotion categorization of a single face but also the ensemble emotion judgment of multiple faces. These findings suggest that the quality of individual members' representations significantly impacts ensemble perception. Previous research also showed that reduced individual emotion intensities due to face inversion impaired observers' overall emotion intensity judgments²⁴. Conversely, when the fast flicker adaptation enhanced individual emotion perception, the sensitivity of overall mood judgments increased, both for happy and angry expressions⁴⁴. Additional evidence supports the finding that there is a strong relationship between ensemble face processing and individual face processing. When specific facial information (such as gaze direction) cannot be ignored during the categorization of individual face emotion⁴⁵, the same applies to ensemble face processing, as evident in the significant error correlation between ensemble representations of gaze directions and facial expressions⁴⁶. The facial information extracted from a single face transfers to the processing of an ensemble of faces. These findings underscore the interdependence of individual face processing and ensemble perception, highlighting how the properties of individual faces shape the overall perception of the emotion of a crowd of those individuals.

Ensemble representations are a form of efficient coding that allows observers to create a rapid, unified impression of multiple individual elements. This ensemble coding shows the wisdom of crowds effect, indicating that when measurements are averaged, benefits arise because the errors caused by noisy representations of individuals cancel out each other^{24,47-54}. Thus, the benefits of averaging likely depend on the number of individual measurements averaged, with more measurements resulting in greater benefits. Our results confirm the power of averaging by showing that the more individual faces are present in a crowd, the more quickly participants can respond without sacrificing accuracy. This is the case even when perceiving the emotions of individual faces becomes more challenging when people are wearing facial masks. In addition, that participants' performance is unaffected by the increasing set sizes suggests that ensemble coding can be achieved in parallel. The facial expressions extracted from a group of people are useful for promoting communication with others and regulating subsequent social behaviors^{16,18}. A person may react more quickly to a social context involving a large than a small group, but further research is required to empirically test this. Altogether, the current study provides a novel implication: under ambiguous social situations where masks interfere with the perception of others' emotions, perceiving a set of faces could be more advantageous than relying on an individual face.

Previous studies have found set size effects in the opposite direction from our results, particularly in crowds of faces^{55,56}. Ji and Pourtois⁵⁵ found that the accuracy of averaging facial expressions decreased as the set size increased. The side-by-side presentation might have decreased the accuracy of a larger set due to confounding factors such as crowding and cluttered display. In contrast, in the current study, we presented the two facial

crowds sequentially so that individual faces were less likely to be crowded or cluttered. Another possibility is the interaction between set size effects and stimuli variability. The visual system cannot process more information than its capacity limit. The capacity limit constrains the effect of noise cancellation in large sets, which is reflected in the late noise of averaging^{48,54}. When stimuli variability is high, the effect of noise cancellation as the set size increases is weakened because the stimuli are not densely centered around the representative value. Even in the work of Ji and Pourtois³⁵, when emotion variance was low, the ability to average facial expressions did not depend on set size.

In the present work, we investigated how people's overall judgments of the mood of a crowd are influenced by the occlusion of part of the faces with masks. We found that participants' judgments were impaired by masks, and such impairment was more pronounced for happy emotions than for angry emotions. Nevertheless, due to the power of averaging, observers could determine the overall mood of a large crowd more quickly than a small crowd. Altogether, our results suggest that people's ability to apprehend social-emotional contexts and react to them can be efficiently aided by large crowds, even when individual emotional representations are impaired or partially occluded.

Data availability

The datasets generated and/or analyzed during the current study are available in the Open Science Framework repository, <https://osf.io/75zwr>.

Received: 28 February 2023; Accepted: 25 August 2023

Published online: 31 August 2023

References

- Freud, E., Stajduhar, A., Rosenbaum, R. S., Avidan, G. & Ganel, T. The COVID-19 pandemic masks the way people perceive faces. *Sci. Rep.* **10**, 22344. <https://doi.org/10.1038/s41598-020-78986-9> (2020).
- Calbi, M. *et al.* The consequences of COVID-19 on social interactions: An online study on face covering. *Sci. Rep.* **11**, 2601. <https://doi.org/10.1038/s41598-021-81780-w> (2021).
- Marini, M., Ansani, A., Paglieri, F., Caruana, F. & Viola, M. The impact of facemasks on emotion recognition, trust attribution and re-identification. *Sci. Rep.* **11**, 5577. <https://doi.org/10.1038/s41598-021-84806-5> (2021).
- Ruba, A. L. & Pollak, S. D. Children's emotion inferences from masked faces: implications for social interactions during COVID-19. *PLOS ONE* **15**(12), e0243708. <https://doi.org/10.1371/journal.pone.0243708> (2020).
- Farah, M. J., Wilson, K. D., Drain, M. & Tanaka, J. N. What is "special" about face perception?. *Psychol. Rev.* **105**(3), 482–498. <https://doi.org/10.1037/0033-295X.105.3.482> (1998).
- Maurer, D., Le Grand, R. & Mondloch, C. J. The many faces of configural processing. *Trends Cogn. Sci.* **6**(6), 255–260. [https://doi.org/10.1016/S1364-6613\(02\)01903-4](https://doi.org/10.1016/S1364-6613(02)01903-4) (2002).
- Tanaka, J. W. & Farah, M. J. Parts and wholes in face recognition. *Q. J. Exp. Psychol.* **46**(2), 225–245. <https://doi.org/10.1080/14640749308401045> (1993).
- Lieberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* **54**(5), 358–368. <https://doi.org/10.1037/h0044417> (1957).
- Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L. & Rowland, D. Categorical perception of morphed facial expressions. *Vis. Cogn.* **3**(2), 81–118. <https://doi.org/10.1080/713756735> (1996).
- Etcoff, N. L. & Magee, J. J. Categorical perception of facial expressions. *Cognition* **44**(3), 227–240. [https://doi.org/10.1016/0010-0277\(92\)90002-Y](https://doi.org/10.1016/0010-0277(92)90002-Y) (1992).
- Roberson, D. & Davidoff, J. The categorical perception of colors and facial expressions: The effect of verbal interference. *Mem. Cognit.* **28**(6), 977–986. <https://doi.org/10.3758/BF03209345> (2000).
- Young, A. W. *et al.* Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition* **63**(3), 271–313. [https://doi.org/10.1016/S0010-0277\(97\)00003-6](https://doi.org/10.1016/S0010-0277(97)00003-6) (1997).
- Beale, J. M. & Keil, F. C. Categorical effects in the perception of faces. *Cognition* **57**(3), 217–239. [https://doi.org/10.1016/0010-0277\(95\)00669-x](https://doi.org/10.1016/0010-0277(95)00669-x) (1995).
- Cowen, A., Sauter, D., Tracy, J. L. & Keltner, D. Mapping the passions: Toward a high-dimensional taxonomy of emotional experience and expression. *Psychol. Sci. Public Interest* **20**(1), 69–90. <https://doi.org/10.1177/1529100619850176> (2019).
- Watson, D. & Stanton, K. Emotion blends and mixed emotions in the hierarchical structure of affect. *Emot. Rev.* **9**(2), 99–104. <https://doi.org/10.1177/1754073916639659> (2017).
- Goldenberg, A., Sweeny, T. D., Shpigel, E. & Gross, J. J. Is this my group or not? The role of ensemble coding of emotional expressions in group categorization. *J. Exp. Psychol. Gen.* **149**(3), 445–460. <https://doi.org/10.1037/xge0000651> (2020).
- Huis in 't Veld, E. M. & de Gelder, B. From personal fear to mass panic: The neurological basis of crowd perception. *Hum. Brain Mapp.* **36**(6), 2338–2351. <https://doi.org/10.1002/hbm.22774> (2015).
- Ristic, J. & Capozzi, F. Mechanisms for individual, group-based and crowd-based attention to social information. *Nat. Rev. Psychol.* **1**(12), 721–732. <https://doi.org/10.1038/s44159-022-00118-z> (2022).
- van Kleef, G. A. & Côté, S. The social effects of emotions. *Annu. Rev. Psychol.* **73**, 629–658. <https://doi.org/10.1146/annurev-psych-020821-010855> (2022).
- Alt, N. P. & Phillips, L. T. Person perception, meet people perception: Exploring the social vision of groups. *Perspect. Psychol. Sci.* **17**(3), 768–787. <https://doi.org/10.1177/17456916211017858> (2022).
- Haberman, J. & Whitney, D. Seeing the mean: Ensemble coding for sets of faces. *J. Exp. Psychol. Hum. Percept. Perform.* **35**(3), 718–734. <https://doi.org/10.1037/a0013899> (2009).
- Im, H. Y. *et al.* Differential hemispheric and visual stream contributions to ensemble coding of crowd emotion. *Nat. Hum. Behav.* **1**(11), 828–842. <https://doi.org/10.1038/s41562-017-0225-z> (2017).
- Im, H. Y., Cushing, C. A., Ward, N. & Kveraga, K. Differential neurodynamics and connectivity in the dorsal and ventral visual pathways during perception of emotional crowds and individuals: A MEG study. *Cogn. Affect. Behav. Neurosci.* **21**, 776–792. <https://doi.org/10.3758/s13415-021-00880-2> (2021).
- Sun, J. & Chong, S. C. Power of averaging: Noise reduction by ensemble coding of multiple faces. *J. Exp. Psychol. Gen.* **149**(3), 550–563. <https://doi.org/10.1037/xge0000667> (2020).
- Levitani, C. A. *et al.* Mask wearing affects emotion perception. *i-Perception* **13**, 20416695221107390. <https://doi.org/10.1177/20416695221107391> (2022).
- Schönbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M. & Perugini, M. Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychol. Methods* **22**(2), 322–339. <https://doi.org/10.1037/met0000061> (2017).

27. Brainard, D. The psychophysics toolbox. *Spatial Vision* **10**, 433–436 (1997).
28. Pelli, D. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision* **10**, 437–442 (1997).
29. Yang, J.-W., Chung, K. & Chong, S. C. *The Yonsei face database* (Yonsei University, Seoul, Korea, 2015).
30. Langeslag, S. J., Gootjes, L. & van Strien, J. W. The effect of mouth opening in emotional faces on subjective experience and the early posterior negativity amplitude. *Brain Cogn.* **127**, 51–59. <https://doi.org/10.1016/j.bandc.2018.10.003> (2018).
31. Calvo, M. G. & Nummenmaa, L. Detection of emotional faces: Salient physical features guide effective visual search. *J. Exp. Psychol. Gen.* **137**(3), 471–493. <https://doi.org/10.1037/a0012771> (2008).
32. Cui, S., Song, S., Si, J., Wu, M. & Feng, J. The influence of mouth opening and closing degrees on processing in NimStim facial expressions: An ERP study from Chinese college students. *Int. J. Psychophysiol.* **162**, 157–165. <https://doi.org/10.1016/j.ijpsycho.2021.01.013> (2021).
33. Horstmann, G. & Bauland, A. Search asymmetries with real faces: Testing the anger-superiority effect. *Emotion* **6**, 193–207. <https://doi.org/10.1037/1528-3542.6.2.193> (2006).
34. Horstmann, G., Lipp, O. V. & Becker, S. I. Of toothy grins and angry snarls—Open mouth displays contribute to efficiency gains in search for emotional faces. *J. Vis.* **12**(5), 1–15. <https://doi.org/10.1167/12.5.7> (2012).
35. Bürkner, P. C. brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28. <https://doi.org/10.18637/jss.v080.i01> (2017).
36. Makowski, D., Ben-Shachar, M. S. & Lüdtke, D. bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *J. Open Source Softw.* **4**, 1541. <https://doi.org/10.21105/joss.01541> (2019).
37. Wagenmakers, E. J. & Brown, S. On the linear relation between the mean and the standard deviation of a response time distribution. *Psychol. Rev.* **114**(3), 830–841. <https://doi.org/10.1037/0033-295X.114.3.830> (2007).
38. Hinne, M., Gronau, Q. F., van den Bergh, D. & Wagenmakers, E. J. A conceptual introduction to Bayesian model averaging. *Adv. Methods Pract. Psychol. Sci.* **3**(2), 200–215. <https://doi.org/10.1177/2515245919898657> (2020).
39. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795. <https://doi.org/10.1080/01621459.1995.10476572> (1995).
40. van den Bergh, D. et al. A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. *Ann. Psychol.* **120**, 73–96. <https://doi.org/10.3917/anpsy1.201.0073> (2020).
41. Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I. & Tapp, R. Featural processing in recognition of emotional facial expressions. *Cogn. Emot.* **28**(3), 416–432. <https://doi.org/10.1080/02699931.2013.833500> (2014).
42. Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G. & Caldara, R. Cultural confusions show that facial expressions are not universal. *Curr. Biol.* **19**, 1543–1548. <https://doi.org/10.1016/j.cub.2009.07.051> (2009).
43. Kim, G., Seong, S. H., Hong, S. S. & Choi, E. Impact of face masks and sunglasses on emotion recognition in South Koreans. *PLOS ONE* **17**(2), e0263466. <https://doi.org/10.1371/journal.pone.0263466> (2022).
44. Lee, J. & Chong, S. C. Quality of average representation can be enhanced by refined individual items. *Atten. Percept. Psychophys.* **83**(3), 970–981. <https://doi.org/10.3758/s13414-020-02139-3> (2021).
45. Adams, R. B. & Kleck, R. E. Perceived gaze direction and the processing of facial displays of emotion. *Psychol. Sci.* **14**(6), 644–647. <https://doi.org/10.1046/j.0956-7976.2003.psci.1479.x> (2003).
46. Kwon, D. & Chong, S. C. The relationship between ensemble representations of facial information. *Vis. Res.* **203**, 108156. <https://doi.org/10.1016/j.visres.2022.108156> (2023).
47. Alvarez, G. A. Representing multiple objects as an ensemble enhances visual cognition. *Trends Cogn. Sci.* **15**(3), 122–131. <https://doi.org/10.1016/j.tics.2011.01.003> (2011).
48. Baek, J. & Chong, S. C. Distributed attention model of perceptual averaging. *Atten. Percept. Psychophys.* **82**(1), 63–79. <https://doi.org/10.3758/s13414-019-01827-z> (2020).
49. Baek, J. & Chong, S. C. Ensemble perception and focused attention: Two different modes of visual processing to cope with limited capacity. *Psychon. Bull. Rev.* **27**(4), 602–606. <https://doi.org/10.3758/s13423-020-01718-7> (2020).
50. Galton, F. One vote, one value. *Nature* **75**, 414–414. <https://doi.org/10.1038/075414a0> (1907).
51. Haberman, J. & Whitney, D. The visual system discounts emotional deviants when extracting average expression. *Atten. Percept. Psychophys.* **72**(7), 1825–1838. <https://doi.org/10.3758/APP.72.7.1825> (2010).
52. Lee, H., Baek, J. & Chong, S. C. Perceived magnitude of visual displays: Area, numerosity, and mean size. *J. Vis.* **16**(3), 12. <https://doi.org/10.1167/16.3.12> (2016).
53. Robitaille, N. & Harris, I. M. When more is less: Extraction of summary statistics benefits from larger sets. *J. Vis.* **11**(12), 18. <https://doi.org/10.1167/11.12.18> (2011).
54. Utochkin, I. S., Choi, J. & Chong, S. C. A population response model of ensemble perception. *Psychol. Rev.* <https://doi.org/10.1037/rev0000426> (2023).
55. Ji, L. & Pourtois, G. Capacity limitations to extract the mean emotion from multiple facial expressions depend on emotion variance. *Vision Res.* **145**, 39–48. <https://doi.org/10.1016/j.visres.2018.03.007> (2018).
56. Ji, L., Pourtois, G. & Sweeny, T. D. Averaging multiple facial expressions through subsampling. *Vis. Cogn.* **28**(1), 41–58. <https://doi.org/10.1080/13506285.2020.1717706> (2020).

Author contributions

All authors conceived the research and designed the Experiments, J.C. and Y.Y. conducted the experiments, and all authors analysed the results. All authors reviewed the manuscript.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2022R1A2C3004133) to S.C. and Basic Science Research Program through the NRF funded by the Ministry of Education (2022R111A2067731) to S.J.J. The data are available at <https://osf.io/75zwr>. This study was not preregistered. Part of the data and ideas in this research article was previously presented at the ECVP 2021.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-41366-0>.

Correspondence and requests for materials should be addressed to S.C.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023