



OPEN

Judgment of crowds as emotional increases with the proportion of black faces

Amit Goldenberg^{1,3,4,5}✉, Kyle LaFollette^{2,5}, Zi Huang¹, Erika Weisz³ & Mina Cikara³

Decades of research have sought to characterize how racialization affects the evaluation of faces. However, research has mostly focused on the perception of single faces. In the current project we address this gap by examining whether racialization affects crowd perception, even in the absence of perceived differences at the single-face level. In Experiments 1–2, participants viewed crowds of faces expressing different degrees of happiness or anger and different ratios of lighter-skinned faces (racialized as White) and darker-skinned faces (racialized as Black). A higher proportion of Black faces led to a greater likelihood of perceiving the crowd as emotional. Experiment 3 confirmed this effect was not solely driven by the contrast between White and Black faces within a crowd, as crowds composed of entirely Black faces were more likely to be evaluated as emotional than crowds of White and Black faces. Using hierarchical drift diffusion models, we compared potential mechanisms and found that Black faces expressing emotions weighed more heavily than White faces in judging whether a crowd was emotional. These results shed light on an important way in which racialization impacts judgments of collective emotion.

Imagine standing in front of a crowd and trying to evaluate how emotional it is. Your eyes scan the audience, capturing some of the faces expressing emotion, and you then decide that the crowd is not really emotional yet. Now consider parallel scenarios in which the perceived racial composition of the audience varies—one in which people are perceived as predominantly White and one in which people are perceived as predominantly Black. Assuming that the actual emotional intensity exhibited by the demonstrators is equal, does the racial composition of the crowd affect one's evaluation of the crowd's emotionality? If so, what perceptual mechanisms could account for differences in this judgment?

Evaluating crowds' emotions

When people assess the emotions of crowds, they tend to automatically generate summary statistics, particularly averages, to summarize emotions quickly and with relative accuracy^{1–5}. However, perceivers' errors do not seem to be normally distributed around the true average of the crowd's emotion. Instead, they are biased towards assuming that groups are more emotional than they actually are, a phenomenon known as the crowd-emotion-amplification effect^{6,7}. The crowd-emotion-amplification effect increases with crowd size. It also tends to be greater when people are judging the emotionality of crowds expressing anger as compared to happiness, although this effect tends to be weak.

People are well tuned to the racial composition of crowds, which can affect how individuals perceive crowd emotionality^{8–11}. In past experimental work, perceivers tend to associate Black individuals with aggression and anger more than White individuals^{12–15}. In some cases perceivers are quicker at evaluating anger expressed by Black targets while quicker at evaluating happiness expressed by White targets^{16–19}. These findings suggest that anger expressed by crowds of Black people would be evaluated as more intense than happiness expressed by crowds of Black people. However, it is unclear how the racial composition of the crowd would affect the evaluation of their emotions. Research on attention bias, such as dot probe tasks examining bias toward White and Black faces, provides some indirect evidence on which faces in a crowd might attract attention^{20,21}. These studies suggest that attentional bias depends on several factors, including initial priming, target expression, exposure time, and motivation to respond without prejudice. Due to the complexity of these many factors, it is challenging to make clear predictions regarding emotional evaluation of crowds from findings based on evaluating individuals.

¹Harvard Business School, Harvard University, Boston, USA. ²The University of Chicago, Booth School of Business, Chicago, USA. ³Department of Psychology, Harvard University, Cambridge, USA. ⁴Digital, Data, and Design Institute, Harvard University, Boston, USA. ⁵Amit Goldenberg and Kyle LaFollette contributed equally to this work. ✉email: agoldenberg@hbs.edu

Using drift diffusion to decipher between bias and sensitivity

Irrespective of the direction of effect, it is important to understand whether evaluations of emotionality are driven by *bias* or *sensitivity*²². To exemplify this potential difference, take the case of increased probability of evaluating crowd's emotionality when the proportion of Black faces increases: *Bias* in this case means that merely realizing that there are more Black faces in the crowd should immediately lead to stronger attributions of emotionality. By contrast, *sensitivity* in this case means that each emotional expression detected in a Black face will count more toward an 'emotional' decision compared to a White face, making observers more likely to evaluate the crowd as emotional quicker when there are more Black faces. Said another way, bias is driven by the assumption that the crowd is more emotional if it contains Black people, regardless of the emotions they express. Sensitivity refers to the increased impact of emotional Black faces towards an evaluation of emotionality. Both are forms of prejudice, but these are very different types of prejudice, and differentiating between them can potentially help us understand which form of prejudice needs to be corrected²⁰. In other words, should we address the belief that merely containing Black people makes a crowd seem emotional, or should we address perceivers' heightened sensitivity to emotional expressions of Black people in a crowd?

One way to differentiate bias and sensitivity is by examining people's decisions using a drift diffusion model. Drift diffusion models (DDM) are designed to evaluate inputs into decisions as a function of participants' response times. A key assumption in DDMs is that people accumulate relevant evidence from their environment until such accumulation reaches a certain decision threshold, which here is the decision that the crowd is either emotional or not (marked by the black boundaries of Fig. 1). Decisions may be affected by bias, sensitivity, or both^{23,24}. Bias, represented by z (Fig. 1A), is captured by the intercept of the model. A bias towards evaluating a crowd with more Black faces as emotional means that the intercept starts closer to the 'emotional' threshold when the proportion of Black faces in the crowd increases (Z_1 compared to Z_2). The second is evidence accumulation or sensitivity, represented by δ (Fig. 1B). Sensitivity attempts to capture the rate at which evidence is accumulated before a decision is reached. A steeper δ when the proportion of Black faces increases suggests that evidence from Black faces regarding the crowd's emotionality is accumulated faster compared to White faces (δ_1 compared to δ_2).

In the past decade, drift diffusion models have been used to characterize instances of bias and sensitivity in perceived race^{25,26}. Several studies suggest that the effect of perceived race on social decisions is driven by increased sensitivity, rather than bias, to Black faces. Other studies have examined crowd evaluation using DDM, particularly looking at people's categorization of the crowd's valence, indirectly suggesting that both bias and sensitivity drive decision making^{27,28}. However, existing studies do not provide direct evidence on the question of how the perceived race of the crowd members affects attributions of emotionality.

The present research

We conducted a set of 3 empirical experiments to examine how the proportion of Black/White faces in a crowd contributes to its evaluation as emotional or not. Our experiments had four preregistered hypotheses (Experiment 1: <https://osf.io/6smu8/>, Experiment 2: <https://osf.io/ctjz6/>, Preregistration for Study 1 was uploaded to the system on August 26, 2021, as a document but was not formally submitted as a preregistration due to an error. The document with the appropriate date can be found at the link. We did not formally submit the preregistration upon discovering the error to avoid modifying the upload date.) Our first three hypotheses were derived from existing evidence of the crowd-emotion-amplification effect⁶. First (preregistered for both E1 and E2), we predicted that people would be more likely to evaluate crowds with more emotional faces as emotional than non-emotional. Second (preregistered for E1, but not for E2 as results were found to be non-significant in E1), we predicted that larger crowds would increase the probability that the crowd was judged emotional. And third (preregistered in E2 after discovered in E1), we expected that crowds expressing anger, compared to happiness, would be more likely to be evaluated as emotional.

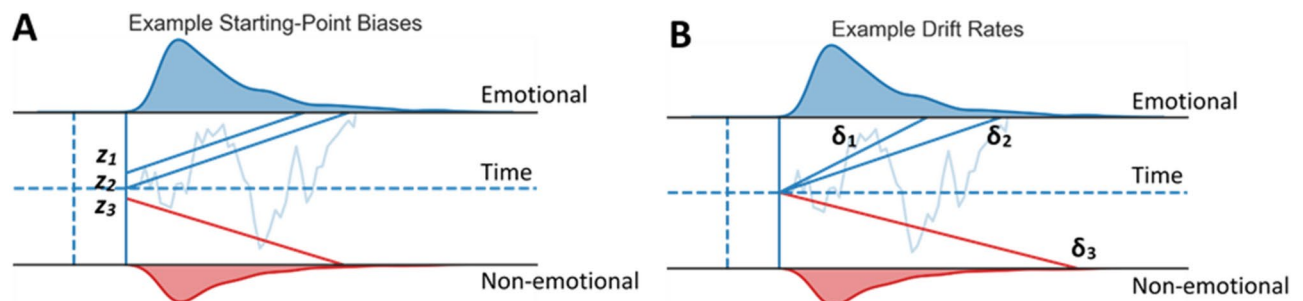


Fig. 1. Example schematics of the drift diffusion process. The black lines represent a decisions threshold, in this case the decision to evaluate the crowd as emotional or non-emotional. The top and bottom distributions reflect response times for emotional and non-emotional inferences. The light blue line represents the noisy diffusion process of evidence accumulation, which results in a sampled response time once it reaches the evidence threshold of either choice. Both differences in the starting-point biases (A) and average rates of evidence accumulation (B) can affect response time. Here, we show how disparate effects of race ratio on response time and choice could be explained by differences in the drift slope and bias intercept.

The fourth prediction (preregistered both in E1 and E2) and the most relevant to the current project relates to the probability of evaluating a crowd as emotional as a function of targets' perceived race. Our hypothesis on this topic has changed as a function of our findings in Study 1. Based on our piloting study, which was done with a limited sample and close to the social movement related to the death of George Floyd, we estimated that an increase in the proportion of Black faces would lead to a *decrease* in the probability of evaluating the crowd as emotional. However, based on results of Experiment 1, which suggested that crowd emotionality *increased* with the proportion of Black faces, we initiated a new pre-registered experiment and conducted a direct replication of the study in Experiment 2 showing similar results to Study 1 (see full comparison of hypotheses in SM). Experiment two was designed to test another preregistered hypothesis, which is that people who are more "warm" towards black people (according to the survey thermometer) would be less affected by the proportion of black faces in evaluating the crowd emotion. We added this hypothesis in Experiment 2 as a result of findings of Experiment 1, but did not find this effect in experiment 2 and therefore do not report it in the main manuscript.

Experiment 3 was designed to replicate the results and to eliminate the possibility that our results were merely driven by the contrast between White and Black faces *within* a crowd. It was not preregistered but the structure was almost identical to that of Experiments 1 and 2. We therefore tested similar hypotheses.

Drift diffusion models were applied to all of the experiments to specifically examine differences between bias and sensitivity, though note that drift diffusion analyses were not preregistered.

Experiments 1–2: testing the effect of target race on judgments of crowd emotionality

The goal of experiments 1 and 2 was to examine the effect of racial composition of a crowd on judgments of the crowd's emotionality.

Method

Participants

This study (as all other studies) was approved by the ethics committee at Harvard Business School. All methods were carried out in accordance with guidelines and all participants filled out an informed consent before participating in this and all other studies. We used data from a recent pilot similar to the current experiment to evaluate the appropriate sample size for the experiment (see⁷ for a detailed description). Based on these findings, we concluded that 150 participants completing 50 trials would be enough to power our analysis to 99.99%.

Participants were recruited from Mechanical Turk and compensated \$3. All of our participants were American and spoke English as their first language. Out of the 150 that we collected, 148 participants completed the full task in Experiment 1 and 146 in Experiment 2. Looking at the data of Experiment 1, we decided to apply another selection criteria that was not preregistered for Experiment 1 but was included in the pre-registration for Experiment 2: we removed participants who made the exact same emotionality choice for 45 of the 50 trials (45 emotional or 45 not emotional). Such choice-patterns are likely to be caused by rushing through the task by choosing the same outcome. Removing these participants did not change the overall outcome and improved the quality of the data (see SM for full description and comparison). Our final sample of Experiment 1 was therefore 141 (men: 76, women: 63, other: 2; age: $M = 37.60$, $SD = 10.20$) and 138 for Experiment 2 (men: 78, women: 59, other: 1; age: $M = 39.57$, $SD = 12.04$). We decided not to limit recruitment to monoracial White participants but rather to examine the results after the fact to test whether our findings were moderated by participants' self-reported race (see full analysis in SM). Participants were allowed to mark all racial/ethnic categories with which they identify (Experiment 1: Black = 18, Hispanic = 6, White = 111, Asian = 9, other = 4; Experiment 2: Black = 13, Hispanic = 5, White = 117, Asian = 6).

Stimuli

To create the stimuli set for our task, we used a morph from a face set recently developed for an investigation of ensemble face perception¹⁰, which was based on eight exemplar faces of men from the NimStim face set²⁹. Four faces were White and four were Black. For each identity in the set, and for two emotions for each identity—happiness and anger—the morphing program created a morphed scale of 0%—completely neutral to 100%—completely emotional in increments of 2% (1–50 scale, see Fig. 2). This meant that each identity had two sets of 50 faces expressing emotions from neutral-to-happy and neutral-to-angry, respectively.

Pilot

One of the most important preconditions of this experiment was that there would be no perceptual differences at the single-face level between the White and Black faces: specifically that there were no systematic differences in the threshold for judging black versus white racialized faces as emotional. Differences at the single face level could aggregate to differences in crowd evaluation in ways that are unrelated to the mechanisms we intended to test. We therefore conducted a pilot experiment to ensure there were no race-related differences at baseline of *single* face emotional evaluations (see SM for full description and analysis). Participants saw a single White or Black face whose intensity was randomly drawn from the face morphs of either neutral to happy or neutral to angry. After viewing each face, participants indicated if they judged the face as emotional or not emotional. This test allowed us to evaluate the morph threshold at which each face *identity* was considered emotional.

Our pilot analysis indicated that different identities were indeed associated with different thresholds. Of the eight face identities we initially had, we selected a subset of four identities in which the emotionality threshold for White and Black racialized-faces was equal. This process ensured that any differences we observed at the crowd level would be a result of aggregation rather than individual level evaluation. Our final stimulus set was therefore comprised of four face identities, two White and two Black, expressing either varying degrees of anger or happiness.

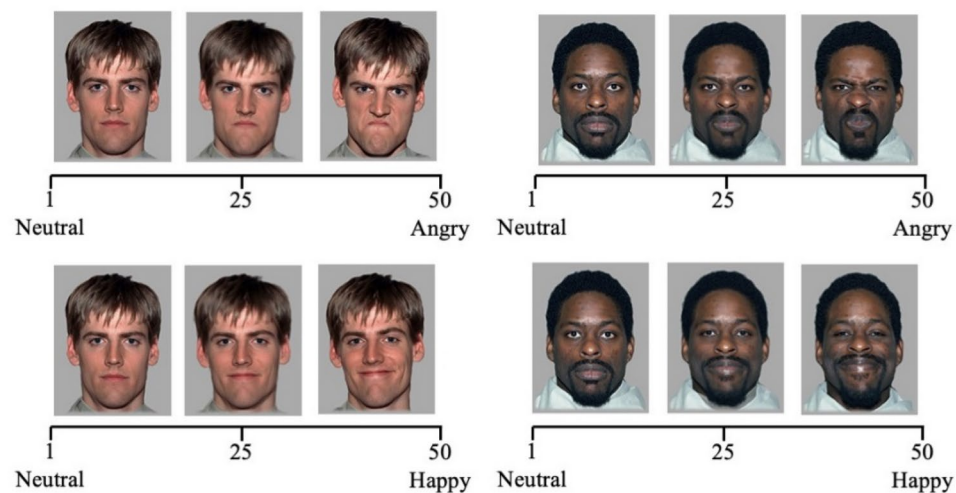


Fig. 2. A sample of three faces from the neutral-to-angry scale (top) and from the neutral-to-happy scale (bottom) that were used in the experiments. Values of 25 and 50 correspond to 50 and 100% intensities in our morph range, respectively.

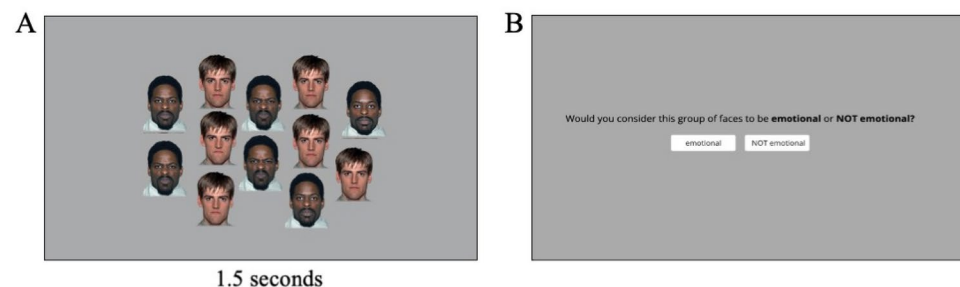


Fig. 3. The task used in Experiments 1–3. Participants saw a crowd of 4, 8 or 12 faces comprised of two identities and expressing either different degrees of anger or happiness that appeared on the screen for 1.5 s (A). Participants then indicated whether they would judge this group of faces to be emotional or not emotional (B).

Procedure

Ethics committee approval was obtained prior to the data collection. The experiment consisted of 50 trials. In each trial, participants first saw an array containing 4, 8, or 12 faces divided into two identities: one White and one Black (Fig. 2a). Our primary manipulation was the proportion of Black faces in the crowd which was 25, 50, or 75%. We did not have a 0 or 100% proportion in these experiments because we were concerned that these trials would reveal the purpose of the experiment and potentiate social desirability responding, but we added these conditions in Experiment 3. We chose to use only two face identities in each trial so that the judgment of emotionality would be as simple as possible and that any effect of crowd estimation could not be caused by participants attending to some identities in the crowd more than others. In addition to manipulating the proportion of Black/White faces in the crowd and the size of each array, the intensity of *each face* in the face array was randomly drawn from a 1–50 morph, 1 corresponding to a neutral face, and 50 to the most-angry/most-happy face (Fig. 3A). The distribution of the actual emotional intensity of all arrays was Gaussian $N(25.52, 5.32)$.

Importantly, each trial contained either neutral-to-angry and neutral-to-happy faces. We did not mix the happy and angry faces in the same trial for two reasons. First, doing so could undermine our ability to interpret the findings: if participants fixated on one extremely negative and one extremely positive face, then they may be confused as to how to respond. Second, the most-happy and most-angry faces may not be considered equal in intensity, thus making the average between the two different from zero.

Participants viewed each face array for 1.5 s. This length was chosen based on previous experiments which find that the amplification effect is greatest at this exposure time⁶. After observing the array, participants were asked: “would you consider this group of faces to be emotional or NOT emotional” (Fig. 3B). We decided to use a binary outcome variable for two reasons. First, a binary model allowed us to examine the mechanism—bias or sensitivity—using a DDM. Second, it is likely that decisions based on crowd emotionality are driven by a threshold of the crowd as emotional or not, and we wanted our experiment to mirror these contexts. After

viewing each face array and choosing whether it was emotional, participants saw a fixation cross for a duration between 400–600 ms (randomly determined on each trial) before moving on to the next trial.

After completing the main task, participants filled out a short survey which included the social dominance orientation scale³⁰, a feeling thermometer towards various groups³¹, a Big-Five personality scale³², and demographic questions including age, gender, race and education level. These scales were administered in order to examine potential moderators and led to inconsistent findings which are reported in SM.

Drift diffusion model

Crowd emotion-amplification was modeled as a drift diffusion process. Drift diffusion modeling assumes that peoples' responses are functions of a noisy process of evidence accumulation toward one of two decision thresholds (Fig. 1). Here, one threshold corresponds with an inference that the crowd is emotional, and the other being that the crowd is non-emotional.

Generally speaking, the evaluation of decision making processes via drift diffusion models can be decomposed into four parameters: starting-point z (bias), drift rate δ (sensitivity), boundary separation α , and non-decision time τ . Because our goal was to examine how starting point/bias and sensitivity/drift rate differed as a function of proportion Black faces in a crowd, we fixed the two other parameters: boundary separation and non-decision time. However, we also tested models in which boundary separation was allowed to vary, and found no differences in that parameter across conditions (see SM). We allowed our free parameters to differ as function of proportion of race and array emotion valence to examine how both impacted participants' decisions. To evaluate these parameters, we inferred posterior parameter distributions using a Hamiltonian Monte Carlo No-U-Turn sampler with the Stan probabilistic programming language³³. The HMC algorithm is a Markov chain Monte Carlo (MCMC) sampling method used to estimate the joint posterior of the four DDM parameters.

Hyperpriors for the group-level parameter means and standard deviations were weakly informative and set to standard normal ($N(0,1)$) and half-Cauchy ($HCauchy(0, 5)$), respectively. Four MCMC chains were run in parallel for 20,000 iterations each, of which the first 10,000 were discarded as warm-up. We confirmed convergence by checking the within-chain and between-chain variances of our four MCMC chains. The Gelman-Rubin convergence diagnostic $\hat{R}$ denotes the difference between these two types of variance – we determined that $\hat{R}$ was less than or equal to 1.01 for each posterior distribution, indicating practical equivalence between variances³⁴. For DDM analyses only, we excluded reaction times less than 100 ms and greater than 5000 ms, as these observations were likely false starts or distracted, respectively. Hierarchical DDMs are also particularly valuable for the analysis of outlier subject-level observations. Hierarchical models put less weight on individual extreme observations and instead bias subject-level estimates toward the group-level³⁵. The influence of outlier response times was further mitigated with half-Cauchy distributed group-level variances.

Results and discussion

Hypothesis 1: evaluation of emotionality

Our first hypothesis was the arrays, in general, would be more likely to be evaluated as emotional than non-emotional. In order to examine this hypothesis, we tested the association between the average intensity of the emotion expressed in the crowd and the probability of evaluating that crowd as emotional. This model allowed us to examine the general probability of evaluating arrays as emotional at an average intensity (intercept of the model) and to look at the general connection between the array intensity and the probability of evaluating the array as emotional (slope—manipulation check). We conducted a mixed generalized linear model using the average intensity of the arrays as the independent variable predicting the probability of categorizing the crowd as emotional, our dependent variable. We also used three random intercepts: participant id, the identity of the White face, and the identity of the Black face (preregistered specifically in Experiment 2). We also tested an alternative model with a different random variable structure that nested White and Black facial identities within participants, but this model did not turn out to be superior in the current context (models were compared using likelihood ratio test; $\chi^2 = 5.98$, $p = 0.11$). Results indicated that the intercept of the model was significantly positive, although the effect was weaker in Experiment 2 (Experiment 1: $b = 0.36$ [0.03, 0.70], $SE = 0.12$, $z = 3.05$, $p = 0.002$, $R^2c = 0.18$, Experiment 2: $b = 0.30$ [0.08, 0.83], $SE = 0.15$, $z = 2.01$, $p = 0.04$, $R^2c = 0.11$). We then examined the association between array intensity and the probability of evaluating the array as emotional which served as a manipulation check. Results indicated a very strong association between the actual intensity of the array and the probability that it would be categorized as emotional (Experiment 1: $b = 0.64$ [0.59, 0.70], $SE = 0.02$, $z = 22.35$, $p < 0.001$, $R^2c = 0.18$, Experiment 2: $b = 0.72$ [0.66, 0.78], $SE = 0.03$, $z = 24.28$, $p < 0.001$, $R^2c = 0.11$).

Model specification

After establishing the expected association between the actual intensity to participants' decision, we tested all other hypotheses in a second model by conducting a three way interaction between array size, valence, and proportion of Black faces in the crowd. This model was slightly different than the one we preregistered, in which we said we would test each of the variables in a separate model. We chose an interaction model because it was more parsimonious and conservative, but the results are identical in both analytic approaches. The analysis did not reveal any three-way or two way interactions, either in Experiment 1 or 2 (see full details of the model in SM), suggesting that neither array size nor valence interacted with the proportion of Black faces in the crowd. We then turned to test the main effects of the model.

Hypothesis 2: crowd size

Next, we tested our hypothesis that bigger crowds were more likely to be evaluated as emotional. Results indicated that the association between array size and the probability of evaluating the crowd as emotional was not significant in Experiment 1 ($b = 0.05$ [−0.01, 0.12], $SE = 0.03$, $z = 1.56$, $p = 0.11$, $R^2c = 0.12$) or in Experiment

2 ($b=0.03$ [$-0.03, 0.10$], $SE=0.03$, $z=0.88$, $p=0.37$, $R^2c=0.11$). These results are incongruent with previous findings suggesting that crowd size impacts the evaluation of emotionality⁶. However, in the original set of studies that established the crowd-emotion-amplification effect, array sizes ranged from 1 to 12 with 12 possible variations, while in the current design, arrays only comprised 4, 8, and 12 faces with only three variations. Reducing the number of variations in crowd size likely impacted the ability to detect a significant effect. **Hypothesis 3: Crowd valence.** We then examined whether valence was associated with increased likelihood of evaluating the crowd as emotional. Results indicated that crowds expressing anger (relative to happiness) were more likely to be evaluated as emotional both in Experiment 1 ($b=0.70$ [$-0.01, 0.09$], $SE=0.05$, $z=13.83$, $p<0.001$, $R^2c=0.12$) and in Experiment 2 ($b=0.74$ [$0.64, 0.84$], $SE=0.5$, $z=14.40$, $p<0.001$, $R^2c=0.11$).

Hypothesis 4: perceived race composition

Finally, we tested our central hypothesis: that the proportion of Black faces in the crowd was associated with increased likelihood of evaluating the crowd as emotional. More Black faces in the crowd were associated with increased probability of evaluating the crowd as emotional both in Experiment 1 ($b=0.12$ [$0.05, 0.19$], $SE=0.03$, $z=3.54$, $p<0.001$, $R^2c=0.12$, Fig. 4) and in Experiment 2 ($b=0.08$ [$0.01, 0.15$], $SE=0.05$, $z=2.34$, $p=0.01$, $R^2c=0.11$, Fig. 4).

We also examined whether the tendency to amplify the emotions of crowds with more Black faces was moderated by feeling thermometer ratings towards Black people or White people, or by SDO. The results were different between Experiments 1 and 2 (see SM for full analysis) suggesting that, if there is such moderation, it cannot be detected reliably using the current sample sizes. We finally examined if self-identifying as Black versus not affected results. We found no significant differences in the effect between these subsamples, but again that may be related to sample size limitations and to the fact that the vast majority of the participants were White (12.7% identified as Black Experiment 1 and 9.4% in Experiment 2. see SM for full analysis).

Drift diffusion model

Drift diffusion modeling was used to further explore the observed effects of race on crowd-amplification. However, it's important to note that this analysis was not preregistered and was conceived after the data was collected. In the following analysis we first examine the association between race proportion and starting-point bias, then examine the association with sensitivity (drift rate). We pooled the data provided for these models from Experiments 1 and 2 because their designs were identical and results comparable.

Starting-point/bias We first examined bias as a function of valence. Bayesian equivalence tests indicated that participants were more biased initially toward emotional inferences on neutral-to-angry arrays than on neutral-to-happy arrays. This effect was consistent across all target-race ratio conditions; $z_{25\%Black}$ mean = 0.038, 95% HDI = [0.025, 0.052]; $z_{50\%Black}$ mean = 0.035, 95% HDI = [0.022, 0.048]; $z_{75\%Black}$ mean = 0.049, 95% HDI = [0.037, 0.062]. We did not observe any effect of race ratio on starting-point bias within any of the neutral-to-happy valenced trials. We did find that participants were more biased toward angry emotional inferences for the 75% Black crowds relative to 50% Black crowds (z_{angry} mean = 0.024, 95% HDI = [0.011, 0.037]; right-most distribution of Fig. 5A), but this was an isolated difference and did not reflect a trend across the race ratio conditions more broadly.

Drift rate/sensitivity Next, we examined differences in drift rate as a function of crowd valence. Results indicated that drift rate was greater when judging neutral-to-angry arrays than neutral-to-happy arrays, suggesting that participants accumulated evidence more rapidly when crowd valence was more negative than positive. This difference was present across all race ratio conditions; $\delta_{25\%Black}$ mean = 0.43, 95% HDI = [0.36, 0.50]; $\delta_{50\%Black}$ mean = 0.40, 95% HDI = [0.32, 0.48]; $\delta_{75\%Black}$ mean = 0.41, 95% HDI = [0.33, 0.48].

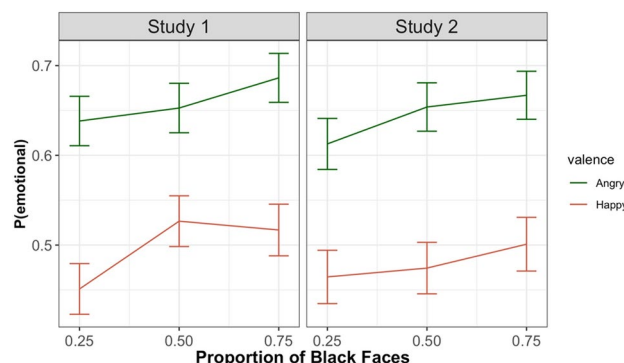


Fig. 4. Results from Experiments 1 ($n=141$) and 2 ($n=139$). The x axis represents the ratio of Black faces within the crowds. The y axis represents the probability of evaluating the crowd as emotional. Results in both experiments indicate two main findings. The first is that crowds expressing anger were more likely to be evaluated as emotional. The second is that increase in the ratio of Black faces led to an increase in the probability of evaluating the crowd as emotional. Error bars represent 95% confidence intervals.

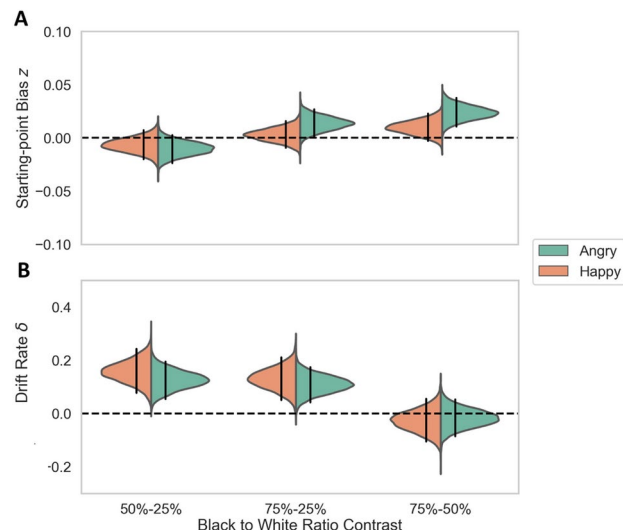


Fig. 5. Posterior predictive density plots for Experiment 1–2 simple effects of race ratio on starting-point bias (A) and drift rate (B) for happy and angry crowd-emotion valences. Vertical black bars are the 95% most credible values for the mean difference between race conditions. With the exception of the contrast between the angry valence 75 and 50% Black trials, all estimates of starting-point bias contained the null-point value, suggesting that starting-point bias did not differ across race proportions. Conversely, drift rates were notably larger than the null-point when comparing the 50 and 75% Black trials to the 25% Black trials, suggesting that evidence accumulated faster toward emotional inferences when crowds were composed of 50% or more Black individuals.

As for the evaluation of sensitivity as a function of race, we observed that target race proportions *did* have an effect on drift rate, particularly where the 25% Black trials were compared to the 50% Black trials (δ_{happy} mean = 0.16, 95% HDI = [0.077, 0.24]; δ_{angry} mean = 0.13, 95% HDI = [0.054, 0.19]) and to the 75% Black trials (δ_{happy} mean = 0.13, 95% HDI = [0.05, 0.21]; δ_{angry} mean = 0.11, 95% HDI = [0.041, 0.17]; Fig. 5B). However, no substantial difference could be detected between the 50% and the 75% Black trials (δ_{happy} mean = -0.027, 95% HDI = [-0.11, 0.055]; δ_{angry} mean = -0.017, 95% HDI = [-0.086, 0.053]). This suggested that evidence accumulated faster for emotional inferences where there were a greater proportion of Black faces relative to White faces, but that this effect saturated at higher proportions of Black faces. Estimated means and highest density intervals for all Experiment 1 and 2 model parameters are reported in (Supplementary Table S5).

Experiment 3: adding all-black and all-white crowds

One question that is raised by the previous findings is whether crowds including more Black faces were more likely to be evaluated as emotional due to the contrast between Black and White faces. In Experiment 3 we examined this question by adding 0% Black target and 100% Black target conditions to the existing design.

Method

Participants

In deciding our sample size for the experiment, we wanted to make sure that the sample size for the mix crowds was kept similar to that of Experiments 1 and 2. Given that we introduced two new conditions (all-White and all-Black), we increased the sample size to 225. Participants completed the task on Mechanical Turk in exchange to \$3. All of our participants were Americans who spoke English as their first language. Out of the 225 that we collected, 222 participants completed the full task. Similar to Experiments 1 and 2, we applied the selection criteria of removing participants who made the same emotionality choice for 45 of the 50 trials. Our final sample was 209 (men: 116, women: 92, other: 1; age: M = 41.60, SD = 12.60). As in Experiments 1 and 2, we decided not to limit recruitment to monoracial White participants. Participants identified as follows: Black = 31, Hispanic/Latin = 9, White = 166, Asian = 10, other = 5.

Procedure

Ethics committee approval was obtained prior to the data collection. The experiment was identical to Experiments 1 and 2 with one difference: in addition to seeing crowds of Black and White face arrays, participants also saw crowds of all-Black and all-White faces. All of the crowds included 2 face identities to keep the structure as similar as possible to the previous experiments (two Black: 100% Black, one White and one Black where proportion Black was 25, 50, or 75%, and two White: 0% Black).

Results and discussion

Hypothesis 1: evaluation of emotionality

Our first hypothesis was the arrays in general will be more likely to be evaluated as emotional than non-emotional. In order to examine this hypothesis, we tested the association between the average intensity of the emotion expressed in the crowd and the probability of evaluating that crowd as emotional. This model allowed us both examine the general probability of evaluating arrays as emotional at an average intensity (intercept of the model) as well as to look at the general connection between the array intensity and the probability of evaluating the array as emotional (slope—manipulation check). We used the same model as in Experiment 1. Results indicated that the intercept of the model was significantly positive ($b = 0.33$ [0.12, 0.54], $SE = 0.09$, $z = 3.48$, $p < 0.001$, $R^2c = 0.20$). Results also indicated a very strong association between the actual intensity of the array and the probability that it would be categorized as emotional ($b = 0.65$, [0.60, 0.70], $SE = 0.02$, $z = 27.24$, $p < 0.001$, $R^2c = 0.20$).

Model specification

Next, we tested all other hypotheses in a second model by conducting a three-way interaction between size, valence, and proportion of Black faces in the crowd. The model did show a weak three-way interaction ($b = -0.08$ [-0.17, 0.01], $SE = 0.04$, $z = -2.01$, $p = 0.04$, $R^2c = 0.16$), but no two-way interactions (see full details of the model in SM) suggesting that both array size and valence did not interact with the proportion of Black faces in the crowd.

Hypothesis 2: crowd size

We then turned to test the main effects of the model. First, we tested whether bigger crowds were likely to be evaluated as emotional. Results indicated that the association between array size and the probability of evaluating the crowd as emotional was marginally significant ($b = 0.05$ [-0.004, 0.11], $SE = 0.02$, $z = 1.79$, $p = 0.07$, $R^2c = 0.16$), which is more or less similar to Experiments 1 and 2 and may be driven by the low variance of face size, and the fact that it was only 4,8 or 12 faces.

Hypothesis 3: crowd valence

Next, we examined whether valence was associated with increased likelihood of evaluating the crowd as emotional. Results suggested that crowds expressing anger were more likely to be evaluated as emotional ($b = 1.04$ [0.95, 1.12], $SE = 0.04$, $z = 24.10$, $p < 0.001$, $R^2c = 0.16$).

Hypothesis 4: perceived race composition

Finally, we examined whether the proportion of Black faces in the crowd was associated with increased likelihood of evaluating the crowd as emotional. Unlike the previous model which had only mixed arrays of White and Black faces, the current model had 5 levels of proportion of Black faces in the crowd, from 0 to 100%, incrementing by 25%. As predicted, more Black faces in the crowd were associated with increased probability in evaluating the crowd as emotional ($b = 0.16$ [0.07, 0.24], $SE = 0.04$, $z = 3.81$, $p < 0.001$, $R^2c = 0.16$, see Fig. 6).

To learn more about the differences between our all-Black, mixed, and all-White conditions we conducted an additional model in which we compared the all-Black condition to the three mixed and all-White conditions in predicting emotionality evaluation. The all-Black condition led to a significantly higher probability of evaluating the crowd as emotional compared to both the mixed conditions ($b = -0.17$ [-0.28, -0.06], $SE = 0.05$, $z = -3.05$, $p = 0.002$, $R^2c = 0.08$) and the all-White conditions ($b = -0.41$ [-0.55, -0.27], $SE = 0.07$, $z = -5.77$, $p < 0.001$, $R^2c = 0.08$). We then reversed the order of the model to compare the all-White condition to the mixed conditions. People were more likely to evaluate the mixed-race crowd as emotional compared to the all-White crowd ($b = 0.23$ [0.12, 0.34], $SE = 0.05$, $z = 4.25$, $p < 0.002$, $R^2c = 0.08$). These results provide additional support to the notion that the occurrence of Black faces in the crowd contributes further to the tendency to evaluate the crowd as emotional.

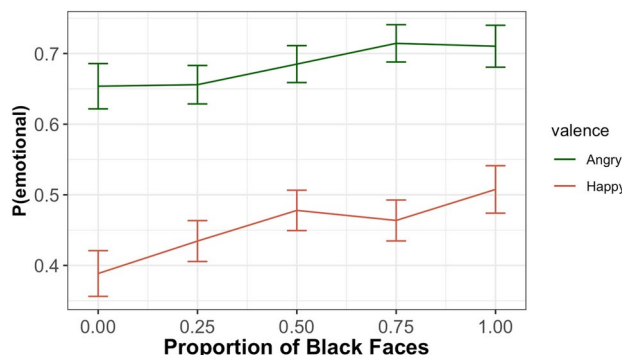


Fig. 6. Results from Experiment 3 ($n = 209$). The x axis represents the ratio of Black faces within the crowds. The y axis represents the probability of evaluating the crowd as emotional. Results in both experiments indicate to main findings. The first is the crowd's expressing anger were more likely to be evaluated as emotional. the second is that increase in the ratio of Black faces led to an increase in the probability of evaluating the crowd as emotional. error bars represent 95% confidence intervals.

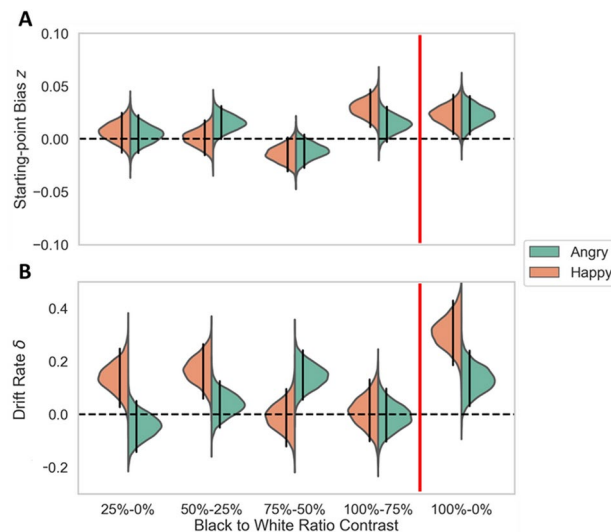


Fig. 7. Posterior predictive density plots for Experiment 3 simple effects of race ratio on starting-point bias (A) and drift rate (B) for Happy and Angry crowd-emotion valences. Vertical black bars are the 95% most credible values for the mean difference between race conditions. With the exception of the contrast between the Happy valence 100 and 75% Black trials, all estimates of starting-point bias contained the null-point value, suggesting that starting-point bias did not differ linearly across race proportions. That said, a small global effect of race on starting-point bias was observed when contrasting the 100 and 0% Black trials. Drift rate steadily increased with Black to White ratio for Happy valence trials, up to Black majority where the effect saturated. For Angry valence trials, drift rate became more positive after reaching Black majority, but not before.

Drift diffusion model

Starting-point bias As with experiments 1–2, we observed similar estimates of the bias parameter in Experiment 3. There was a strong effect of valence on starting-point bias, such that participants were more biased to judge neutral-to-angry arrays as emotional than they were neutral-to-happy arrays, irrespective of target-race proportions; $z_{0\%Black}$ mean = 0.031, 95% HDI = [0.010, 0.050]; $z_{25\%Black}$ mean = 0.043, 95% HDI = [0.027, 0.058]; $z_{50\%Black}$ mean = 0.030, 95% HDI = [0.013, 0.048]; $z_{75\%Black}$ mean = 0.028, 95% HDI = [0.011, 0.046]; $z_{100\%Black}$ mean = 0.046, 95% HDI = [0.028, 0.061]. Similar to Experiments 1 and 2, race ratio had no effect on bias, with the exception of 75–100% Black contrast for neutral-to-happy arrays (z_{happy} mean = 0.029, 95% HDI = [0.011, 0.047]; second to the rightmost distribution of Fig. 7A). However, we also observed a small global effect of race on starting-point bias when contrasting 100 and 0% Black trials (z_{happy} mean = 0.023, 95% HDI = [0.004, 0.042]; z_{angry} mean = 0.022, 95% HDI = [0.004, 0.040]).

Drift rate/sensitivity Finally, we estimated drift rates for valence and target-race proportions from Experiment 3. Similar to the estimates from Experiments 1–2, we found a strong effect of valence on drift rate across all five race conditions, suggesting that participants accumulated evidence more rapidly when crowd valence was more negative than positive: $\delta_{0\%Black}$ mean = 0.75, 95% HDI = [0.64, 0.86]; $\delta_{25\%Black}$ mean = 0.57, 95% HDI = [0.47, 0.66]; $\delta_{50\%Black}$ mean = 0.44, 95% HDI = [0.34, 0.53]; $\delta_{75\%Black}$ mean = 0.59, 95% HDI = [0.48, 0.69]; $\delta_{100\%Black}$ mean = 0.57, 95% HDI = [0.45, 0.68].

More important, we also found an effect of race on drift rate that saturated with higher percentages of Black faces in the neutral-to-happy arrays, such that drift rate steadily increased toward emotion as more Black faces were added to majority White crowds, up to 50% ($\delta_{happy, 25\%-0\%Black}$ mean = 0.14, 95% HDI = [0.027, 0.25]; $\delta_{happy, 50\%-25\%Black}$ mean = 0.17, 95% HDI = [0.059, 0.26]; Fig. 7B). This effect saturated as Black faces became the majority ($\delta_{happy, 75\%-50\%Black}$ mean = -0.008, 95% HDI = [-0.12, 0.096]; $\delta_{happy, 100\%-75\%Black}$ mean = 0.011, 95% HDI = [-0.10, 0.13]). In fact, participants drifted toward non-emotional inferences when neutral-to-happy arrays were majority White – this became more balanced as the Black-to-White ratio increased.

Unlike experiments 1–2, however, drift rate was less linearly related to race for the neutral-to-angry arrays. Here, drift rate remained unchanged as the proportion of Black faces increased in majority White crowds, but then became much more positive in majority Black crowds ($\delta_{angry, 75\%-50\%Black}$ mean = 0.15, 95% HDI = [0.055, 0.24]). This deviates from the findings in Experiments 1 and 2, in that the relationship between race and emotion perception appears less linear with saturation and more stepwise with the indicator being race majority. When limiting the analysis to only 100% and 0% Black trials, a small global effect of race was observed on angry valence trials ($\delta_{angry, 100\%-0\%Black}$ mean = 0.13, 95% HDI = [0.03, 0.24]), as well as a much stronger global effect on happy valence trials ($\delta_{happy, 100\%-0\%Black}$ mean = 0.31, 95% HDI = [0.19, 0.43]). Estimated means and highest density intervals for all Experiment 3 model parameters are reported in (Supplementary Table S6).

In sum, Bayesian equivalence tests revealed that both drift rate and starting-point bias affected ‘emotional’ judgments when crowds expressed neutral-to-angry expressions. However, only drift rate was influenced by the

ratio of Black to White faces in those crowds. Drift rate was faster toward 'emotional' judgments when there were 50% or more Black faces in the crowd.

General discussion

In three experiments we found that increases in the proportion of Black faces in the crowd were associated with increased likelihood of evaluating the crowd as emotional. Results of Experiment 3 suggested that this effect was not driven by the contrast between White and Black faces. Analysis using a hierarchical DDM suggested that this was driven by increased sensitivity, such that every emotional Black face counted more than equally emotional White faces towards a faster decision.

These findings contribute to a broader understanding of social cognition as a function of racialization. Previous work has generally examined the evaluation of single faces^{14,17,20,34,36–38}. There is a growing recognition of the importance of aggregating social information when looking at crowds or groups of people^{8–11,27}. However, existing studies have not yet examined how emotions impact perceptions of crowd emotionality. The present experiments advance understanding of the attentional dynamics associated with racialization in social perception; they reveal that differences in information encoding and averaging occur when participants can choose to which targets they wish to attend. Our diffusion modeling results further elucidate how crowd emotional evaluations are made, which is with greater sensitivity to increasingly Black crowds. This sensitivity increases non-linearly with the number of Black persons in the crowd. Across our experiments, sensitivity shifted at the point of Black majority; in lieu of perceiving emotionality on a continuum, participants instead became more sensitive to emotion expression categorically, up to or after the point of Black majority. This is critically different from an effect of starting-point bias, which would instead reflect a bias for emotionality prior to any decision-making process. Sensitivity, or drift rate, notably reflects the passing of associative memory content into decision systems²³, so our findings may shed light on an overgeneralization of existing racial biases about individuals onto crowds.

We also found that crowds expressing anger were more likely to be evaluated as emotional compared to crowds expressing happiness. However, we did not find an interaction between emotion and the proportion of Black faces in the crowds; sensitivity to Black faces was equal in evaluating crowds expressing anger and crowds expressing happiness. These findings may seem to contradict literature suggesting that perceivers are more likely to categorize racially ambiguous faces expressing anger as Black compared to White^{12,15}, or that latency time in identifying emotions may be impacted by race^{17–19}. However, these previous findings have compared the expression of single faces and did not directly examine whether, in a mixed crowd, Black faces were more likely to be attended to when expressing emotions. The current findings suggest that there is a sensitivity to Black faces expressing both positive and negative emotions, which advances our understanding of how mixed-race crowds are evaluated.

Our findings also have important implications for how social gatherings, demonstrations, and social movements are evaluated based on if they are predominantly populated by Black versus White individuals. Previous work suggests that ethnic minorities striving toward political change are perceived as more violent and that violence seems to be less useful in gaining public support^{39–41}. The current results reveal an important potential mechanism for this differential treatment, which is the evaluation of emotionality. If police officers and the general public are more likely to perceive demonstrators as emotional when the majority of demonstrators are Black, they may be more likely to escalate conflict, or be more supportive of violent counter-protest tactics.

Limitations and future directions

The current experiments revealed a tendency for perceiving emotionality as a function of a crowd's racial composition. Nevertheless, these experiments have several notable limitations and leave open questions regarding how these effects translate to behavior outside the lab.

The first major limitation relates to the face stimuli we used in these experiments. One corresponding question is whether our findings generalize despite the fact that participants viewed expression arrays that were all derived from four faces. In real life, people analyze expressions of many different individuals when estimating the emotions of crowds. We chose to use as few face identities as possible to ensure that attentional biases are not driven by the specific faces in the crowd, and to ensure that the evaluation of intensity was equal between the White and Black faces. However, future work should examine the evaluation of crowds' emotions based on race in more naturalistic settings. Furthermore, as noted in the introduction, the crowds in the present experiments were all made up of male faces. Recent empirical work suggests that people are more likely to evaluate sequences of emotions expressed by women as emotional compared to those expressed by men⁴². It is therefore possible that the current results could be modified as a function of gender, such that increasing the proportion of women should increase the tendency to evaluate crowds as emotional. However, we do not have specific predictions regarding a race by gender interaction as we have already noted that findings from individual face studies may not extend in an additive manner to crowd studies. And finally, crowds' average emotions may not be normally distributed in actuality. It could be that case that a small group of people expresses a lot of emotion, whereas others in the crowd do not. Subsequent work should examine how different emotional distributions affect emotion evaluation. For example, recent work suggests that people tend to both discount² and overweight outliers⁴³ depending on their degree of extremity. It would be interesting to test whether racial composition may moderate these effects.

A second limitation is that crowd emotionality was measured using a binary outcome, whereas in the real world, emotional judgments may be continuous. We cannot be certain that the experimental requirement to convert perceptual summaries into binary decisions did not affect the results. We chose to use binary judgments both to try to mirror the structure of decision-making surrounding crowd emotionality in the real world (i.e., in which observers have to reach some threshold in their judgments to change their behavior, such stay near the

crowd or leave to seek safety) and to allow the application of a drift diffusion model (DDM). However, future research is required to extend our understanding of how measurement may affect the results.

A third limitation of the experiment relates to how these results may differ among people with different social preferences, races, personalities and skills. Differences in prejudice and political affiliation were not consistently associated with judgments of crowd emotionality. However, it is possible that increasing the diversity of the sample could reveal some of these effects. Future research should examine these moderating variables more closely.

To conclude, this project extends recent efforts to examine the role of ensemble coding in processes social decision-making. Learning more about how people rapidly evaluate complex social information may not only explain important aspects of social behavior, but may also highlight specific channels for future interventions.

Data availability

Code and data for study 1 is available here: <https://osf.io/xhpi7/>. Code and data for Study 2 is available here: <https://osf.io/38xhg/>. Code and data for Study 3 is available here: <https://osf.io/7rn2q/>.

Received: 25 October 2023; Accepted: 17 March 2025

Published online: 08 October 2025

References

- Wolfe, B. A., Kosovicheva, A. A., Wood, K. & Whitney, D. Foveal input is not required for perception of crowd facial expression. *J. Vis.* **15**, 11–11 (2015).
- Haberman, J. & Whitney, D. The visual system discounts emotional deviants when extracting average expression. *Atten. Percept. Psychophys.* **7**, 1825–1838 (2010).
- Haberman, J. & Whitney, D. Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* **17**, R751–R753 (2007).
- Elias, E., Dyer, M. & Sweeny, T. D. Ensemble perception of dynamic emotional groups. *Psychol. Sci.* **28**, 193–203 (2017).
- Whitney, D. & Yamanashi Leib, A. Ensemble perception. *Annu. Rev. Psychol.* **69**, 105–129 (2018).
- Goldenberg, A., Weisz, E., Sweeny, T., Cikara, M. & Gross, J. J. The crowd emotion amplification effect. *Psychol. Sci.* **32**, 437–450 (2021).
- 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.* <https://doi.org/10.1037/xge0000651> (2019).
- Alt, N. P. & Phillips, L. T. Person perception, meet people perception: Exploring the social vision of groups. *Perspect. Psychol. Sci.* **17**, 768–787 (2022).
- Kardosh, R., Sklar, A. Y., Goldstein, A., Pertzov, Y. & Hassin, R. R. Minority salience and the overestimation of individuals from minority groups in perception and memory. *Proceedings of the National Academy of Sciences of the United States of America* **119**, (2022).
- Lamer, S. A., Sweeny, T. D., Dyer, M. L. & Weisbuch, M. Rapid visual perception of interracial crowds: Racial category learning from emotional segregation. *J. Exp. Psychol. Gen.* **147**, 683–701 (2018).
- Phillips, L. T., Slepian, M. L. & Hughes, B. L. Perceiving groups: The people perception of diversity and hierarchy. *J. Pers. Soc. Psychol.* **114**, 766–785 (2018).
- Hugenberg, K. & Bodenhausen, G. V. Ambiguity in social categorization: The role of prejudice and facial affect in race categorization. *Psychol. Sci.* **15**, 342–345 (2004).
- Correll, J., Wittenbrink, B., Crawford, M. T. & Sadler, M. S. Stereotypic vision: How stereotypes disambiguate visual stimuli. *J. Pers. Soc. Psychol.* **108**, 219–233 (2015).
- Duncan, B. L. Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of Blacks. *J. Pers. Soc. Psychol.* **34**, 590–598 (1976).
- Hutchings, P. B. & Haddock, G. Look black in anger: The role of implicit prejudice in the categorization and perceived emotional intensity of racially ambiguous faces. *J. Exp. Soc. Psychol.* **44**, 1418–1420 (2008).
- Hutcherson, C. A., Bushong, B. & Rangel, A. A Neurocomputational model of altruistic choice and its implications. *Neuron* **87**, 451–462 (2015).
- Hugenberg, K. Social categorization and the perception of facial affect: Target race moderates the response latency advantage for happy faces. *Emotion* **5**, 267–276 (2005).
- Bijlstra, G., Holland, R. W. & Wigboldus, D. H. J. The social face of emotion recognition: Evaluations versus stereotypes. *J. Exp. Soc. Psychol.* **46**, 657–663 (2010).
- Craig, B., Mallan, K. & Lipp, O. The effect of poser race on the happy categorization advantage depends on stimulus type, set size, and presentation duration. *Emotion (Washington, D.C.)* **12**, 1303 (2012).
- Lloyd, E. P. & Hugenberg, K. Beyond bias: response bias and interpersonal (in)sensitivity as a contributors to race disparities. *Eur. Rev. Soc. Psychol.* **32**, 201–234 (2021).
- Eberhardt, J. L., Goff, P. A., Purdie, V. J. & Davies, P. G. Seeing black: Race, crime, and visual processing. *J. Pers. Soc. Psychol.* **87**, 876–893 (2004).
- Richeson, J. A. & Trawalter, S. The threat of appearing prejudiced and race-based attentional biases: Research report. *Psychol. Sci.* **19**, 98–102 (2008).
- Ratcliff, R., Smith, P. L., Brown, S. D. & McKoon, G. Diffusion decision model: Current issues and history. *Trends Cogn. Sci.* **20**, 260–281 (2016).
- Voss, A., Rothermund, K. & Voss, J. Interpreting the parameters of the diffusion model: An empirical validation. *Mem. Cognit.* **32**, 1206–1220 (2004).
- Johnson, D. J., Hopwood, C. J., Cesario, J. & Pleskac, T. J. Advancing research on cognitive processes in social and personality psychology: A hierarchical drift diffusion model primer. *Soc. Psychol. Personal. Sci.* **8**, 413–423 (2017).
- Röhner, J. & Lai, C. K. A diffusion model approach for understanding the impact of 17 interventions on the race implicit association test. *Pers. Soc. Psychol. Bull.* **47**, 1374–1389 (2021).
- Mihalache, D. et al. General anger bias in the evaluation of crowds anger. *J. Exp. Psychol. Gen.* **150**, 1870 (2021).
- Lerche, V., Bucher, A. & Voss, A. Processing emotional expressions under fear of rejection: Findings from diffusion model analyses. *Emotion* **21**, 184–210 (2021).
- Tottenham, N. et al. The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatr. Res.* **168**, 242–249 (2009).
- Ho, A. K. et al. The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO7 scale. *J. Pers. Soc. Psychol.* **109**, 1003–1028 (2015).
- Haddock, G., Zanna, M. P. & Esses, V. M. Assessing the structure of prejudicial attitudes: The case of attitudes toward homosexuals. *J. Pers. Soc. Psychol.* **65**, 1105–1118 (1993).

32. Gosling, S. D., Rentfrow, P. J. & Swann, W. B. A very brief measure of the big-five personality domains. *J. Res. Pers.* **37**, 504–528 (2003).
33. Carpenter, B. et al. Stan: A probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).
34. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992).
35. Katahira, K. How hierarchical models improve point estimates of model parameters at the individual level. *J. Math. Psychol.* **73**, 37–58 (2016).
36. Amodio, D. M., Harmon-Jones, E. & Devine, P. G. Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *J. Pers. Soc. Psychol.* **84**, 738–753 (2003).
37. Bijlstra, G., Holland, R. W., Dotsch, R., Hugenberg, K. & Wigboldus, D. H. J. Stereotype associations and emotion recognition. *Pers. Soc. Psychol. Bull.* **40**, 567–577 (2014).
38. Plant, E. A. & Devine, P. G. Internal and external motivation to respond without prejudice. *J. Pers. Soc. Psychol.* **75**, 811–832 (1998).
39. Feinberg, M., Willer, R. & Kovacheff, C. The activist's dilemma: Extreme protest actions reduce popular support for social movements. *J. Pers. Soc. Psychol.* **119**, 1086–1111 (2020).
40. Wasow, O. Agenda seeding: How 1960s black protests moved elites, public opinion and voting. *Am. Polit. Sci. Rev.* **114**, 638–659 (2020).
41. Shuman, E., Goldenberg, A., Saguy, T., Halperin, E. & Van Zomeren, M. When are social protests effective?. *Trends Cogn. Sci.* <https://doi.org/10.1016/j.tics.2023.10.003> (2023).
42. Gorges, M., Porat, R., Schöne, J. & Goldenberg, A. Aggregating emotional sequences amplifies the perception of women as more emotional than men. <https://doi.org/10.31219/osf.io/64dru> (2024).
43. Dannals, J. E. & Miller, D. T. General social norm perception in groups with outliers social norm perception in groups with outliers. *J. Exp. Psychol. Gen.* **146**, 1342–1359 (2017).

Author contributions

AG, ZH, EW, and MC conceived and designed the experiments. AG and ZH ran the experiments. AG and ZH, and KL analyzed Studies 1–3. AG wrote the paper and KL, EW, and MC were involved in reviewing and editing the manuscript.

Declarations

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/s41598-025-95029-3>.

Correspondence and requests for materials should be addressed to A.G.

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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025