Is This My Group or Not?

The Role of Ensemble Coding of Emotional Expressions in Group Categorization

Supplementary Materials

Amit Goldenberg[1]

Timothy D. Sweeny[2]

Emmanuel Shpigel[1]

James J. Gross[1]

[1]Stanford University

[2]University of Denver

Corresponding Author: Correspondence concerning this article should be addressed to Amit Goldenberg, Department of Psychology, Stanford University. E-mail: amitgold@stanford.edu.

Study 1

Power Analysis

We conducted a power analysis using the data from the first 10 participants in our study. For these 10 participants, we compared the difference in self-categorization between the same-mean and different-mean trials (similar to the main analysis). We then used the results to conduct a simulated power analysis (1000 simulations for each sample size) for the required amount of participant using the R package simr (Green & Macleod, 2016). As reflected by Figure 1S, results from our simulation suggested that using 30 participants would lead to power well above 80%. We therefore decided to use that sample size in our experiments.

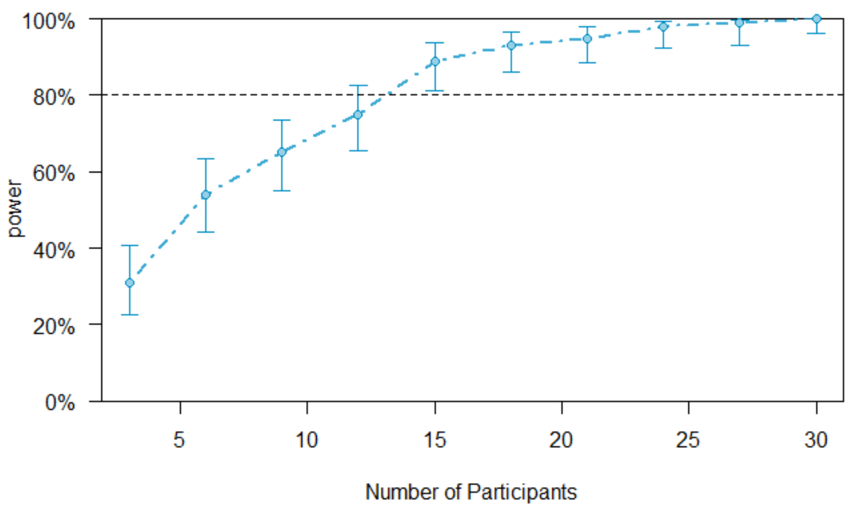

Figure 1S. Estimated power based on the number of participants using the data from the first 10 participants.

Difference in categorization as a result of participants race, gender, political affiliation and identification

Race. Participants were asked to indicate their race by choosing from five options: white, Black, Hispanic, Asian and other. Participants were allowed to mark more than one option. Out of the 30 participants who took part in the experiment, 16 indicated that they were white, six Asian, five Black, two Hispanic and one chose other. Since the faces that were presented to participants were of white males, and because the majority of participants were white (and the numbers of the other ethnic groups were relatively low) we decided to test whether self-identifying oneself as white (compared to other races) predicted participants' degree of self-categorizing. We therefore conducted a generalized mixed model analysis using participants' race (white or not white) as an independent variable and self-categorization as the dependent variable. We also used a by-participant random variable. Results suggested that being white or not did not lead to differences in self-categorization compared to other races (b =-.06 [-.58, .44], SE = .25, z = -.26, p = .79).

Gender. We conducted a generalized mixed-model analysis using participants' gender to predict self-categorization. We further used a by-participant random variable. Results suggested that the difference between males and females was marginally significant, such that males self-categorized themselves more as members of the groups compared to woman (b =.49 [-.07, 1.07], SE = .28, z = 1.75, p = .08). However, notice that similar analysis in Study 3 did not yield significant effects pointing to the fact that if such effect exist it is most likely weak.

Political affiliation. Participants' political affiliation was measured using a one item question ranging from 1 to 7, 1 indicating strongly Liberal and 7 indicating strongly Conservative. We conducted a generalized mixed-model using participants' political affiliation to predict self-categorization. We also used a by-participant random variance. Results suggested a marginally significant main effect for pollical affiliation predicting self-categorization such that more conservative participants tended to self-categorize themselves as group members slightly more (b =.20 [-.04, .44], SE = .12, z = -1.67, p = .09). However, notice that we were not able to replicate this finding in Study 3.

Group identification. Group identification was measured using a five-item scale adapted from work by Roccas and colleagues (Roccas, Klar, & Liviatan, 2006). The items were: '*Being an American is an important part of who I am', 'Its important to me to view myself as an American', 'When I talk about Americans, I usually say we rather than they', 'I am strongly committed to my nation', 'I love the United States.'* (α=.93[.89,.97]). We conducted genialized mixed-model analysis looking at participants' degree of identification as predicting participants' self-categorization. Results suggested that self-categorization was higher for participants who identified more with the United States (b =.29 [.07, .52], SE = .11, z = 2.67, p = .01). These results point to the fact that high identifiers were more open self-categorize people as their group, regardless of the group mean. However, notice that results were in the opposite direction in Study 3.

Self-categorization compared to chance

We conducted a generalized mixed-model in order to compare both of our conditions (same and different group mean) to chance. Results suggested that participants' self-categorization in both the same mean condition (b =1.24 [.94, 1.55], SE = .15, z =8.20, p < .001) and the different mean condition (b =.55 [.28, .83], SE = .13, z =4.03, p < .001) were significantly higher than chance.

Differences between the higher and the lower group mean trials

In order to test differences between the higher and the lower mean group emotion trials we created an orthogonal contrast, first comparing the similar group emotion trials to the two other different group emotion trials (higher and lower), and then comparing the higher and the lower group mean emotion trials to each other. We conducted a generalized mixed model analysis using our orthogonal contrast an independent variable and participants self-categorization as a dependent variable. In addition, we used a by-participant random variable. Looking specifically on the comparison between the higher and the lower trials, results suggested no significant difference between these two conditions (b =-.03 [-.19, .12], SE = .08, z = -.40, p = .69, see Figure 2S).

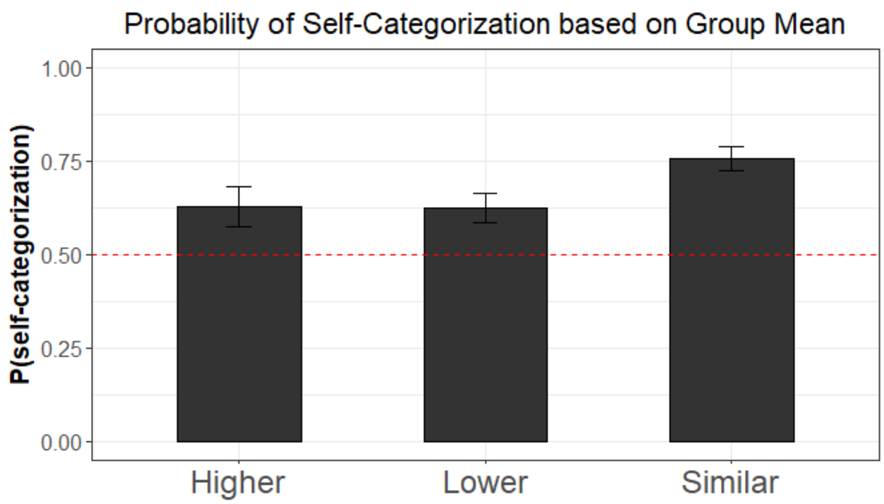

Figure 2S. Participants' self-categorization as a function of whether the mean group emotion was higher, lower or similar to participants' own rating. Error bars represent 95% confidence intervals.

Study 2

Difference in mean evaluation as a result of participants race, gender, political affiliation and group identification

Race. Similar to Study 1, we tested whether self-categorizing oneself as white lead to changes in participants' estimation of the mean group emotion. We conducted a mixed-model analysis looking at whether participants' race (white or not white) predicted participants' evaluation of the mean group emotion. We also used a by-participant random variable. Results suggested that being white did not lead to differences in evaluations of group mean (b =.75 [-1.28, 2.78], SE = 1.04, $t(28)$ = .72, p = .47).

Gender. We conducted a mixed-model analysis using participants' gender to predict participants' estimation of the mean group emotion. We further used a by-participant random variable. Results suggested that the difference between males and females did not lead to differences in participants' evaluation of the mean group emotion (b =-*.33* [-2.39, 1.71], SE = 1.05, t(30) = -.32, p = .75).

Political affiliation. Participants' political affiliation was measured using a one item question ranging from 1 to 7, 1 indicating strongly Liberal and 7 indicating strongly Conservative. We tested whether participants political affiliation predicted the evaluation of the mean by conducting a mixed-model analysis, further using a by-participant random variable. Results suggested that participants' political affiliation did not predict the evaluation of the group mean emotion (b =-.62 [-2.07, .82], SE = .74, t(28) = -.84, p = .40).

Group identification. We used the same scale as in Study 1 to test whether participants' group identification predicted their estimation of the mean (α = .83 [.74, .93]). We conducted a linear mixed-model analysis, predicting participants' evaluation of the mean from their degree of identification. Results suggested no association between the two (b =.38 [-.66, 1.44], SE = .53, t(28) = .72, p = .47).

Study 3

Difference in categorization as a result of participants race or gender

Race. Similar to Study 1, we tested whether self-categorizing oneself as white led to changes in participants' self-categorization. We conducted a generalized mixed model analysis looking at whether participants' race (white or not white) predicted self-categorization. We also used a by-participant random variable. Results suggested that being white (compared to other races) did not lead to differences in self-categorization (b =-.36 [-.97, .21], SE = .28, z = -1.24, p = .21).

Gender. We conducted a generalized mixed-model analysis using participants' gender to predict self-categorization. We further, used a by-participant random variable. Results suggested no significant difference between males and females in degree of self-categorization (b =.15 [-.43, .75], SE = .28, z = .54, p = .58).

Political affiliation. We tested whether participants political affiliation predicted self-categorization by conducting a generalized mixed-model analysis. We further used a by participant random variable. Results suggested that participants' political affiliation did not predict participants' self-categorization (b =-.002 [-.28, .27], SE = .13, z = -.02, p = .98).

Group identification. We used the same scale as in Study 1 to test whether participants' group identification predicted self-categorization (α = .91 [.86, .96]). We conducted a generlized mixed-model analysis, predicting participants' evaluation of the mean from their degree of identification. Results suggested that participants self-categorization was high for participants who identified less with the United States (b =-.30 [-.60, -.03], SE = .13, z = -2.21, p = .03). Notice that these results are in the opposite direction to the findings of Study 1.

Self-categorization compared to chance

We conducted a generalized mixed-model analysis to compare participants' decision of self-categorization in all of the four conditions (same mean high variance, same mean low variance, different mean high variance and different mean low variance) to chance. Results suggested that when the mean was similar to participants' own ratings, the choice in both the high standard deviation trials (b =1.21 [.84, 1.61], SE = .19, z =6.43, $p < .001$) and the low standard deviation trials (*b* =1.90 [1.50, 2.32], *SE* = .20, *z* = 9,21, *p* < .001) was positive and higher than chance. When the mean group emotion was different from participants' own rating, self-categorization was positive and different from chance only in the low standard deviation trials (*b* =.74 [.39, 1.10], *SE* = .17, *z* =4.18, *p* < .001). However, in the high standard deviation condition, the difference was non-significantly different from chance (*b* =.19 [-.14, .54], *SE* = .17, *z* =1.16, *p* = .24). These results support the idea that participants were not able to self-categorize themselves during the different mean, high variance trials.

Differences between the higher and the lower group mean trials

In order to test differences between the higher and the lower mean group emotion trials we created an orthogonal contrast, first comparing the similar mean group emotion trials to the two other different mean group emotion trials, and then later comparing the higher and the lower mean group emotion trials to each other. We conducted a generalized mixed model analysis using our orthogonal contrast an independent variable and participants self-categorization as a dependent variable. In addition, we used a by-participant random variable. Looking specifically on the comparison between the higher and the lower trials, results suggested no significant difference between these two conditions (*b* =-.14 [-.31, .02], *SE* = .04, *z* = -1.62, *p* = .10, see Figure 3S) suggesting no difference in categorization between these two conditions.

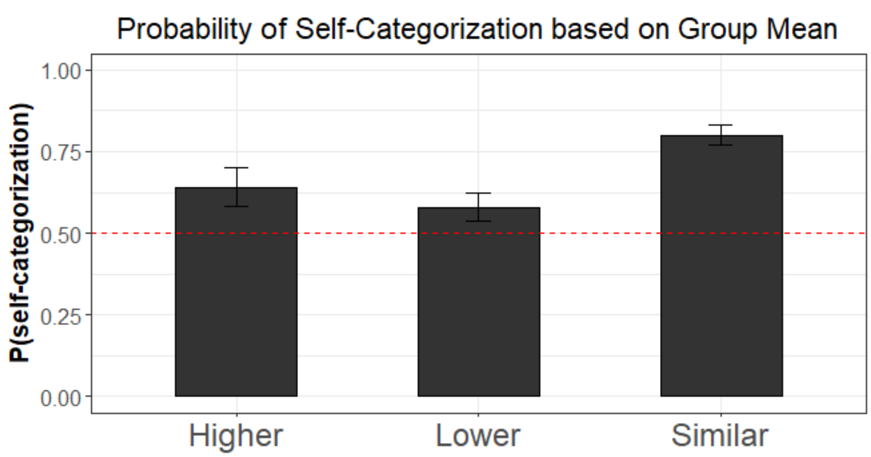

Figure 3S. Participants' self-categorization as a function of whether the mean group emotion was higher, lower or similar to participants' own rating. Error bars represent 95% confidence intervals.

Study 3a

Online Replication of Study 3

The goal this study was to create an online replication of the effects found in Study 3.

Method

Participants. We recruited 100 participants who completed the study online. Sample size was set to 100 participants, much larger than in study 3, under the assumption that effect sizes would be smaller in an online study. Participants were all college students either at Stanford or Foothill College who completed the task in exchange for a course credit. The study included a set of attention checks in which participants were asked to describe the content of the pictures that they responded to. Out of the 100 participants who completed the task, we removed 4 participants for describing the pictures incorrectly. Our final sample was therefore 96 participants (Males: 32, Females: 62; Age: M= 21.37, SD = 6.02).

Procedure. The study procedure was a direct replication of Study 3 with one addition: every five trials participants were asked to describe the picture they saw in the last trial. This request was designed in order to make sure that participants were indeed paying attention to the pictures.

Results and Discussion

We conducted a mixed generalized linear model looking at the interaction between the group mean (similar versus different) and the group standard deviation (low versus high) predicting participants' self-categorization decision. We also used a by-participant random variable. Looking first at the interaction, results suggested that the interaction was significant such that the difference between lower and higher standard deviation was different between the same and different group emotion conditions (*b* =.09 [.02, .16], *SE* = .03, *z* =2.77, *p* = .01, R^2 =.18). These results were different from those of Study 3 in which an interaction was not found. We further examined the main and simple effects (also see Figure 4S).

Looking first at the main and simple effects of the group mean, results show that self-categorization was higher when the group mean was similar to that of participants, compared to when the group mean was different (*b* = -.56 [-.63, -.50], *SE* = .03, *z* = -16.47, *p* < .001, R^2= *.17*). These results replicate our findings from Studies 1 and 3. The significant difference between the same mean and different mean conditions was similar in the low standard deviation group (*b* = -.66 [-.77, -.56], *SE* = .05, *z* = -13.06, *p* < .001, R^2= *.22*) and the high standard deviation group (*b* = -.47 [-.56, -.38], *SE* = .06, *z* = -10.24, *p* < .001, R^2= *.13*).

Next, we examined the main effect of standard deviation on categorization decisions. We found a main effect for standard deviation such that self-categorization was lower for groups with a high standard deviation of emotional intensity regardless of the group mean (*b* = -.19 [-.26, -.12], *SE* = .03, *z* = -5.75, *p* < .001, R^2= *.09*). These results were similar to those of Study 3. We further examined the simple effects of the differences between the low and high standard deviation conditions for both the similar and different-mean conditions. For the same-mean condition, self-categorization was higher for the low standard deviation trials compared to the high standard deviation trials (*b* = -.29 [-.41, -.19], *SE* = .05, *z* = -5.33, *p* < .001, R^2= *.18*). These results were congruent with our second hypothesis and similar to those found in Study 3. Importantly, results of the different group emotions trials showed that self-categorization was higher in the low standard deviation trials as well, suggesting that even in the different-mean emotion trials, participants preferred to self-categorize themselves as members of more homogeneous groups (*b* = -.10 [-.18, -.02], *SE* = .04, *z* = -2.45, *p* = .01, R^2= .09). These results were similar to those in Study 3 as well.

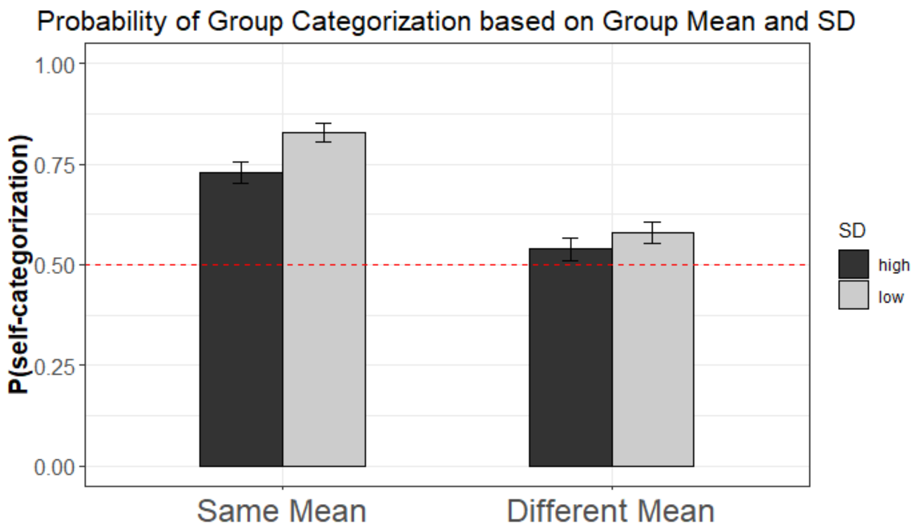

Figure 4S. Self-categorization based on both mean and standard deviation manipulation in Study 3A. Error bars represent 95% confidence intervals.

Overall, results of Study 3A replicate the findings of Study 3 and provide further evidence that a group's emotional variance can have an important and surprising influence on a viewer's self-categorization decisions. Participants preferred to self-categorize into groups with low emotional variance, both when the group's average emotional intensity was similar or different from participants' own emotions.

Study 4

Pre-test 1: Creating the Stimulus Set

We conducted a pre-test before running the actual study in order to generate a stimulus set of pictures depicting happy group-related situations, which we expected would elicit positive emotions like happiness.

We recruited 40 participants who completed the study online on Mechanical Turk (23 males, 17 females, Age: *M* = 38.2, *SD* = 11.30). Participants were all Americans located in the

United States who completed the task on a computer (rather than a smartphone). Selection criteria was similar to previous studies. Participants' received $1.5 for their participation (average time 16.7 minutes). Participants viewed 30 pictures each that were randomly chosen from 70 pictures, with the intention of choosing 50 pictures from this set. These were pictures of people celebrating the 4th of July and of American athletes celebrating having received Olympic medals or wining international competitions. Participants saw each pictures in a random order for first seconds and were then asked to answer a few questions (for the task, see https://github.com/chagag/piloting_emotions). First, participants were asked whether the emotions that were experienced in response to the pictures were negatively valanced (negative feeling) or positively valanced (positive feeling) or natural. Second, participants were asked to indicate their emotional intensity in response to these pictures on a 1-100 sliding scale, 1 indicating no emotional intensity and 100 indicating strong emotional intensity. The valance (positive or negative) and intensity measures were taken separately so as to match with the actual task. Finally, participants were asked to evaluate their distinct emotions in response to the pictures from a list of eight distinct emotional states (Happiness, Excitement, Calmness, Sadness, Anger, Guilt, Fear, Hatred). This was done in order to examine the distinct emotions elicited by the pictures.

Positive versus negative. We first examined whether the pictures were evaluated as positive, negative or neutral. For each picture we calculated the percentage of participants who evaluated the pictures as negative (100% indicating that all participants evaluated the picture as negative). Figure 5S is a histogram of the percent of positive, negative and neutral evaluations for all pictures. As the figure suggests, the majority of the pictures were classified as negative by more than 85% of participants (M = .85, SD = .11) with the majority of the rest categorized as neutral. Very few pictures were evaluated as eliciting negative emotions. Pictures that were evaluated by less than 85% of participants as negative removed from the actual study.

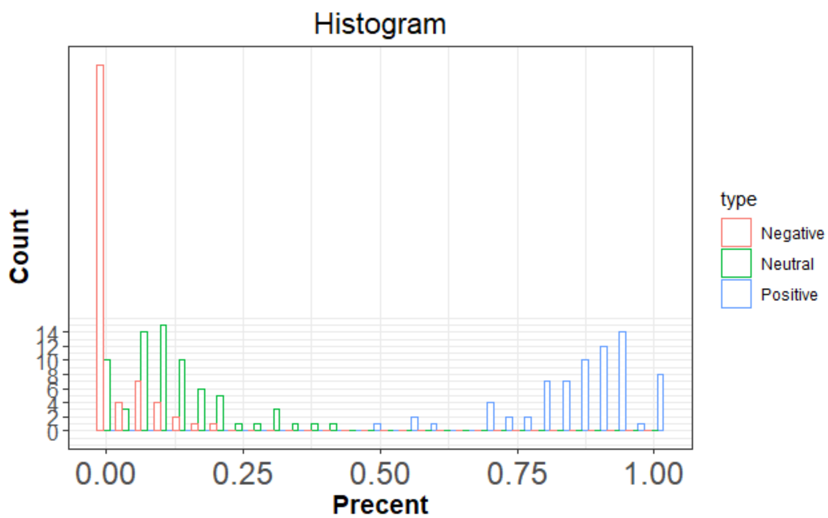

Figure 5S. The percentage of participants who evaluated each picture as positive. As this figure suggests, the vast majority of pictures were evaluated as positive by more than 85% of participants, other pictures were removed from the set.

Discrete emotions. Next, we examined participants' discrete emotions elicited by the pictures. In general participants mainly evaluated the pictures as eliciting positive emotions, predominantly happiness, excitement and calmness (figure 6S). This was expected, considering the nature of the pictures, and indicated that the stimulus set was appropriate to the goals of the study. We removed pictures that elicited negative emotions.

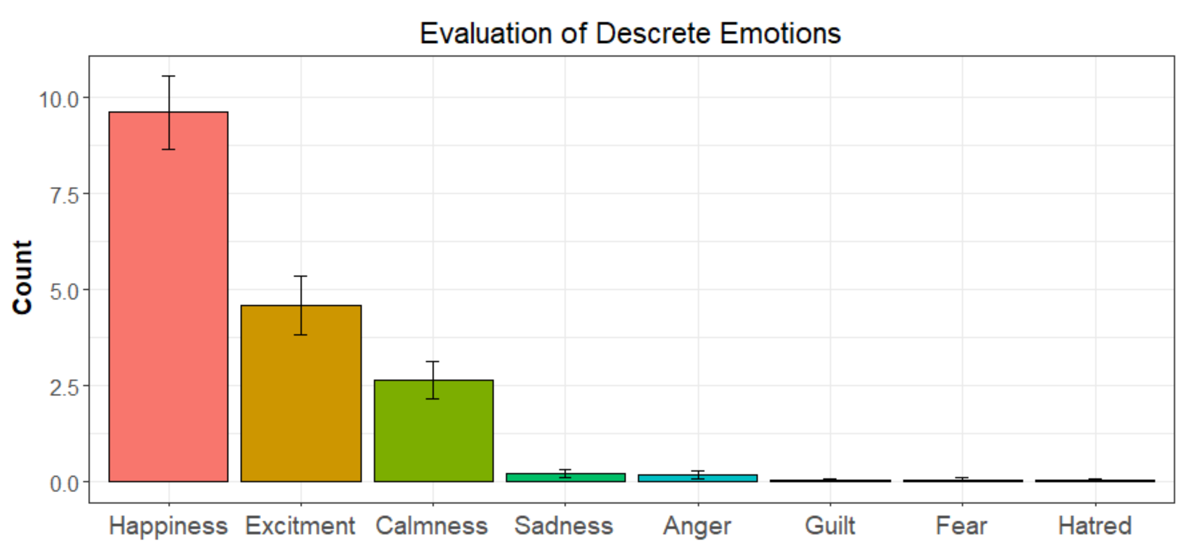

Figure 6S. Count of discrete emotions for the stimulus set of the pre-test.

Difference in categorization as a result of participants race or gender

Race. Similar to Studies 1,3, we tested whether self-categorizing oneself as white led to changes in participants' self-categorization. We conducted a generalized mixed model analysis looking at whether participants' race (white or not white) predicted self-categorization. We also used a by-participant random variable. Results suggested that being white (compared to other races) did not lead to differences in self-categorization (b =-.03 [-.39, .32], SE = .17, z = -.18, p = .85).

Gender. We conducted a generalized mixed-model analysis using participants' gender to predict self-categorization. We further, used a by-participant random variable. Results suggested no significant difference between males and females in degree of self-categorization (b =.10 [-.34, .22], SE = .07, z = 1.51, p = .13).

Political affiliation. We tested whether participants political affiliation predicted self-categorization by conducting a generalized mixed-model analysis. We further used a by participant random variable. Results suggested that participants' political affiliation did not predict participants' self-categorization (b =.10 [-.03, .24], SE = .07, z = 1.51, p = .13).

Group identification. We used the same scale as in Studies 1 and 3 to test whether participants' group identification predicted self-categorization (α = .96 [.94, .97]). We conducted a generalized mixed-model analysis, predicting participants' evaluation of the mean from their degree of identification. Results suggested that participants self-categorization was not predicted by identification (b =.06 [-.07, .20], SE = .07, z = .97, p = .33).

Self-categorization compared to chance

We conducted a generalized mixed-model analysis to compare participants' decision of self-categorization in all of the four conditions (same mean high variance, same mean low variance, different mean high variance and different mean low variance) to chance. Results suggested that when the mean was similar to participants' own ratings, the choice in both the high standard deviation trials (b =1.28 [1.08, 1.49], SE = .10, z =12.39, p < .001) and the low standard deviation trials (b =1.59 [1.38, 1.80], SE = .10, z = 14.821, p < .001) was positive and higher than chance. When the mean group emotion was different from participants' own rating, self-categorization was positive and different from chance both in the low standard deviation trials (*b* =.43 [.25, .61], *SE* = .09, *z* =4.75, *p* < .001) and the high standard deviation (*b* =.21 [.03, .39], *SE* = .09, *z* =2.37, *p* = .02). These results were slightly different compared to those of Study 3 in which the high standard deviation condition was not significantly different from chance.

Differences between the higher and the lower group mean trials

In order to test differences between the higher and the lower mean group emotion trials we created an orthogonal contrast, first comparing the similar mean group emotion trials to the two other different mean group emotion trials, and then later comparing the higher and the lower mean group emotion trials to each other. We conducted a generalized mixed model analysis using our orthogonal contrast an independent variable and participants self-categorization as a dependent variable. In addition, we used a by-participant random variable. Looking specifically on the comparison between the higher and the lower trials, results suggested a significant difference between the low and the high conditions such that the higher group emotion was more likely to categorized as one's group (*b* =-.26 [-.36, -.16], *SE* = .04, *z* = -5.49, *p* < .001, see Figure 7S). These results are different from those of Study 3 and reflect a bias towards higher group emotions.

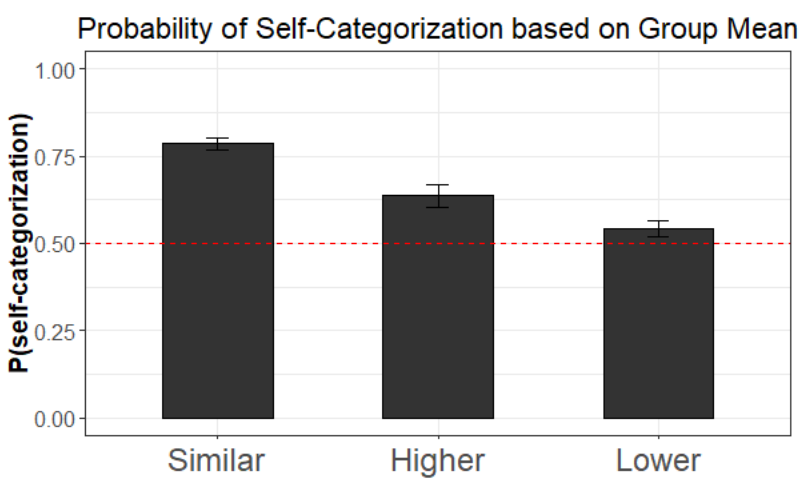

Figure 7S. Participants' self-categorization as a function of whether the mean group emotion was higher, lower or similar to participants' own rating. Error bars represent 95% confidence intervals.

Study 5

Self-categorization compared to chance

We conducted a generalized mixed-model analysis to compare participants' decision of self-categorization in all of the four conditions (same mean high variance, same mean low variance, different mean high variance and different mean low variance) to chance. Results suggested that when the mean was similar to participants' own ratings, the choice in both the high standard deviation trials (*b* =1.59 [1.3, 1.86], *SE* = .13, *z* =11.80, *p* < .001) and the low standard deviation trials (*b* =1.72 [1.45, 2.00], *SE* = .13, *z* = 12.37, *p* < .001) was positive and higher than chance. When the mean group emotion was different from participants' own rating, self-categorization was positive and different from chance both in the low standard deviation trials (*b* =.86 [.67, 1.06], *SE* = .09, *z* =8.76, *p* < .001) and the high standard deviation (*b* =1.25 [1.05, 1.46], *SE* = .10, *z* =12.20, *p* < .001).

Differences between the higher and the lower group mean trials

In order to test differences between the higher and the lower mean group emotion trials we created an orthogonal contrast, first comparing the similar mean group emotion trials to the two other different mean group emotion trials, and then later comparing the higher and the lower mean group emotion trials to each other. We conducted a generalized mixed model analysis using our orthogonal contrast an independent variable and participants self-categorization as a dependent variable. In addition, we used a by-participant random variable. Looking specifically on the comparison between the higher and the lower trials, results suggested a no significant difference between the low and the high conditions (*b =-.07* [-.15, .01], *SE* = .04, *z* = -1.85, *p* = .06, see Figure 8S). These results make sense as participants were merely completing a simple face categorization task in which there should not be a difference between the low and the high group mean condition.

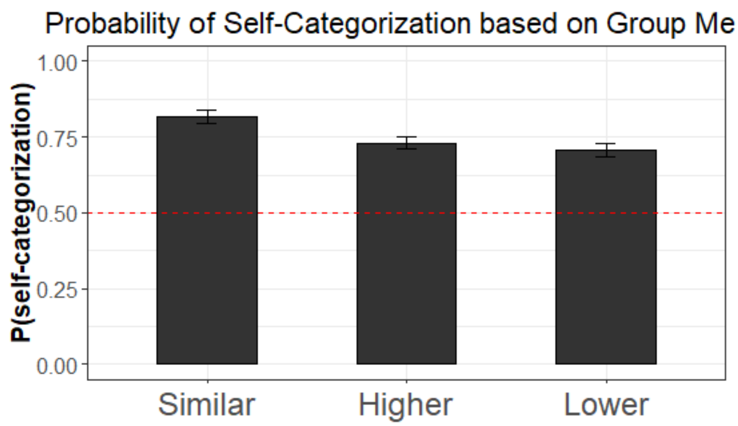

Figure 8S. Participants' self-categorization as a function of whether the mean group emotion was higher, lower or similar to participants' own rating. Error bars represent 95% confidence intervals.