

---

**ORIGINAL RESEARCH REPORT**

# Biased by the Group: Memory for an Emotional Expression Biases Towards the Ensemble

Jonathan C. Corbin and L. Elizabeth Crawford

---

An emotional expression can be misremembered as more similar to previously seen expressions than it actually was – demonstrating inductive category effects for emotional expressions. Given that memory is influenced over time, we sought to determine whether memory for a single expression would be similarly influenced by other expressions viewed simultaneously. In other words, we test whether the ability to encode statistical features of an ensemble (i.e., ensemble encoding) is leveraged when attempting to recall a single expression from the ensemble. In three preregistered experiments, participants saw an ensemble of 4 expressions, one neutral and the three either happy or sad. After a delay, participants were asked to reproduce the neutral face by adjusting a response face's expression. In Experiment 1, the ensemble comprised images of the same actor; in Experiments 2 and 3, images were comprised of individuals varying race and gender. In each experiment we demonstrated that even after only a single exposure, memory for the neutral expression in the happy group was biased happier relative to the same expression in the sad group. Data and syntax can be found at <https://osf.io/gcbez/>.

---

**Keywords:** Memory; Ensemble Encoding; Emotion; Facial Expressions

---

Memory for an emotional facial expression is influenced by knowledge gained from prior experience. Specifically, when people view and then reproduce individual facial expressions, their memory for each face is biased toward the central expression of faces they previously viewed (Corbin, Crawford, & Vavra, 2017). This result is consistent with a Bayesian model in which uncertain memory of a particular object is combined with prior knowledge about the distribution from which it is drawn (cf., Hemmer & Steyvers, 2009; Huttenlocher, Hedges & Vevea, 2000). In this line of research, the prior knowledge used to inform memory is acquired inductively, i.e., through experience with a series of objects presented one-at-a time. However, faces are social stimuli, and in many circumstances we encounter several people at once. Research has shown that when people see a group of faces, they encode its statistical properties in an ensemble representation (Herman & Whitney, 2007, 2009), but it is not known if such ensembles are used like inductive categories to adjust memories of individual faces. Here we present three studies examining whether memory of an individual face is biased by the mean ensemble expression.

Research using simpler stimuli that vary in size, shade, or orientation has shown that people capture the statistical structure in presented sets. Whether objects

are shown sequentially or simultaneously, when people are directly asked to estimate (or select from a set of choices) the average object, they do so with impressive accuracy (Ariely, 2001; Chong & Treisman, 2003; Oriet & Hozempa, 2016; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). This statistical information is thought to be computationally valuable for both perception and memory. When perceiving a group of objects, it affords efficient, compressed encoding (Alvarez & Oliva, 2008; Haberman & Whitney, 2012). When estimating a single object from memory, it can be used to reduce error in responses by combining noisy trace memory with prior information about what the stimulus was likely to be. In both areas of research, Bayesian models have been used to describe the combination of an inexact trace with a prior distribution.

A number of studies in the inductive category learning literature have shown evidence for this kind of combination when objects are studied individually (Ashourian & Loewenstein, 2011; Corbin et al., 2017; Crawford, Huttenlocher, & Engebretson, 2000; Hemmer & Steyvers, 2009; Huttenlocher et al., 2000; Olkkonen & Allred, 2014; Jazayeri & Shadlen, 2010), but there is less work examining whether ensembles are used in the same way. Brady and Alvarez (2011) showed that, under some conditions, estimates for the size of a colored circle were biased toward the mean size of similarly colored circles that had been presented simultaneously. Huang and Sekuler (2010) and Dubé, Zhou, Kahana, and Sekuler (2014) found that memory for individual gabor

patches varying in spatial frequency was biased towards task-irrelevant stimuli which were presented in tandem with the task relevant stimuli. Prior work on recognition memory for individual faces embedded in ensembles has found that participants were more likely to erroneously select expressions that were closer to the ensemble mean as having been shown previously than those that were farther away (Haberman & Whitney, 2007, 2009). These experiments typically rely on very brief exposures to ensembles in order to emphasize the ability to extract mean information without having access to corresponding information about specific features of the individual stimuli. However, Li et al. (2016) examined the role of processing time in recognizing individual expressions from an ensemble with a 2 AFC task and found that recognition of individual expressions did improve when individuals were given longer than 50 ms to view the ensembles. Whereas this work examined accuracy for individual items and ensembles, more recently, Griffiths, Rhodes, Jeffery, Palermo, and Neumann (2018) did show that in a recognition task, the intensity of individual expressions within an emotional spectrum (either happy or angry) were more likely to be misremembered as closer to the average intensity of the ensemble.

We are interested in how ensembles affect memory of an individual facial expression for two reasons. First, to examine the generality of ensemble encoding effects. There is reason to think that facial expressions might operate differently than the lower level stimuli often used in visual memory and ensemble encoding research. Faces are perceptually special; they communicate a wide range of socially relevant information that humans have prior experience interpreting (Roberson, Damjanovic, & Pilling, 2007). Furthermore, unlike simpler stimuli such as line length, or shade, emotions are inferred rather than perceived directly. Ensemble encoding research that tested the same participants on both low level features and on facial expressions showed only small correlations in performance (.05 to .29; Haberman, Brady, & Alvarez, 2015), suggesting largely independent mechanisms for encoding stimuli of differing complexity. Second, by extending prior work to facial expressions of emotions, we take a step closer to real world application of basic research. People are often encountered in groups, but it is not known how others in a group may influence our judgment and memory of an individual.

In three pre-registered experiments, we examine whether exposure to an ensemble of emotional facial expressions biases memory for one of the constituent expressions. By presenting only a single trial, the study isolates the cause of such a bias to the presented ensemble, and prevents people from acquiring information from repeated viewing of stimuli (c.f., Crawford, Corbin, & Landy, under review). Unlike prior work in ensemble encoding, which typically relies on recognition tasks, we rely on a method more consistent with the inductive category literature, in which participants must choose expressions from a continuous range, allowing us to avoid recognition or forced-choice designs, which require repetition to estimate memory accuracy. Furthermore, whereas prior

work has presented stimuli that are both perceptually and socially homogenous (e.g., removing hair, using one race or gender, grey scale images; see Griffiths et al., 2018; Haberman & Whitney, 2007), we intentionally present ensembles that include people of different apparent races and genders to determine if memory is adjusted by ensembles formed from a diverse group. If estimates are biased as predicted, this would suggest that inductively learned categories and ensembles have similar impacts on memory, even with only a single exposure in a task that includes a diverse group.

## Experiment 1

### Method

#### Participants

Two hundred thirty-six participants were recruited from Mechanical Turk (our preregistration aimed for 200, but some mturkers began the hit but never submitted, and thus weren't counted during data collection). This sample size was estimated based on pilot testing which yielded an effect size of  $d = .61$ , with alpha set at .01 and power = 0.95 (see <https://osf.io/eqph8/> for preregistered sample size and analysis plans. All data and syntax for analyses can be found at <https://osf.io/gcbez/>.) Mturkers were paid \$0.30 for their participation. Forty-one participants failed to complete the demographic questionnaire, and thus are not accounted for in the following statistics (possibly due to need to enter a new website to fill out demographic questions). The sample was 63% male, 80.5% White, 8.7% Asian, 7.7% African American, ~1% American Indian or Alaskan Native and ~1% Native Hawaiian or Pacific Islander (1.5% preferred not to answer) and, 87.2% Non-Hispanic or Latino, 9.7% Hispanic or Latino, and 3.1% preferred not to answer. The average age of our sample was 34.91 years ( $SD = 10.63$ , min = 20, max = 70).

#### Materials

Images used in the present study were drawn from the NimStim face stimulus set,<sup>1</sup> a database of stock photographs of young adults, varying in ethnicity and gender, depicting various emotional expressions. Photos of a female making sad, neutral, and happy expressions were used to create the stimuli. Using FantaMorph software (Abrosoft, 2002), two sets of morphs were created: one set changed from the model's sad expression to her neutral expression, the second changed from neutral to happy at 5% increments. From these sets, we extracted 41 evenly distributed expressions of each model's face ranging from sad (expression -20), to neutral (expression 0), to happy (expression 20). Given that changes in hair position could lead to distracting artifacts in the morph, we edited the initial images prior to morphing to maintain consistent hair placement and maintain a seamless, but realistic morph.

#### Procedure

The experiment was designed and implemented using GameMaker software (Version 1.4; YoYo Games, 2015). Participants viewed a group of four expressions (all the same face), which were shown in a 2 × 2 grid pattern

at the center of the screen (location of each expression was randomized). Participants were randomly assigned to either the Sad condition, in which the expressions of the four faces were 0, -5, -10, and -15 or to the Happy condition in which expressions were 0, 5, 10, and 15. The expressions were displayed for ~4 seconds, followed by a ~2 second delay in which a blank screen was displayed (timing variables may have slightly varied due to users' internet connection speeds.) Next, the response face reappeared in the area of the grid that contained the neutral expression (expression 0) and participants were asked to recall the expression that was previously in that location. Participants were not told prior to this which response face would be tested and position of the expressions was randomized. Participants were instructed to scroll through the morphed expressions in order to select the expression they believe was the one they saw in the designated position (participants could scroll through the entire range of expressions and the expressions would circle back if they reached the extreme expression (i.e., they could scroll from 19 to 20, back to 19 with the same key).) The response face was already set at neutral, so no actual change was needed for a correct response. Given that this experiment only consisted of a single trial, participants were given a practice trial prior to the actual experiment that mimicked this procedure exactly except that shaded squares were used rather than expressions, and instructions were included to help with understanding.

#### Data Preparation

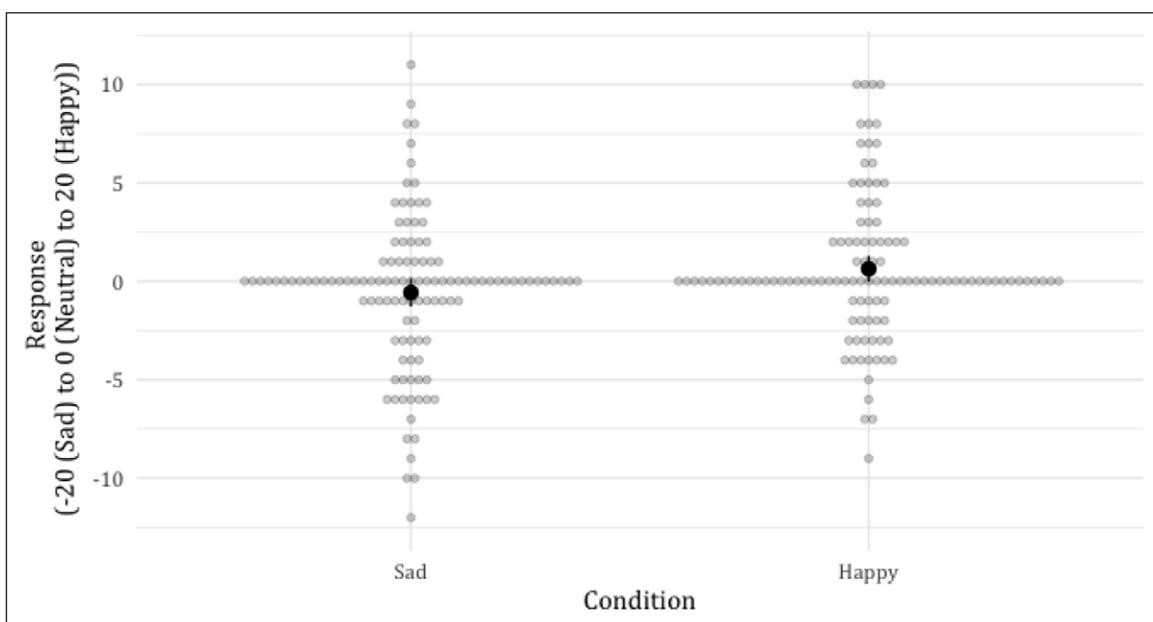
Prior to analyses, we computed error scores for each participant by taking the absolute value of participant's responses (i.e., the absolute distance from the correct response of 0). Next, we removed participants whose responses were greater than 2.5 *SDs* from the mean error

of the group that the participant was assigned to. Three participants were removed from the Sad condition (the cutoff was 12.15), and four participants were removed from the Happy condition (the cutoff was 10.97), leaving the sample at  $N = 229$ .

#### Results and Discussion

All analyses were conducted using the R software package (R Core Team 2012) and RStudio (RStudio Team, 2016). As specified in our preregistration, alpha was set at a .05 threshold (we relied on a stricter criterion for power analysis given uncertainty about the effect size estimated in pilot studies.) In support of our prediction, participants assigned to the Sad condition misremembered the neutral face as sadder than participants assigned to the Happy condition ( $M_{Sad} = -0.57$   $SD = 3.82$ ;  $M_{Happy} = 0.63$ ,  $SD = 3.57$ ), Welch's  $t(226.22) = 2.47$ ,  $p = 0.014$ ,  $d = .33$  (see **Figure 1**). We also conducted a JZS Bayes factor t-test (Morey & Rouder, 2015) with a Cauchy distributed prior ( $\delta = .707$ ), which yielded a  $BF_{10} = 2.48$ , meaning that the data are 2.48 times more likely under the alternative hypothesis than under the null. Despite the mean difference between conditions, it is worth noting that 40.17% of participants chose the correct expression. The fact that so many participants had perfect accuracy is likely due to the fact that the starting expression was the correct response.

These results show a bias towards the ensemble even after only a single exposure. They also show that this bias can persist even after a relatively long exposure (as compared to previous ensemble studies). The single-trial design eliminates the possibility that effects could stem from any prior learning over time. It also allowed us to generalize bias towards the ensemble mean to single-exposure situations. Large within-participant designs have many advantages over a between-participant approach for



**Figure 1:** Means and 95% CIs for Happy and Sad Conditions along with the distributions of raw data for each condition. Plots were built using ggplot2 (Wickham, 2009).

ensemble encoding research, including maximizing one's sample-to-power ratio (and therefore requiring fewer resources to complete a study) and allowing for within-participant estimates of an effect. However, despite these advantages, in order to answer the question as to whether these effects are reliable for single-exposure situations, one must design a study meant for this purpose.

## Experiment 2

In Experiment 1, we showed that on average, memory for a single emotional expression is biased towards that of other expressions in an ensemble. Furthermore, by only presenting a single trial, we demonstrated that these effects could be found even within the context of a single exposure. However, it is important to note that the strength of the evidence for this hypothesis was weak (i.e.,  $BF_{10} = 2.48$ ), and replication is warranted. In Experiment 2, we examine whether the same effects from Experiment 1 will replicate when presenting multiple unique individuals. In line with prior work on ensemble encoding (e.g., Li et al., 2016; Haberman & Whitney, 2007, 2009), our stimuli for Experiment 1 consisted of a single individual's face making four different expressions. However, relying on a single individual's face has a number of drawbacks. First, it limits the extent to which we can generalize these effects to situations involving different people. Second, it does not differentiate whether individuals are relying on an ensemble representation of a particular organization of facial features or an ensemble representation of an emotional state. Whereas Griffiths et al., 2018 did rely on different individual's faces in examining effects of ensembles on memory for individual facial expressions, their stimuli were all young, Caucasian males who had many individuating features removed. Replicating Experiment 1's findings with multiple visually distinct faces would help tease apart these mechanisms (though not fully, as emotional expressions will by necessity share structural similarities across individuals.) Finally, this approach has the benefit of providing participants with a cue beyond position for which expression they must remember (the unique individual's face).

### Participants

Two hundred thirty participants were recruited from Mechanical Turk. Mturkers were paid \$0.30 for their participation. Forty-five Mturkers failed to complete the demographic questionnaire, and thus are not accounted for in the following statistics. The sample was 63% male, 77.7% White, 10.3% Asian, 5.4% African American, 4.3% American Indian or Alaskan Native (2.2% preferred not to answer) and, 87.5% Non-Hispanic or Latino, 11.4% Hispanic or Latino, and 1.1% preferred not to answer. The average age of our sample was 33.41 years ( $SD = 8.94$ ,  $min = 20$ ,  $max = 62$ ).

### Materials

Images used in the present study were drawn from the NimStim face stimulus set, a database of stock photographs of young adults, varying in ethnicity and gender, depicting various emotional expressions. In the same procedure as

Experiment 1, morphs were created for sixteen different individuals, varying evenly on the dimensions of sex (Female, Male) and race (African American, White), totaling 4 unique faces per demographic group. Ensembles of four faces were constructed by randomly assigning one of the faces from each demographic to an ensemble.

### Procedure

The procedure for Experiment 2 was similar to the first experiment's except that four different individuals were shown, each assigned to a different expression. At test, the response face matching one of the individuals appeared in the same location as that individual and participants were instructed to make the expression match the one that individual had seen previously. Thus location and identity both served as retrieval cues for the expression. Every ensemble included a white female, white male, black male, and a black female and the assignment of identity to expression and location was randomized. Participants participated in three additional trials after the initial trial in which they were asked to reproduce the happy extreme, sad extreme, as well as the neutral expression one more time, but these were collected as pilot data for future work. Finally, prior to collecting demographic data at the end of the study, participants were given a short form measuring Big 5 personality traits, which was also meant as pilot data.

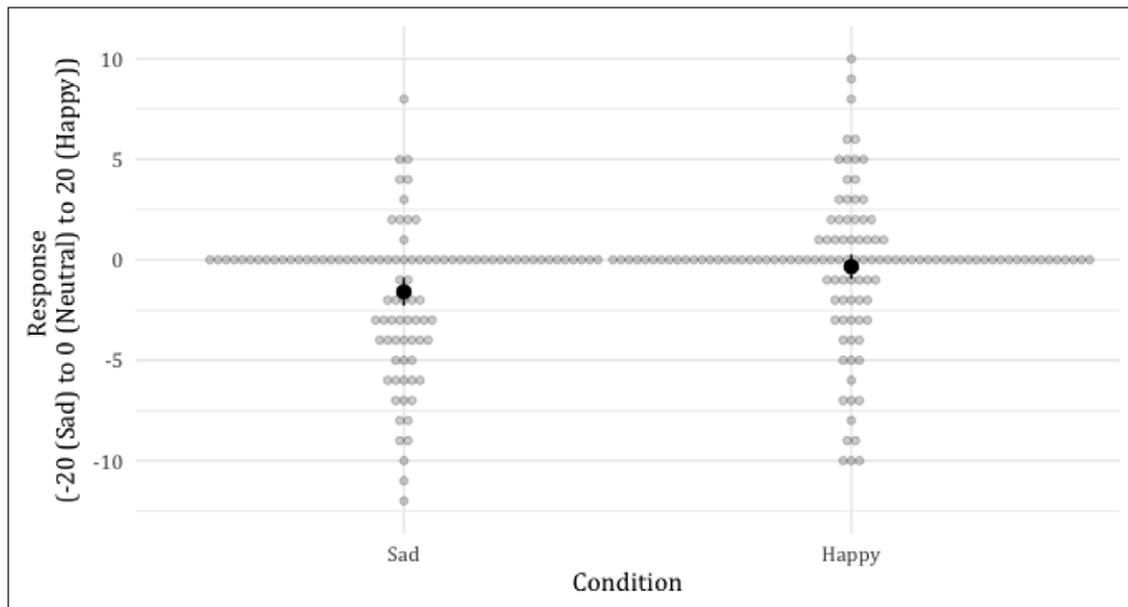
### Data Preparation

Prior to analyses, we computed error scores for each participant by taking the absolute value of participant's responses. Next, we removed participants whose responses were greater than 2.5  $SDs$  from the mean error of the group that the participant was assigned to. Three participants were removed from the Sad condition (the cutoff was 9.82), and four participants were removed from the Happy condition (the cutoff was 8.93), leaving the sample at  $N = 223$ .

### Results and Discussion

As in Experiment 1, the neutral expression was remembered as sadder in the Sad condition as compared to the Happy condition ( $M_{sad} = -1.59$   $SD = 3.46$ ;  $M_{happy} = -0.34$ ,  $SD = 3.40$ ), Welch's  $t(210.32) = 2.70$ ,  $p = 0.007$ ,  $d = .37$  (see **Figure 2**).<sup>2</sup> A Bayes t-test with a Cauchy prior of  $\delta = .33$  (adjusting for the effect size in Experiment 1) yielded a  $BF_{10} = 5.89$ . Similar to Experiment 1, a large percentage of participants (48.88%) chose the correct expression.

As in the first experiment, Experiment 2 showed that estimates of neutral face depended on the surrounding ensemble. When it was embedded in the Sad ensemble, it was estimated to be sadder than when it was embedded in the Happy ensemble. Unlike Experiment 1, which showed a roughly symmetrical bias towards the ensemble mean for Happy and Sad conditions, here the difference between the two conditions is asymmetric and appears to be driven by the bias in the Sad condition. While the difference between conditions was as predicted, this asymmetry was surprising, and it suggests that there may be additional sources of bias at play in this reproduction task.



**Figure 2:** Means and 95% CIs for Happy and Sad Conditions along with the distributions of raw data for each condition.

### Experiment 3

Experiment 2 demonstrated that memory for an individual expression depends on others in the group. The same face was remembered as sadder when embedded amongst sad expressions than when it was embedded amongst happy expressions. Furthermore, Experiment 2 still showed a large percentage of participants choosing the correct expression. Whereas this could simply be due to the ease of the task (they were provided with the correct expression from the start), this could also be due to individuals failing to adequately engage the task. Experiment 3 accounts for both possibilities by changing the starting expression of the response face. This removes the potential benefit of providing the correct expression and will allow us to account for participants who fail to change the expression due to lack of engagement.

#### Participants

Two hundred thirty-one participants were recruited from Mechanical Turk. Mturkers were paid \$0.30 for their participation. Twenty-three mturkers failed to complete the demographic questionnaire, and thus are not accounted for in the following statistics. The sample was 57.6% male, 83.3% White, 8.4% Asian, 5.9% African American, <1% American Indian or Alaskan Native and Native Hawaiian or Pacific Islander (~1% preferred not to answer) and, 89.2% Non-Hispanic or Latino, 10.3% Hispanic or Latino, and <0.5% preferred not to answer. The average age of our sample was 35.48 years ( $SD = 9.88$ ,  $min = 22$ ,  $max = 67$ ).

#### Materials and Procedure

Materials and procedure for Experiment 3 were the same as Experiment 2, except that the starting value for the response expression was set at either  $-20$  (the saddest expression) or  $20$  (the happiest expression). Like Experiment 2, the identity of the response expression was identical to that of the target (neutral) expression in

the study phase. It is worth noting that the participants themselves were not told that they were starting at the most extreme expression, as the morphs scrolled continuously from  $-20$  to  $20$ , and back to  $-20$  (or vice versa).

#### Data Preparation

Prior to analyses, 31 participants were removed (13.42% of the sample), because they failed to change the starting expression. In the culled sample, no participant's response was greater than 2.5  $SDs$  from the mean error in the Sad (19.81) or Happy (21.04) conditions, so no other participants were removed, leaving the sample at  $N = 200$ .

#### Results and Discussion

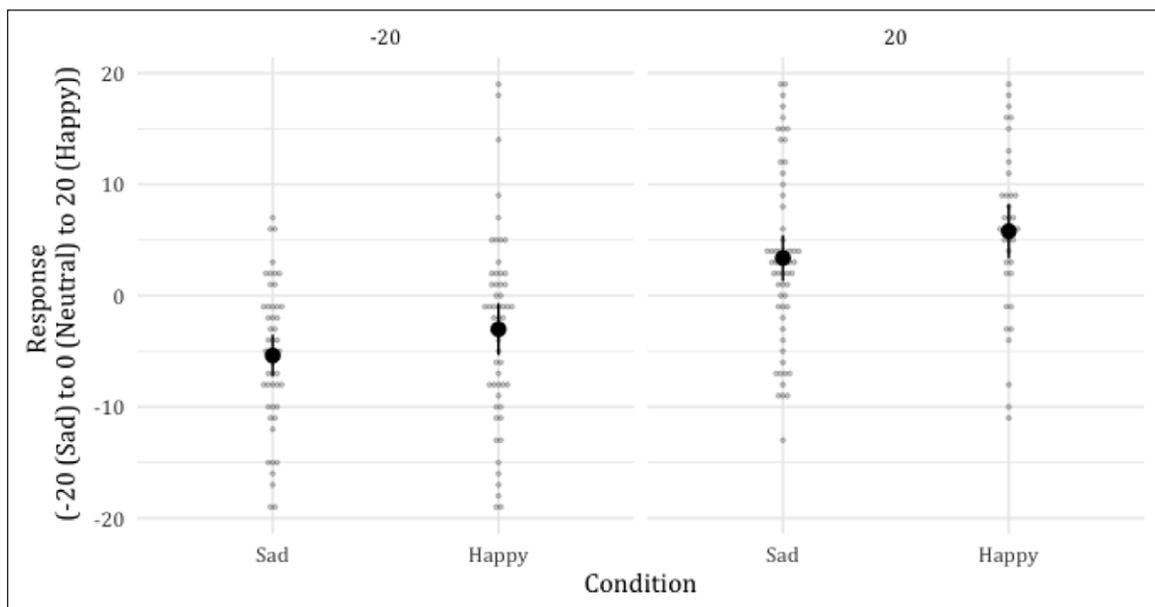
We conducted a linear regression with condition (Sad, Happy) predicting participants' chosen expressions, controlling for start value. As shown in **Table 1**, at  $\alpha = .05$  (see <https://osf.io/emrtf/> for preregistered analysis plan), responses in the Sad condition were remembered as significantly sadder than those in the Happy condition ( $M_{sad} = -1.01$ ,  $SE = 0.74$ ;  $M_{Happy} = 1.38$ ,  $SE = 0.80$ ; note: means are least-squared estimates controlling for start value). Responses also anchored on start value ( $M_{Expression -20} = -4.20$ ,  $SE = 0.75$ ;  $M_{Expression 20} = 4.58$ ,  $SE = 0.79$ ). A model including an interaction term failed to yield a significantly better fit to the data ( $F < .001$ ,  $p = .983$ ; but see **Figure 3** for raw data split by Start Value and Condition.) A JZS Bayes factor ANOVA with start value treated as a nuisance variable and the Cauchy prior for Happy set at  $\delta = .37$  (based on effect size from Experiment 2) yielded a  $BF_{10} = 1.64$ .<sup>3</sup>

Unlike Experiment 2 (but consistent with Experiment 1), bias towards the ensembles in each condition was approximately symmetrical. Whereas we were only interested in the effect of the ensemble on memory for the individual expression, the large effect of start

**Table 1:** Ensemble Expression and Start Value Predicting Response.

	<b>B</b> <b>(95% CI)</b>	<b>std.</b> <b>Error</b>	<b>Beta</b> <b>(95% CI)</b>	<b>std.</b> <b>Error</b>	<b>p</b>
Intercept	-5.40 (-7.27 – -3.53)	0.95			<.001
Condition (Happy)	2.39 (0.23 – 4.56)	1.10	0.14 (0.01 – 0.26)	0.06	.030
Start Value (20)	8.78 (6.62 – 10.94)	1.09	0.50 (0.38 – 0.62)	0.06	<.001
$R^2/adj. R^2$		.250/.242			
$F$		32.85			<.001

Note. Table was generated using sjplot package (Lüdtke, 2017).



**Figure 3:** Means and 95% CIs for Happy and Sad Conditions separated by Start Value along with the distributions of responses for each condition.

value demonstrated another factor besides memory that influences responses (see Allred, Crawford, Duffy, & Smith, 2016; Corbin et al., 2017 for other examples of start value effects). Finally, in contrast to the prior experiments, only 2.0% of participants chose the correct expression, suggesting that forcing participants to adjust the expression (rather than simply allowing them to recognize the expression without having to change the morph) made the task substantially more difficult.

**General Discussion**

Prior research on ensemble encoding for emotional expressions has demonstrated that people rapidly and reliably encode representations that capture average expression in a group of faces (Haberman & Whitney, 2007, 2009; Li et al., 2016). Extending this work, here we tested whether such representations influence memory for an individual face. Our hypothesis stemmed from work on inductive category learning, which has demonstrated that memory for a single facial expression is biased toward the average expression of the group

from which it was sampled (Corbin et al., 2017). If inductively learned categories are used to minimize error in estimating an expression (i.e., an estimate of an expression is systematically biased towards the central value), it seems reasonable that ensemble statistics would operate in the same way. Results from three experiments supported this hypothesis, demonstrating that on average, memory for a single expression is biased by the surrounding expressions of the group in which that face was embedded.

Experiments 1 and 2 provide a conservative test of our hypothesis by giving participants the correct expression as the response expression. Participants could have achieved the correct estimate simply by leaving the response expression alone. Whereas this led to a substantial proportion of participants estimating correctly (though under 50%), the majority of participants did alter the expression. It could be that, because the first two experiments implied a need to alter the expression, some of these adjustments reflect demand characteristics. However, demand characteristics alone would not account

for the direction of these adjustments. Experiment 3 replicated the effect when the starting expression was set at one of the two extremes (sad or happy), and also showed that estimates were additionally biased towards each start value. Whereas Griffiths et al. (2018) showed that individuals bias memory for an individual expression towards mean facial expressions of a group of different individual faces (in their case, all young male Caucasian identities), Experiments 2 and 3 demonstrated that ensembles can influence memory even when the group is socially diverse (in this case, a mix of Black and White men and women.) This result suggests that ensembles can be formed that abstract across differences in the identity of faces to capture commonalities in emotional expression.

Prior research examining the relationship between individual stimulus and the ensemble has typically relied on multiple repetitions – an approach that is warranted given the goal of maximizing measurement precision. However, it is also important to determine whether an effect is robust across various situations. Our experiments focused on a single trial in order to estimate the average size of the ensemble bias for a novel experience. This result allows us to extend ensemble-encoding effects on single stimuli to situations that involve only a single exposure, eliminating the possibility that these effects rely on participants' learning statistical regularities or developing task-specific strategies over many trials. Also, the real world doesn't always present us with hundreds of opportunities to view an individual. For example, plenty of eyewitness scenarios involve brief, single exposures to individuals who were part of a group at the time of exposure. Our results suggest that features common to the group in a scenario like this may bias memory when it comes time to single out an individual. Whereas we focused on emotional expressions, this work could be extended to examine other perceptual attributes such as height, weight, or apparent racial identity, which may be relevant in legal settings.

### Data Accessibility Statement

All data can be accessed at <https://osf.io/gcbez/>.

### Notes

- <sup>1</sup> Development of the MacBrain Face Stimulus Set was overseen by Nim Tottenham and supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. Please contact Nim Tottenham at [tott0006@tc.umn.edu](mailto:tott0006@tc.umn.edu) for more information concerning the stimulus set.
- <sup>2</sup> Although not preregistered, we looked to see if there was an interaction between target identity and emotion condition and found no significant effect. We note that that these studies were not sufficiently powered for such an analysis and this question would be better addressed by a multi-trial experiment.
- <sup>3</sup> We also applied this culling rule and analysis to an identical experiment for which they were not preregistered (see Group Study Footnote 3 in <https://osf.io/gcbez/>). With  $n = 103$ , results were

comparable to those of Experiment 3, with estimated bias (controlling for start value) in the Sad condition remembered as significantly sadder than those in the Happy condition ( $M_{sad} = -1.51$ ,  $SE = 1.11$ ;  $M_{Happy} = 1.91$ ,  $SE = 1.16$ ;  $t(100) = -2.11$ ,  $p = .037$ ,  $BF_{10} = 1.45$ ). Combining the two datasets and re-computing the Bayesian t-test yielded a  $BF_{10} = 10.63$ .

### Competing Interests

The authors have no competing interests to declare.

### Author Contributions

- Contributed to conception and design: JCC, BC
- Contributed to acquisition of data: JCC, BC
- Contributed to analysis and interpretation of data: JCC, BC
- Drafted and/or revised the article: JCC, BC
- Approved the submitted version for publication: JCC, BC

### References

- Abrosoft.** (2002). FantaMorph SE (Version 5.4.6) [Software]. Available from: <http://www.fantamorph.com/>.
- Allred, S. R., Crawford, L. E., Duffy, S., & Smith, J.** (2016). Working memory and spatial judgments: Cognitive load increases the central tendency bias. *Psychonomic bulletin & review*, *23*, 1825–1831. DOI: <https://doi.org/10.3758/s13423-016-1039-0>
- Alvarez, G. A., & Oliva, A.** (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological science*, *19*(4), 392–398. DOI: <https://doi.org/10.1111/j.1467-9280.2008.02098.x>
- Ariely, D.** (2001). Seeing sets: Representation by statistical properties. *Psychological science*, *12*(2), 157–162. DOI: <https://doi.org/10.1111/1467-9280.00327>
- Ashourian, P., & Loewenstein, Y.** (2011). Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS ONE*, *6*, e19551. DOI: <https://doi.org/10.1371/journal.pone.0019551>
- Brady, T. F., & Alvarez, G. A.** (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*, 384–392. DOI: <https://doi.org/10.1177/0956797610397956>
- Chong, S. C., & Treisman, A.** (2003). Representation of statistical properties. *Vision Research*, *43*, 393–404. DOI: [https://doi.org/10.1016/S0042-6989\(02\)00596-5](https://doi.org/10.1016/S0042-6989(02)00596-5)
- Corbin, J. C., Crawford, L. E., & Vavra, D. T.** (2017). Misremembering emotion: Inductive category effects for complex emotional stimuli. *Memory & Cognition*, *45*, 691–698. DOI: <https://doi.org/10.3758/s13421-017-0690-7>
- Crawford, L. E., Corbin, J. C., & Landy, D.** (under review). Prior experience informs ensemble encoding.
- Crawford, L. E., Huttenlocher, J., & Engebretson, P. H.** (2000). Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science*, *11*, 280–284. DOI: <https://doi.org/10.1111/1467-9280.00256>

- Dubé, C., Zhou, F., Kahana, M. J., & Sekuler, R.** (2014). Similarity-based distortion of visual short-term memory is due to perceptual averaging. *Vision research*, *96*, 8–16. DOI: <https://doi.org/10.1016/j.visres.2013.12.016>
- Griffiths, S., Rhodes, G., Jeffery, L., Palermo, R., & Neumann, M. F.** (2018). The average facial expression of a crowd influences impressions of individual expressions. *Journal of Experimental Psychology: Human Perception and Performance*, *44*, 311–319. DOI: <https://doi.org/10.1037/xhp0000446>
- Haberman, J., Brady, T. F., & Alvarez, G. A.** (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, *144*, 432. DOI: <https://doi.org/10.1037/xge0000053>
- Haberman, J., & Whitney, D.** (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*, R751–R753. DOI: <https://doi.org/10.1016/j.cub.2007.06.039>
- Haberman, J., & Whitney, D.** (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 718. DOI: <https://doi.org/10.1037/a0013899>
- Haberman, J., & Whitney, D.** (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In: Wolfe, J., & Robertson, L. (Eds.), *Oxford series in visual cognition. From perception to consciousness: Searching with Anne Treisman*, 339–349. New York, NY, US: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:osobl/9780199734337.003.0030>
- Hemmer, P., & Steyvers, M.** (2009). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin & Review*, *16*, 80–87. DOI: <https://doi.org/10.3758/PBR.16.1.80>
- Huang, J., & Sekuler, R.** (2010). Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision*, *10*, 24–24. DOI: <https://doi.org/10.1167/10.2.24>
- Huttenlocher, J., Hedges, L. V., & Vevea, J. L.** (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*, 220–241. DOI: <https://doi.org/10.1037/0096-3445.129.2.220>
- Jazayeri, M., & Shadlen, M. N.** (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*, 1020–1026. DOI: <https://doi.org/10.1038/nn.2590>
- Li, H., Ji, L., Tong, K., Ren, N., Chen, W., Liu, C. H., & Fu, X.** (2016). Processing of individual items during ensemble coding of facial expressions. *Frontiers in psychology*, *7*. DOI: <https://doi.org/10.3389/fpsyg.2016.01332>
- Lüdecke, D.** (2017). sjPlot: Data Visualization for Statistics in Social Science. *R package version 2.3.1*, <URL: <https://CRAN.R-project.org/package=sjPlot>>.
- Morey, R. D., & Rouder, J. N.** (2015). BayesFactor (Version 0.9.10-2) [Computer software].
- Olkkonen, M., & Allred, S. R.** (2014). Short-term memory affects color perception in context. *PloS One*, *9*, e86488. DOI: <https://doi.org/10.1371/journal.pone.0086488>
- Oriet, C., & Hozempa, K.** (2016). Incidental statistical summary representation over time. *Journal of Vision*, *16*, 1–14. DOI: <https://doi.org/10.1167/16.3.3>
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M.** (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*, 739. DOI: <https://doi.org/10.1038/89532>
- R Core Team.** (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. <https://www.R-project.org>.
- Roberson, D., Damjanovic, L., & Pilling, M.** (2007). Categorical perception of facial expressions: Evidence for a category adjustment model. *Memory & Cognition*, *35*, 1814–1829. DOI: <https://doi.org/10.3758/BF03193512>
- RStudio Team.** (2016). *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc. <http://www.rstudio.com/>.
- Wickham, H.** (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. DOI: <https://doi.org/10.1007/978-0-387-98141-3>
- YoYo Games Ltd.** (2015). GameMaker [Computer Software]. Retrieved from: <https://www.yoyogames.com/gamemaker>.

**Peer review comments**

The author(s) of this paper chose the Open Review option, and the peer review comments are available at: <http://doi.org/10.1525/collabra.186.pr>

**How to cite this article:** Corbin, J. C., & Crawford, L. E. (2018). Biased by the Group: Memory for an Emotional Expression Biases Towards the Ensemble. *Collabra: Psychology*, *4*(1): 33. DOI: <https://doi.org/10.1525/collabra.186>

**Senior Editor:** Rolf Zwaan

**Editor:** Akira O'Connor

**Submitted:** 03 August 2018

**Accepted:** 31 August 2018

**Published:** 15 October 2018

**Copyright:** © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.