CrossMark

# Misremembering emotion: Inductive category effects for complex emotional stimuli

**Jonathan C. Corbin**[1] · **L. Elizabeth Crawford**[1] · **Dylan T. Vavra**[1]

© Psychonomic Society, Inc. 2017

**Abstract** Memories of objects are biased toward what is typical of the category to which they belong. Prior research on memory for emotional facial expressions has demonstrated a bias towards an emotional expression prototype (e.g., slightly happy faces are remembered as happier). We investigate an alternate source of bias in memory for emotional expressions – the central tendency bias. The central tendency bias skews reconstruction of a memory trace towards the center of the distribution for a particular attribute. This bias has been attributed to a Bayesian combination of an imprecise memory for a particular object with prior information about its category. Until now, studies examining the central tendency bias have focused on simple stimuli. We extend this work to socially relevant, complex, emotional facial expressions. We morphed facial expressions on a continuum from sad to happy. Different ranges of emotion were used in four experiments in which participants viewed individual expressions and, after a variable delay, reproduced each face by adjusting a morph to match it. Estimates were biased toward the center of the presented stimulus range, and the bias increased at longer memory delays, consistent with the Bayesian prediction that as trace memory loses precision, category knowledge is given more weight. The central tendency effect persisted within and across emotion categories (sad, neutral, and happy). This article expands the scope of work on inductive category effects to memory for complex, emotional stimuli.

✉ Jonathan C. Corbin
jcorbin@richmond.edu

[1] Department of Psychology, University of Richmond, Sara Brunet Hall, 28 Westhampton Way, Richmond, VA 23173, USA

**Keywords** Memory · Central tendency bias · Emotion · Facial expressions

Memory is rarely perfect, and in many cases, the distance from perfection tends to grow over time. Not only does it become increasingly noisy, it also tends to become biased in a way that systematically favors categories. For instance, episodic memory favors semantic categories, resulting in false memories that capture overall meaning at the expense of the details (Brainerd, Yang, Reyna, Howe, & Mills, 2008). Memory for spatial locations is biased by structural categories carved out in space, such as the four quadrants of a computer screen (Crawford, Landy, & Salthouse, 2016; Simmering, Spencer & Schöner, 2006), and memory for stimulus attributes are biased towards the center of the category that they belong to (i.e., the central tendency bias; Huttenlocher, Hedges, & Vevea, 2000). The central tendency bias has been attributed to a Bayesian combination of an imprecise memory trace of the object combined with prior information about its category (e.g., Feldman, Griffiths, & Morgan, 2009; Hemmer & Steyvers, 2009; Huttenlocher, et al., 2000).

In many studies of the central tendency bias, participants view and immediately reproduce individual stimuli that vary with regard to a specific attribute, such as length (Ashourian & Loewenstein, 2011; Duffy, Huttenlocher, Hedges, & Crawford, 2010), size (Huttenlocher et al., 2000), hue (Olkkonen & Allred, 2014), and time (Jazayeri & Shadlen, 2010). The results show that estimates tend to be biased toward a central value of the presented set, suggesting that participants form an inductive category out of their experience with stimuli and use that category to reconstruct memories. Work has confirmed that the central tendency bias is indeed a phenomenon of reconstructive memory rather than a bias in encoding (Crawford, Huttenlocher, & Engebretson, 2000),

 Springer

and that the bias increases along with cognitive load (Allred, Crawford, Duffy, & Smith, 2016) and over a lengthening delay (Olkkonen & Allred, 2014; Olkkonen, McCarthy, & Allred, 2014). Whereas previous studies have focused on simple, concrete attributes, the present article extends this prior work to a new domain – memory for emotional facial expressions. Facial expressions are socially meaningful and visually complex stimuli with which people have extensive prior experience, and as a result, it is not clear that the kind of inductive category effects shown in prior research would occur for these stimuli.

Studies of memory for facial expressions has focused on other sources of bias for memory for facial expressions. For example, Halberstadt and Niedenthal (2001) showed that memory for expressions is suggestible (see also Halberstadt, 2001, 2005). In these studies, experimenters took photos of actors expressing happiness and sadness and, using morphing software, created 100 pictures blending happy and sad expressions ranging from one end of the spectrum (100% happy) to the other (100% sad). Participants were shown a photo that represented the mid-point between emotions (i.e., the most ambiguous expression) and were asked to explain why the face was either happy or sad (participants received multiple trials of both happy and sad explanation conditions). After a 30-min delay, participants were given a recognition task in which they used a slider to progress through the expressions (from sad to happy). Results showed that the act of explaining the emotion-biased memory of the ambiguous expressions (angry-happy or angry-sad blends) in the direction of the explanation.

Another potential source of bias is people's prior knowledge about emotional categories. In a two-alternative forced-choice paradigm, Roberson, Damjanovic, and Pilling (2007) found that participants were more likely to make memory errors favoring emotional expressions that were more similar to an emotion category prototype (e.g., a 90% happy/10% angry as opposed to 60% happy/40% angry blend). This effect also held for caricatures of emotions, such that a face that was made to over-emphasize features associated with happiness would be falsely remembered as less caricatured (and closer to the emotion prototype), suggesting that individuals do not simply exaggerate expression in memory, but rather bias toward an emotion category prototype that reflects real-world experience of emotions.

Given that people already have a great deal of experience with facial expressions, it is not clear if they will form inductive categories to capture a distribution of presented faces as they do with other kinds of stimuli. Here we explore this question using the immediate reproduction paradigm commonly used in studies of the central tendency bias, but not previously used to study facial expression memory. As in earlier work on the central tendency bias, we have participants view and estimate individual faces taken from a set of face

stimuli that range continuously between neutral and a given emotion. If participants use their pre-existing emotion categories to inform their memories of individual faces, their estimates of intermediate faces should be biased toward emotional category prototypes (as in Roberson et al., 2007). Alternatively, if participants use an inductively formed category based on their experience with the presented distribution, their estimates should be biased toward the center of that distribution. In either case, if categories are used to reconstruct uncertain memories, then category effects should be more pronounced when uncertainty is greater. Here we manipulate uncertainty by varying the delay between encoding and retrieval. We test these hypotheses in four experiments, varying several aspects of the design. In Experiments 1a and b, we examine memory for mildly happy to happy emotional expressions for a variety of male and female faces. In Experiment 2, we look at whether the effects in the prior studies replicate for a neutral to sad distribution of expressions. Finally, in Experiment 3 we look at a distribution of expressions that crosses from sad to neutral to happy. In each experiment we also manipulate delay between the stimulus and response.

## Experiments 1a and b

### Participants

In Experiment 1a,[1] 84 participants were recruited from Mechanical Turk. Twenty-six were removed because they failed to complete the study, leaving a total of 58 participants. Seventy-two participants were recruited for Experiment 1b, and seven of these participants failed to complete the study, leaving the sample at 65.

### Materials

Images used in the present study were drawn from the NimStim[2] face stimulus set, a database of stock photographs of young adults, varying in ethnicity and gender, depicting various emotional expressions. Ten models (five male, five female) were chosen from the database to use in this study, and the sad, neutral, and happy expressions of each model were used to create the stimuli. Because in some cases changes in hair position led to distracting artifacts in the morphed sets, we edited the initial images to maintain consistent hair

---

[1] Data from an unreported pilot study as well as the studies in this manuscript are available at osf.io/2wuz3.

[2] 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 for more information concerning the stimulus set.

placement and create seamless morphs. Using FantaMorph software (Abrosoft, 2002), two sets of morphs were created for each model: one set changed from the model's sad expression to his neutral expression, the second changed from neutral to happy. From these sets, we extracted 41 evenly distributed expressions of each model's face ranging from Saddest (Expression 0), to Neutral (Expression 20), to Happiest (Expression 40), enabling us to express the change across frames in 5% increments.

## Design

The experiment was a 2 Delay (500 ms, 4 s) × 10 Stimulus Expression (Mildly Happy [Expression 26] to Happy [Expression 35]) × 2 Block (Block 1, Block 2) design. All factors were within-participant. The ten expressions were therefore shown four times, twice with a 500-ms delay and twice with a 4-s delay. In other words, a Block consisted of 20 trials – with each expression being shown at both delays within a Block (no expressions were repeated before all ten were shown). Each expression was given by a unique face, with five female and five male faces. In Experiment 1a, we randomized the expression of the starting face (i.e., the expression that participants saw at the beginning of the response phase). In Experiment 1b, the starting expression was always set outside of the distribution of presented expressions. Each experiment provided a different way to account for potential anchoring effects.

## Procedure

In each trial, participants were presented with an expression on the top left side of the screen for 2 s. Presented expressions ranged from a mildly happy expression (26) to Happy (35). A buffer of six expressions was left unpresented on the neutral end of the distribution and a buffer of five expressions on the happy end to give participants room for error at each end. Expressions 0 through 20 represented an unpresented sad-to-neutral range (which are examined in Experiments 3 and 4). Next, the stimulus was removed and the screen was blank for either 500 ms or 4 s. Following this, a response face was shown in the middle of the screen. Participants were instructed to "use the right and left arrow keys to change the expression of the face to match the expression of the previous photograph." Clicking the right arrow key made the expression cycle through happy then neutral then sad expressions whereas clicking left cycled neutral-sad-happy. After 40 trials were completed, participants were asked to fill out the demographic questionnaire, and then were provided with a completion code, signaling the end of the experiment.

## Data preparation

We computed a bias score for each trial as the difference between the participant's recalled expression and the stimulus expression. Positive bias indicates that the expression was remembered as happier than the stimulus; negative indicates that it was biased towards sad. We also took the absolute value of bias scores as a measure of error. Prior to analysis, we decided that participants with average error scores over all trials greater than 2.5 *SD*s from the average of each participant's average error would be culled. In Experiment 1a, the average for each participant's error (absolute value of bias) was 2.53 (*SD* = 0.71), which means our line for culling was 4.3. One participant was removed, with an average error above this line (*M* = 4.41). Furthermore, individual trials greater than 2.5 *SD*s above the overall mean error (*M* = 2.5, *SD* = 2.23) were also removed. This resulted in removing observations that had an error greater than 8.1: 41 out of 2,280 total observations (or 1.8% of the sample). For Experiment 1b, two participants were removed for having average bias scores greater than the 2.5 *SD*s of the mean average error ($M_{Error}$ = 2.74, $SD_{Error}$ = 0.90, Cutoff = 4.98). Forty-one observations out of 2,520 (1.6% of the sample) were removed for being greater than 2.5 *SD*s over the mean bias across all observations.

## Results

All analyses were conducted using the R software package (Core Team 2012). We conducted a linear mixed model with Stimulus Expression and the interaction of Stimulus Expression and Delay predicting bias in order to determine the direction of memory bias across expressions and Delay. We also included the Block, Gender, and Starting Expression factors in the analysis as well as a Block × Expression interaction term (and including participant as a random variable; see Tables 1 and 2).[3] The analysis was conducted using the lme4 (Bates, Maechler, Bolker, & Walker, 2015) and lmertest (Kuznetsova, Brockhoff, & Christensen, 2016) R packages.

The results showed a significant bias toward the center of the distribution even when controlling for the starting value of the response face, suggesting that this bias is not due only to an anchoring effect. Analyses revealed significant Expression × Delay interactions, with a larger central tendency bias for a longer delay (see Fig. 1a and b).

The central tendency bias was also stronger for the second half of the study (though this interaction missed significance for Experiment 1a). For Experiment 1a, we also found a

---

[3] The structure of the model was determined based on analyses of pilot data (see osf.io/2wuz3).

**Table 1**  Predicting memory bias in Experiment 1a

| Predictors | B | 95% CI | SE B | t | p |
|---|---|---|---|---|---|
| **Fixed** | | | | | |
| Intercept | -0.11 | −2.29, 2.06 | 1.11 | −0.10 | .918 |
| Expression | −0.12 | −0.18, −0.05 | 0.03 | -3.33 | .001 |
| Delay (4s) | 2.67 | 0.26, 5.08 | 1.23 | 2.17 | .030 |
| Block (2) | 2.06 | −0.35, 4.47 | 1.23 | 1.67 | .094 |
| Gender (Male) | 0.25 | 0.03, 0.48 | 0.12 | 2.19 | .029 |
| Start Expression | 0.13 | 0.11, 0.15 | 0.01 | 13.03 | .001 |
| Exp X Del | −0.08 | −0.16, −0.00 | 0.04 | −2.00 | .046 |
| Exp X Block | −0.08 | −0.16, 0.00 | 0.04 | −1.94 | .053 |
| **Random** | | | | | |
| $\tau_{00, \text{Participant}}$ | 0.59 (0.77) | | | | |
| $\sigma^2$ | 7.40 (2.72) | | | | |
| $R^2 / \Omega_0^2$ | .19/.19 | | | | |

*Note.* Betas for Exp X Del and Exp X Block are identical due to rounding to the nearest hundredth. Exp. = Expression (26 (Mildly Happy) to 35 (Happy)), Del. = Delay (.5 s, 4 s). Standard errors for random effects are in parenthesis

significant anchoring effect, such that memory was biased toward the starting value of the response expression.

We also examined the effect of the actor's gender on bias. Although we only varied gender for the purpose of generalizability, we recognized that reporting effects of gender may be useful for other researchers. As shown in Tables 1 and 2, there was a significant main effect of gender, with male actor's faces biased happier in memory as compared to females for both Experiments.

**Table 2**  Predicting memory bias in Experiment 1b

| Predictors | B | 95% CI | SE B | t | p |
|---|---|---|---|---|---|
| **Fixed** | | | | | |
| Intercept | 1.97 | −0.00, 3.94 | 1.01 | 1.96 | .051 |
| Expression | −0.09 | −0.15, −0.03 | 0.03 | −2.80 | .005 |
| Delay (4s) | 2.78 | 0.53, 5.03 | 1.15 | 2.42 | .016 |
| Block (2) | 7.75 | 5.50, 10.01 | 1.15 | 6.75 | .001 |
| Gender (Male) | 0.46 | 0.25, 0.67 | 0.11 | 4.25 | .001 |
| Exp X Block | −0.08 | −0.15, −0.01 | 0.04 | −2.17 | .030 |
| Exp X Rep | −0.25 | −0.32, −0.17 | 0.04 | −6.59 | .001 |
| **Random** | | | | | |
| $\tau_{00, \text{Participant}}$ | 1.47 (1.21) | | | | |
| $\sigma^2$ | 7.11 (2.67) | | | | |
| $ICC_{\text{Participant}}$ | 0.17 | | | | |
| $R^2 / \Omega_0^2$ | .25/.25 | | | | |

*Note.* Exp. = Expression (26 (Mildly Happy) to 35 (Happy)), Del. = Delay (.5 s, 4 s). Standard errors for random effects are in parenthesis
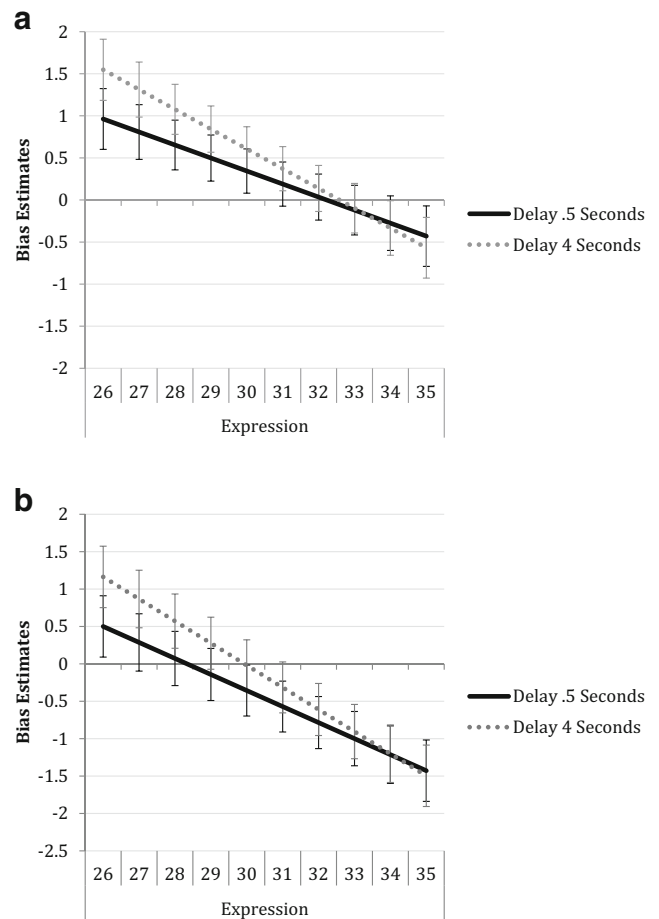
**Fig. 1**  **a** Experiment 1a bias estimates (Response Expression – Stimulus Expression) predicted by which Expression was shown (Expression 26 (Mildly Happy) to 35 (Happy)) and whether the delay between Stimulus and Response was .5 s or 4 s. The slope for the longer delay is steeper, indicating a larger central tendency effect as compared to the shorter delay. *Error bars* are 95% CIs. **b** Experiment 1b bias estimates (Response Expression – Stimulus Expression) predicted by which Expression was shown (Expression 26 (Mildly Happy) to 35 (Happy)) and whether the delay between Stimulus and Response was .5 s or 4 s. The slope for the longer delay is steeper, indicating a larger central tendency effect as compared to the shorter delay. *Error bars* are 95% CIs

## Discussion

We sought to determine whether central tendency effects in spatial memory occurred for emotional materials – in this case, faces expressing varying degrees of happiness. An alternative hypothesis is that remembered expressions would be more prototypical of the emotion category represented (i.e., faces categorized as happy would be recalled as happier than they actually were). Results show support for a central tendency bias for these types of materials. This suggests that recall of stimuli requiring attention to multiple interacting visual features (i.e., an expression) is prone to the same biases in memory that affect the recall of simpler stimuli such as line lengths. Increased bias with a longer delay between

stimulus and recognition is consistent with Bayesian memory models which predict that, as memory becomes less precise, category information will be relied on more (Crawford et al., 2000). This result mirrors prior work showing that the central tendency bias is larger for individuals put under cognitive load (Allred et al., 2016) or delay (Olkkonen et al., 2014).

Experiment 1a revealed a significant anchoring effect (i.e., responses were biased towards the initial expression), which did not account for any of the central tendency effects. Furthermore, by placing the starting expression completely outside the distribution, Experiment 1b confirmed that central tendency effects for these materials were independent from any effects of anchors.

The finding that the central tendency bias was larger in the second half of the study indicates that participants relied more on the inductive category over time, something not addressed in prior studies of central tendency bias. This could be due to the fact that the complex materials required a significant number of trials to form an accurate category, or possibly overcome an alternate prior category (e.g., an emotion expression prototype). The increased bias would not be expected if participants were becoming better at encoding perceptual details of the expressions over time.

## Experiment 2

Thus far, under a number of varying conditions, we have found that individuals show a central tendency bias in memory for emotional faces, which are more complex than the stimuli used in prior studies. We have also found evidence to support that this is indeed a memory phenomenon in line with Bayesian models of reconstructive memory, given that delays exacerbate the bias (Allred et al., 2016; Crawford et al., 2000). However, up to this point, we have only examined the neutral-to-happy range of expressions. Therefore, in Experiment 2, we determine whether this effect extends to a distribution of sad emotional expressions.

### Participants

Sixty-nine participants were recruited from Amazon Mechanical Turk. Nine of these participants failed to complete the study, leaving the sample at 60.

### Materials and design

The materials and design were identical to Experiment 1b, except that the distribution of expressions ranged from expression 6 (seven expressions from the saddest (which was expression 0)) to expression 15 (five expressions from neutral

(expression 20)). As in Experiment 1b, the start value for the expressions was set at neutral.

### Data preparation

Three participants were removed for having average bias scores greater than the 2.5 $SDs$ of the mean average error ($M_{Error} = 4.15$, $SD_{Error} = 0.96$, Cutoff = 6.56). Forty-eight observations out of 2,280 (2.1% of the sample) were removed for being greater than 2.5 $SDs$ over the mean bias ($M_{Bias} = 4.01$, $SD_{Bias} = 3.24$, Cutoff = 12.10) across all observations.

## Results and discussion

As shown in Fig. 2, the central tendency bias was found for the sad range of expressions. This effect was stronger when the delay between presentation of the stimulus and recognition increased. As in the prior Experiments, the effect was also stronger for trials in the second half of the experiment (see Table 3). Finally, male faces in this study were remembered as less sad as compared to female faces.

## Experiment 3

Our final experiment examines memory performance when the distribution of expressions cross emotional categories. In this case we show participants a distribution of expressions ranging from mildly sad to neutral to mildly happy. This approach may encourage participants to focus on contrasting emotion categories, therefore encouraging bias toward emotion expression prototypes as opposed to inductive categories.
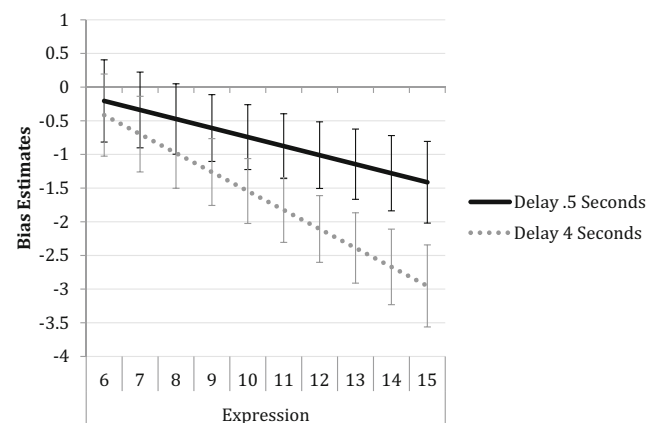


Fig. 2 Experiment 2 bias estimates (Response Expression – Stimulus Expression) predicted by which Expression was shown (Expression 6 (Sad) to 15 (Mildly Sad)) and whether the delay between Stimulus and Response was .5 s or 4 s. The slope for the longer delay is steeper, indicating a larger central tendency effect as compared to the shorter delay. *Error bars* are 95% CIs

**Table 3**  Predicting memory bias in Experiment 2

| Predictors | B | 95% CI | SE B | t | p |
|---|---|---|---|---|---|
| Fixed | | | | | |
| Intercept | −0.99 | −2.19, 0.22 | 0.61 | −1.61 | .108 |
| Expression | 0.01 | −0.09, 0.12 | 0.05 | 0.27 | .807 |
| Delay (4s) | 0.67 | −0.62, 1.97 | 0.66 | 1.04 | .308 |
| Block (2) | 2.51 | 1.22, 3.81 | 0.66 | 3.80 | .001 |
| Gender (Male) | 0.66 | 0.31, 1.00 | 0.17 | 3.77 | .001 |
| Exp X Del | −0.15 | −0.27, −0.03 | 0.06 | −2.45 | .015 |
| Exp X Rep | −0.29 | −0.41, −0.18 | 0.06 | −4.85 | .001 |
| Random | | | | | |
| $\tau_{00,\ Participant}$ | 2.40 (1.55) | | | | |
| $\sigma^2$ | 16.92 (4.11) | | | | |
| $R^2 / \Omega_0^2$ | .18/.18 | | | | |

*Note.* Exp. = Expression (6 (Sad) to 15 (Mildly Sad)), Del. = Delay (.5 s, 4 s). Standard Errors for random effects are in parenthesis

In other words, a slightly sad face within a distribution of sad and happy expressions may be remembered as sadder than it actually was, because the participant is relying on the category of sad as a cue in memory.

## Participants

Sixty-four participants were recruited from Amazon Mechanical Turk. Five of these participants failed to complete the study, leaving the sample at 59.

## Materials and design

The materials and design were identical to Experiment 1b, except that the distribution of expressions ranged from expression 15 to expression 25, excluding expression 20 (neutral), for a total of ten expressions (five sad, five happy). The starting value in this experiment was either set at 10 or 30 (five expressions outside of the distribution at both ends), and was a between-participant factor in this experiment, meaning that participants would either always move from happy to sad or vice-versa to recreate presented expressions.

## Data preparation

Four participants were removed for having average bias scores greater than 2.5 *SD*s of the mean average error ($M_{Error}$ = 3.55, $SD_{Error}$ = 1.25, Cutoff = 6.66). Sixty-four observations out of 2,320 (2.8% of the sample) were removed for being greater than 2.5 *SD*s over the mean bias ($M_{Bias}$ = 3.43, $SD_{Bias}$ = 3.13, Cutoff = 11.26) across all observations.

## Results and discussion

Even with expressions crossing emotional categories, participants showed a bias toward the center of the distribution. As in prior experiments, the central tendency bias was more pronounced in trials with the longer delay (see Fig. 3) and the bias was also larger in the second half of the experiment (see Table 4).

Also, similar to Experiment 1a, we found an anchoring effect, such that participants whose response expression started on the happy end of the distribution were biased towards happier expressions, and those starting on the sad end were biased toward sadness. In contrast to the results in Experiment 2, there was a significant main effect of gender such that male faces were biased sadder as compared to female faces.

## General discussion

It is well known in the memory literature that individuals tend to bias their memories in toward category centers (e.g., Huttenlocher et al., 2000). Here we examined the sources of category structures people would use to reconstruct memory for emotional facial expressions. Prior research on memory for emotional expressions suggests that under certain conditions, individuals tend to rely on pre-existing categories of emotions, leading them to recall expressions as more reflective of the emotion category than they actually were (mildly happy faces are recalled as happier). However, research has also shown that given a distribution of attributes (i.e., line lengths), people will use their knowledge of that distribution to reconstruct memories, leading estimates to be biased towards the center of the
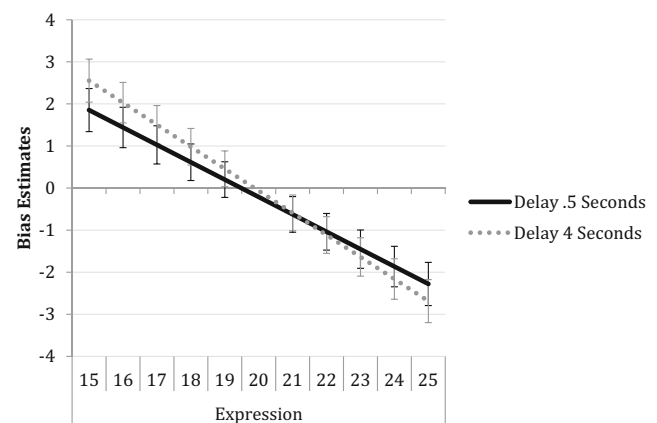


**Fig. 3** Experiment 3 bias estimates (Response Expression – Stimulus Expression) predicted by which Expression was shown (Expression 15 (Mildly Sad) to 15 (Mildly Happy); Expression 20 was not presented) and whether the delay between Stimulus and Response was .5 s or 4 s. The slope for the longer delay is steeper, indicating a larger central tendency effect as compared to the shorter delay. *Error bars* are 95% CIs

**Table 4** Predicting memory bias in Experiment 3

| Predictors | B | 95% CI | SE B | t | p |
|---|---|---|---|---|---|
| Fixed | | | | | |
| Intercept | 6.13 | 4.56, 7.69 | 0.80 | 7.68 | .001 |
| Expression | −0.34 | −0.41, −0.27 | 0.04 | −9.16 | .001 |
| Delay (4s) | 2.39 | 0.67, 4.05 | 0.86 | 2.73 | .006 |
| Block (2) | 3.27 | 1.60, 4.99 | 0.86 | 3.81 | .001 |
| Gender (Male) | −0.29 | −0.57, −0.02 | 0.14 | −2.08 | .038 |
| Start Expression (10) | 0.85 | 0.08, 1.61 | 0.39 | 2.17 | .034 |
| Exp X Del | −0.11 | −0.19, −0.03 | 0.04 | −2.60 | .009 |
| Exp X Block | −0.15 | −0.23, −0.06 | 0.04 | −3.41 | .001 |
| Random | | | | | |
| $\tau_{00,\ Participant}$ | 1.91(1.38) | | | | |
| $\sigma^2$ | 11.20 (3.35) | | | | |
| $R^2 / \Omega_0^2$ | .31/.31 | | | | |

*Note.* Exp. = Expression (15 (Mildly Sad) to 25 (Mildly Happy) – Expression 20 was skipped), Del. = Delay (.5 s, 4 s), and Start Expression (10 (Sad Expression), 35 (Happy Expression)). Standard errors for random effects are in parenthesis

distribution. Given the social relevance and complexity of emotional expressions as compared to stimuli such as line lengths, we sought to determine whether given a distribution of expressions, memory would be biased towards the emotion category, or toward the inductive (i.e., central) category.

Results from 4 experiments showed that – rather than biasing towards an emotion category when given a distribution of expressions within and/or between emotional categories – memory was biased towards the center of the distribution. Furthermore, the bias was most prominent in the second half of each experiment (after all expressions had been shown once). This suggests that perhaps due to the visual complexity of the expressions category formation took longer for these materials as compared to simpler stimuli such as line lengths. An alternative hypothesis may be that pre-existing emotional categories competed with inductive categories during the task. The central tendency bias was also more pronounced when a longer delay was inserted between the presentation of the expression and the recall task, suggesting that this is a bias in memory rather than perception. This effect persisted regardless of whether the distribution ranged within a single emotional category or across categories.

Here we have assumed that participants encode the emotional variation between stimuli, but the stimuli vary on physical dimensions as well, such as mouth shape and size. This suggests the possibility that participants learned the distribution of these physical properties rather than or in addition to the valence distribution. Bayesian models have not addressed this distinction, but an emotion-based interpretation is consistent with prior theoretical work in the false memory literature.

Work testing fuzzy-trace theory predictions showed that valence is encoded as a gist (a memory representation that captures meaning), and that presenting words that share valence increases susceptibility to false memory by cuing feelings of similarity (Brainerd, Stein, Silveira, Rohenkohl, & Reyna, 2008). In another study, participants were shown clips of happy or sad people who were not indicating these emotions with prototypical facial expressions expected for each emotion. In a subsequent recognition task, participants falsely identified prototypical expressions as the ones seen in the video clips (Fernandez-Dols, Carrera, Barchard, & Gacitua, 2008). When examining emotional expressions, participants encoded the emotional semantic information even when the perceptual corollary was absent.

The results of these studies provide the groundwork for future work to examine the formation of socially and emotionally relevant categories and their influence on memory for individuals. Whereas here participants learned the distribution through many individual encounters, other research has examined how participants can quickly and easily ascertain the average level and type of emotion when briefly shown a group of individuals simultaneously (Haberman & Whitney, 2007; Neumann, Schweinberger, & Burton 2013). It is not known if summary representations formed in this way will give rise to the same kinds of biases or if this bias depends on the serial reproduction paradigm. In addition, further studies will examine the impact of social stereotypes on face category learning and memory for expressions.

The memory literature is rife with examples of predictable inaccuracy. We mistakenly recall items that are most representative of a semantic category. We bias recollection of a point away from a boundary in space. We recognize objects as more representative of their categories in many domains (e.g., size, length, hue) than they actually were. This article expands the scope of this work, showing that inductive category effects in memory occur for complex, emotional stimuli. Rather than memory biasing towards the best representation of an emotional category, we demonstrated that it could bias towards an expression that is more ambiguous, yet better representative of a group's average expression.

## 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, 1,* 1–7. doi:10.3758/s13423-016-1039-0

Ashourian, P., & Loewenstein, Y. (2011). Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS ONE, 6,* e19551. doi:10.1371/journal.pone.0019551

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67,* 1-48. doi: 10.18637/jss.v067.i01

Brainerd, C. J., Stein, L. M., Silveira, R. A., Rohenkohl, G., & Reyna, V. F. (2008). How does negative emotion cause false memories? *Psychological Science, 19*(9), 919–925.

Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in "associative" false memory. *Psychonomic Bulletin & Review, 15,* 1035–1053. doi:10.3758/PBR.15.6.1035

Core Team, R. (2012). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing.

Crawford, L. E., Huttenlocher, J., & Engebretson, P. H. (2000). Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science, 11,* 280–284. doi:10.1111/1467-9280.00256

Crawford, L. E., Landy, D., & Salthouse, T. A. (2016). Spatial working memory capacity predicts bias in estimates of location. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* doi:10.1037/xlm0000228

Duffy, S., Huttenlocher, J., Hedges, L. V., & Crawford, L. E. (2010). Category effects on stimulus estimation: Shifting and skewed frequency distributions. *Psychonomic Bulletin & Review, 17,* 224–230. doi:10.3758/PBR.17.2.224

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review, 116,* 752–782. doi:10.1037/a0017196

Fernández-Dols, J. M., Carrera, P., Barchard, K., & Gacitua, M. (2008). False recognition of facial expressions of emotion: Causes and implications. *Emotion, 8,* 530–539. doi:10.1037/a0012724

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17,* R751–R753. doi:10.1016/j.cub.2007.06.039

Halberstadt, J. B., & Niedenthal, P. M. (2001). Effects of emotion concepts on perceptual memory for emotional expressions. *Journal of Personality and Social Psychology, 81,* 587–598. doi:10.1037/0022-3514.81.4.587

Hemmer, P., & Steyvers, M. (2009). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin & Review, 16,* 80–87. doi:10.3758/PBR.16.1.80

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General, 129,* 220–241. doi:10.1037/0096-3445.129.2.220

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience, 13,* 1020–1026. doi:10.1038/nn.2590

Kuznetsova, A., Brockhoff, P. B, & Christensen, R. H. B. (2016). lmerTest: Tests for random and fixed effects for linear mixed effect. R package version 2.0-30. Available at: https://CRAN.R-project.org/package=lmerTest

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition, 128,* 56–63. doi:10.1016/j.cognition.2013.03.006

Olkkonen, M., & Allred, S. R. (2014). Short-term memory affects color perception in context. *PLoS One, 9,* e86488. doi:10.1371/journal.pone.0086488

Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision, 14,* 1–15. doi:10.1167/14.11.5

Roberson, D., Damjanovic, L., & Pilling, M. (2007). Categorical perception of facial expressions: Evidence for a "category adjustment" model. *Memory & Cognition, 35,* 1814–1829. doi:10.3758/BF03193512

Simmering, V. R., Spencer, J. P., & Schöner, G. (2006). Reference-related inhibition produces enhanced position discrimination and fast repulsion near axes of symmetry. *Perception & Psychophysics, 68,* 1027–1046. doi:10.3758/BF03193363