The average facial expression of a crowd influences impressions of individual expressions

Sarah Griffiths[1], Gillian Rhodes[2], Linda Jeffery[2], Romina Palermo[2], and Markus F. Neumann[2]

[1]School of Experimental Psychology, University of Bristol

[2]ARC Centre of Excellence in Cognition and its Disorders, School of Psychology, University of Western Australia

Author Note

Sarah Griffiths, School of Experimental Psychology, University of Bristol, UK; Gillian Rhodes, Linda Jeffery, Romina Palermo, and Markus F. Neumann, ARC Centre of Excellence in Cognition and its Disorders, School of Psychology, The University of Western Australia, Western Australia, Australia.

Sarah Griffiths is now at the Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, CB2 8AH, UK.

This research was supported by the Australian Research Council Centre of Excellence in Cognition and its Disorders (CE110001021), an ARC Discovery Outstanding Researcher Award to Rhodes (DP130102300), and a Worldwide University Network Researcher Mobility Award to Griffiths. We thank Matthew Robson for collecting data for Experiment 2.

Correspondence concerning this article should be addressed to Sarah Griffiths, Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, CB2 8AH, UK. E-mail: slg68@medschl.cam.ac.uk

Abstract

People can accurately assess the "mood of a crowd" by rapidly extracting the average intensity of all the individual expressions, when the crowd consists of a set of faces comprising different expressions of the same individual. Here, we investigate the processes involved when people judge the expression intensity of individual faces that appear in the context of a more naturalistic crowd of different individuals' faces. We show that judgments of the intensity of happy and angry expressions for individual faces are biased towards the group mean expression intensity, even when the faces are all different individuals. In a second experiment, we demonstrate that this bias is not due to a generic tendency to endorse intermediate intensity expressions more frequently than more extreme intensity expressions. Together, these findings suggest that people integrate ensemble information about the group average expression when they make judgments of individual faces' expressions.

Keywords: expression, context, crowd, ensemble coding, summary statistics

Word count: 5291 (excluding references, figure captions and footnote)

Statement of Public Significance

Most studies testing recognition of emotion from facial expressions show faces in isolation. However, faces are often seen in groups, and in this study we showed groups of different people whose faces varied in emotional intensity. We found that the facial expression displayed by an individual is seen as a combination of their actual expression and the average expression of the entire group (known as the ‘ensemble’ expression). For example, the face of a mildly angry person seen as part of a crowd of other angrier people is judged as more angry than it actually was. This means that both individual and ensemble information are used when judging an individual’s emotion. On a more general level, our study shows how our processing of individual expressions is malleable and can be informed by the context, such as surrounding faces in a crowd.

Introduction

Social situations often require us to judge other people's emotional states. One very important source for this information is their facial expressions. However, the perception of facial expressions is malleable, and can be strongly influenced by the situational context (for review, see Barrett, Mesquita, & Gendron, 2011). Viewing a single facial expression in an emotionally congruent scene, or paired with an emotionally congruent object, body posture or word, can facilitate processing of the expression, while viewing the same face paired with emotionally incongruent stimuli can impair it (Aviezer et al., 2008; Aviezer, Trope, & Todorov, 2012; Halberstadt & Niedenthal, 2001; Ngo & Isaacowitz, 2015; Righart & de Gelder, 2008)

The presence of other faces might also change how we process individual facial expressions, because the visual system utilizes a specialized process that computes the average properties of sets of similar objects. This "ensemble coding" is well-established for simple objects (often examined with sets of circles varying in size), and recent studies have suggested that it might also be involved in coding properties of sets of faces including facial expression (Haberman & Whitney, 2012), attractiveness (Walker & Vul, 2013), sex (Haberman & Whitney, 2007), and facial identity (de Fockert & Wolfenstein, 2009; Leib et al., 2014; Neumann, Schweinberger, & Burton, 2013). Ensemble coding is one possible means by which information can be compressed into more abstractive statistical representations in visual working memory, in order to deal with the constant stream of information from our environment (Alvarez, 2011). Ensemble coding of facial expressions is also assumed to be advantageous in social situations, as it could allow us to quickly read "the mood of a crowd" at a glance, maybe without having to attend to each individual face separately (Haberman & Whitney, 2009).

Previous studies on ensemble coding of expression have predominantly focused on the ability to form average representations under rather artificial conditions. Specifically, ensemble coding of expression has almost exclusively been examined for sets of faces containing *one single identity* (Haberman & Whitney, 2007, 2009; Leib et al., 2014; Leib et al., 2012). Faces in these sets were sampled from morph continua between two different emotional expressions (e.g., happiness to sadness; neutral to disgust) displayed on the same individual's face. These studies revealed that people are able to spontaneously and accurately code the average expression of these kinds of sets.

To our knowledge, only one study has investigated expression coding for naturalistic crowds consisting of different people's faces (Yang, Yoon, Chong, & Oh, 2013). Yang and colleagues presented crowds of different individuals, some with happy expressions and some with angry expressions (in different ratios) and participants had to report whether the crowd was overall "positive" or "negative" in emotion. All participants were able to accurately judge the overall mood of the crowd (although participants with low social anxiety exhibited a small bias towards positive emotion). One possible explanation for the accurate judgments of the overall emotion is that participants had used ensemble coding to determine the *average* expression (that is, the ensemble expression) of the set, and then judged whether this average expression was positive or negative. However, an alternative possibility is that participants were able to judge the relative number of faces belonging to each emotion category in the set. Accurate judgments of the relative numerosity of two large groups of items can be made under limited presentations times (Anobile, Cicchini, & Burr, 2016). As all the expressions were full-blown and the two categories of emotion (happy and angry) have very distinct facial features, rapidly distinguishing the two categories would be relatively easy. Judgements of the ratio of happy to angry expressions might be possible without integration of the expressions into an ensemble. The use of ratio judgements could be confined to groups in which faces can be easily assigned to different categories such as angry and happy, because they exhibit full-blown expressions. However, ratio judgements would be less applicable for naturalistic groups of faces that are less easy to categorize, for instance when they all show a single expression with different intensities.

Thus, a novel approach to answering the question of whether ensemble expression coding occurs for groups of different identities would be to test whether memory for the expression intensity of an individual face is systematically influenced by the ensemble expression of the entire group. Ensemble coding can influence perception of basic features (such as orientation of Gabor patches) by making the features of individual group members look more similar to the group average (Ross & Burr, 2008). Ensemble representations can also bias short-term memory for the individual members towards the mean property of a group (Brady & Alvarez, 2011). Specifically, when judging the size of individual circles, participants' responses are slightly biased towards the mean size of the group. When colour is a distinguishing feature between several different groups of circles, participants are not only biased towards the mean size of all circles, but also to the mean size of the group of circles that had the same colour as the to-be-remembered circle. Brady and colleagues (2011) concluded that information about the size of items in a group is represented on multiple levels of abstraction (e.g., individual circle, same-coloured circles, all circles), and integrated across the different levels during encoding or retrieval. Thus, ensemble representations play an important role for the processing of the individual items in a group.

As yet, it is unclear whether a systematic influence of the ensemble on exemplar information, indicating the integration of ensemble and exemplar information, also occurs for high-level information about facial properties such as expression or identity. Previous work has established that individual and ensemble representations of facial identity can be extracted and stored simultaneously (Neumann et al., 2013), but did not examine whether these representations interact. Some evidence for a possible integration of average and individual information comes from other work by Sweeny and colleagues (2009), which have shown that interpretation of the emotion on a face can be influenced by the expression on another face that is seen at the same time. Very briefly presented valence-neutral faces were rated as more positive when paired with a happy face than when paired with an angry face (Sweeny, Grabowecky, Paller, & Suzuki, 2009). This difference was only found when both faces were shown in the same hemifield, but not when the two faces were in opposite hemifields. The authors suggested that this effect is due to a perceptual mechanism that averages expression information within, but not across, receptive fields. However, it is unclear to what extent such a mechanism is related to - or contributes to - ensemble coding, which has been consistently demonstrated for arrays of stimuli that cover different receptive fields (e.g., are presented in both hemifields).

In the present study, we ask whether integration of ensemble and individual information occurs for the expressions of a group of *different identity* faces, by determining whether the ensemble expression influences memory for the expression of individual "target" faces in a group, as has been demonstrated for the sizes of circles (Brady & Alvarez, 2011). We adapted a well-established membership identification paradigm, in which participants are asked whether a probe was a member of the preceding set (e.g., Ariely, 2001; Brady & Alvarez, 2011; Haberman & Whitney, 2009; Walker & Vul, 2013). Membership identification paradigms encourage individuation of the set members, and ensemble coding is inferred from incorrect endorsements of the group average. Endorsements of the group average occur frequently, suggesting that participants engage in ensemble coding even when the task encourages individuation. In our study, participants determined whether a subsequent probe face had the same expression intensity or a different expression intensity compared to a target face. Probe expressions could either be the same intensity as the target, closer to, or further away from, the group's average expression intensity.

If participants code the ensemble expression from groups of different identities, as well as the individual expressions of the group, then we expect that this ensemble expression will influence memory for individual expressions in a group (as shown for the sizes of a group of individual circles, Brady & Alvarez, 2011), which should be systematically shifted towards the *average intensity* of the group. In this case, participants would be more likely to report that they had seen probes with expression intensities that were closer to the group mean, than those that were further away from it. An alternative possibility is that the ensemble expression is either not coded or does not influence memory for the individual expression. In this case, participants would not exhibit a bias towards the average group expression. Finally, it is possible the participants code the ensemble expression, but not the individual group expressions (Haberman & Whitney, 2009). In this case, participants would endorse probe expressions that match the average expression of the group, independent of what expression the individual target face had. We will refer to these alternatives as "Exemplar biased by ensemble", "Exemplar only", and "Ensemble only", respectively.

Experiment 1 - Methods

Participants

Twenty-four students and staff from the University of Western Australia were recruited (mean age = 23.00, SD = 5.08 years; 9 male). Student participants received either course credits or compensation of $5AUD for their time. Sample size was based on related studies (Brady & Alvarez, 2011; Walker & Vul, 2013).

Stimuli

Three images (happy, angry, and neutral) of four young male Caucasian identities were sourced from the Radboud face database (Langner et al., 2010). We selected individuals for which agreement was high regarding the expression displayed (>95% agreement) and which were rated as relatively intense (intensity >3.5, max = 5). We created "weaker" intensity levels for each emotion by morphing each of the original (100%) happy or angry faces with the neutral face of the matching identity, using FantaMorph 5 (Abrosoft, http://www.fantamorph.com/). The full intensity range consisted of 11 steps, including the 100% emotional and neutral (0%) faces for each identity. All face images were transformed into grey scale, adjusted so that their pupils were horizontally aligned, and placed in a mask that covered external face features and hair.

Study groups consisted of four faces (one of each identity) arranged in a 2 x 2 grid. Grid positions were randomly assigned to each identity on each trial. Faces subtended a visual angle of approximately 3.5° x 4.0° with the total grid subtending 8.4° x 9.2° when viewed from a viewing distance of about 65 cm. Each face in a group displayed the same emotion (either all angry or all happy), but faces varied systematically in expression intensity around a "group mean" intensity level which itself was never shown in the study set. Each study set contained two expressions that were more intense than the group mean (+10%,

+20%), and two expressions that were less intense than the group mean (-10%, -20%). On each trial the group mean intensity was randomly selected, with the only restriction being that all faces presented fell within the range of the 11-step expression intensity sequence. The group mean was therefore always between 20%-80%, whereas the target and probe intensities could occupy any position in the 11-step sequence (0%-100%), depending on the relative positions of the target and probe.

Procedure

On each trial, participants were presented with a fixation cross in the centre of the screen for 500 ms, followed by the study group for 2000 ms (Fig. 1a). This study group duration matched that used in previous studies on expression ensemble coding (e.g. Haberman et al., 2009). Participants were instructed to remember the expression intensity for each of the four faces as accurately as possible. Immediately after the study group had disappeared, a cue (a black frame 3.8° x 5.1° VA) appeared for 200 ms at the position of a “target” identity. A single probe face (4.9° x 5.5° VA) of the same identity as the cued target was then presented in the centre of the screen. Participants indicated whether they believed the probe face had the same or a different intensity expression as the target face by pressing “a” for “same”, and “l” for “different”, respectively (keys were labelled with response options). The probe face remained on the screen until the participant made a response.

Participants completed two blocks of 160 trials (320 trials in total) with a break between blocks. Each block contained one trial for all possible combinations of the four target face identities, four target expression intensities (+10%, +20%, -10%, -20% relative to group mean intensity), five probe expression intensities (0%, +10%, +20%, -10%, and -20% relative to target intensity), and two emotions (happy, angry). The whole session lasted approximately 25 minutes. To familiarise participants with the trial procedure, they were given three happy and three angry practice trials in which each of the four identities was shown at least once before the experiment proper. No feedback was provided.

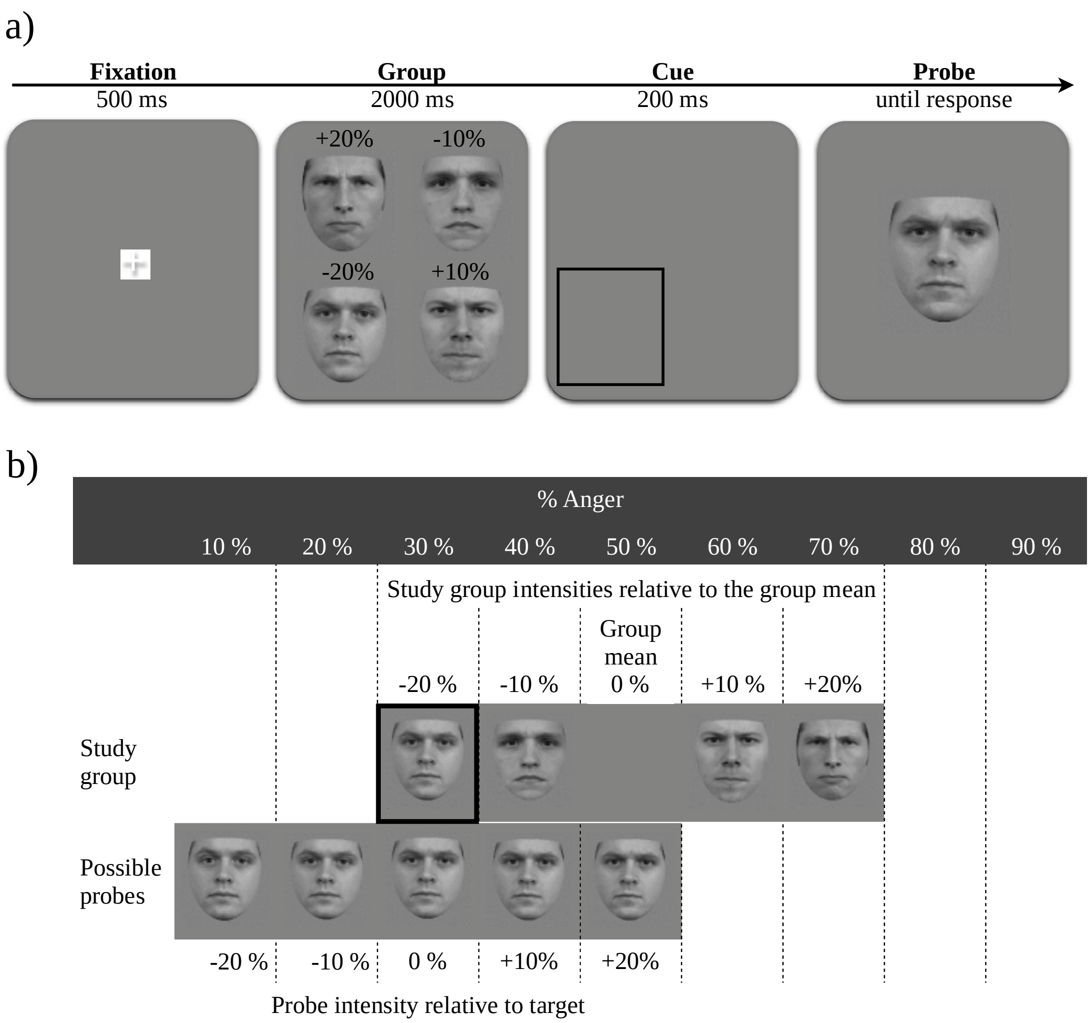

Figure 1. (a) Trial procedure. The intensity values give the intensity of the expressions in the group relative to the group mean, with positive percentage-values indicating higher intensities than the mean, and negative values indicating lower intensities than the mean. Note that these values were not seen by the participants in the experiment. (b) Schematic of the expression intensity range of the example study group, and the expression intensity range of the five possible probes that could follow the example target (here: 30% anger, in black frame). Participants reported whether or not the expression intensity of the probe face was the "same" or "different" than that of the target.

Possible outcomes

Fig. 2 illustrates the three alternative possible outcomes outlined in the introduction. Panel a) shows the expected data pattern if individual expressions were coded accurately, and responses were not affected by the group mean intensity (“exemplar only” outcome, panel a). In this case, “same” responses should occur most frequently for probes that match the target intensity and errors should be normally distributed around the target intensity (that is, errors should be independent of the intensity of the probe relative to the group mean intensity). If however only the group mean intensity was coded (ensemble “only”, panel b), “same” responses should be most frequent for probes that correspond to the group mean intensity, and responses should not be affected by the target intensity (c.f., Haberman & Whitney, 2007). Finally, if participants’ responses reflected a combination of representations for individual target intensities and the group mean intensity (“Exemplar biased by ensemble”, panel c), then “same” responses should be most frequent for probes with expression intensities that lie somewhere between the target intensity and mean group intensity. Furthermore, participants’ error distributions would be skewed towards the group mean intensity, such that errors should be more frequent for mismatch probes that deviate *towards* the group mean intensity, than to mismatch probes that deviate *away* from the group mean intensity. More specifically, for targets of a higher intensity than the mean (+10%, +20%), “same” responses should be more frequent for probes that are less intense than the target, compared to probes that are more intense than the target. Conversely, for targets of lower intensity than the mean (-10, -20%), “same” responses should be more frequent for probes that are of more intense than the target, compared to probes that are less intense than the target.

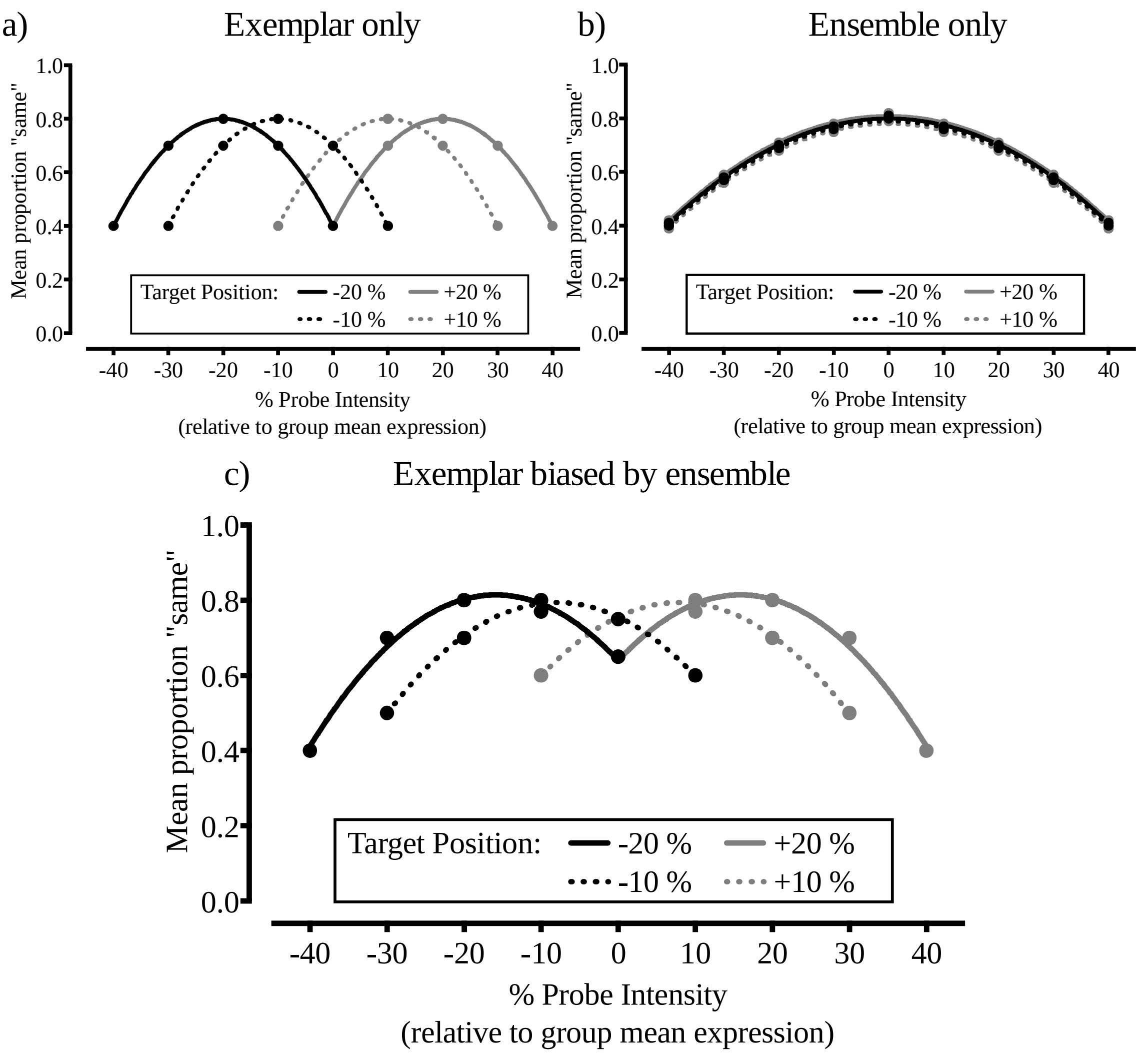

Figure 2. Proportions of “same” responses to probes plotted as a function of probe intensity plotted relative to the group mean expression **(e.g.,** for a -20% target (black solid line), the five levels of probe intensity relative to target intensity (-20%, -10%, 0%, +10%. +20%) are depicted as -40%, -30%, -20%, -10%, and 0%, respectively**)**. Three possible outcomes are shown: (a) *Exemplar only*: Separate curves are plotted for each target position (-20%, -10%, +10%, +20%). Proportion “same” responses is highest for probes that match the target intensities, and errors are independent of the group mean expression (responses are normally distributed around each target intensity); (b) *Ensemble only*: Proportion “same” responses is highest for the group mean expression; (c) *Exemplar biased by ensemble*: Proportion “same”

responses are distributed around the target intensity but skewed towards the group mean intensity.

Results

Figure 3 shows participants' "same" responses as a function of the probe intensity relative to the group mean intensity. Inspection of the pattern shown suggests that memory for exemplar expressions is biased by the ensemble expression (Fig. 2c.). Participants were most likely to respond "same" to probes that either matched the target intensity, or were close to it, illustrated by the peaks occurring close to the centre of each target's response curve. This result suggests that participants had some memory for individual expressions in the group. In addition, memory for the individual expressions appears to be influenced by the group mean expression, because "same" responses to mismatching probes were more frequent when the probe's expression intensity was shifted towards the group mean intensity, compared to when probe intensity was shifted away.

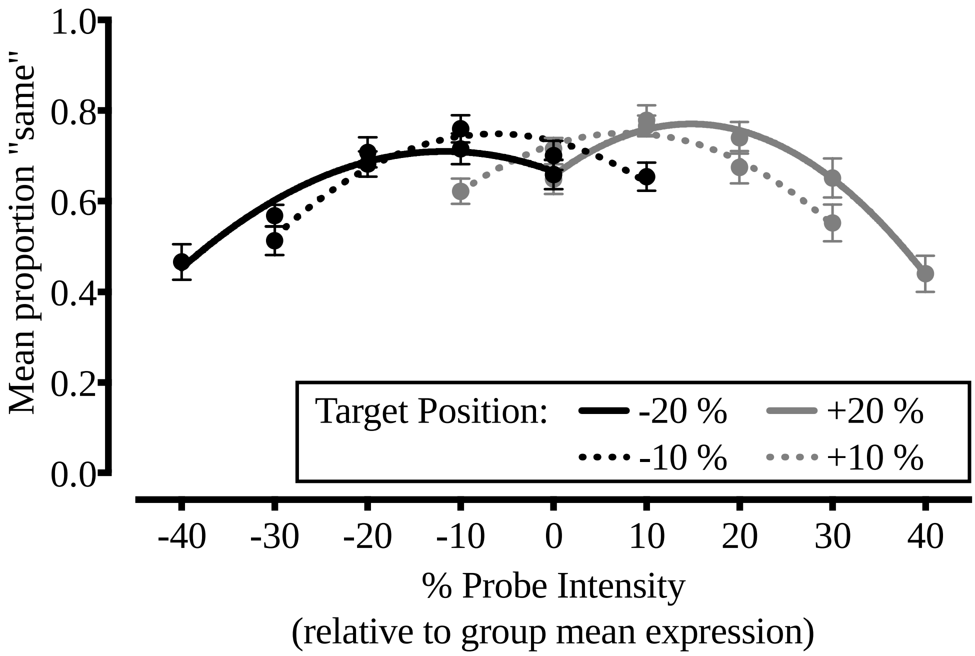

Figure 3. Mean proportion of "same" responses to probes as a function of probe intensity plotted relative to the group mean expression (e.g., for a -20% target (black solid line), the five levels of probe intensity relative to target intensity (-20%, -10%, 0%, +10%. +20%) are

depicted as -40%, -30%, -20%, -10%, and 0%, respectively). Separate curves are plotted for each target position (-20%, -10%, +10%, +20%). Error bars show standard error.

First we tested whether the mean group expression had an effect on responses, or if the data could be explained solely by participants' ability to code exemplars without additionally coding the group mean expression ("exemplar only", Fig. 2a). If only the exemplars are coded then participants' responses should not be differently affected by the target position, that is, the target face's expression intensity relative to the other faces in a group. To test this we entered proportion of "same" responses into a repeated measures ANOVA with Target Position relative to group mean intensity (-20, -10, +10, +20), Probe Intensity (0%, +10%, +20%, -10%, and -20% relative to target intensity), and Emotion (happy versus angry) as factors.

We found no effect of Emotion, and no interaction involving Emotion and Probe Intensity (all *F*s < 2.69, all *p*s > .115) We found a main effect of Probe Intensity, $F(4, 92) = 28.22$ $p < .001$, $\eta^2 = .55$. However, there was also a main effect of Target Position, $F(3, 69) = 3.30$, $p = .025$, $\eta^2 = .126$, qualified by a two-way interaction between Probe Intensity and Target Position, $F(12, 276) = 12.61$, p < .001, $\eta^2 = .354$ (Figure 3). The significant effect of Target Position showed that participants' responses to the different probe faces were affected by how intense a target face's expression was in relation to the group mean expression, thus providing evidence against participants *only* coding the exemplar (see Fig 2b). Separate follow-up ANOVAs confirmed that Probe Intensity effects were significant on each level of Target Position, all $F > 12$, all $p < .001$, all $\eta^2 > .340$, but the interaction between Probe Intensity and Target Position suggests that the effect of probe intensity varied across the different levels of target position.

The interaction of Probe Intensity and Target Position could either indicate that participants code the mean group expression only (‘ensemble only’, Fig. 2b), or that participants code both mean and individual expressions, but are biased towards the group mean expression (‘exemplar biased by ensemble’, Fig. 2c). To distinguish between these two possibilities, we ran two sets of planned contrasts to explore the nature of the interaction, each set addressing one potential outcome specifically.

In the first set of planned contrasts, we determined whether encoding of only an ensemble expression could account for the interaction between Probe Intensity and Target Position. We compared the proportions of “same” responses to matching probes (that corresponded to the target intensity) with performance on mismatching probes that corresponded to the group mean intensity for each target position (see Table 1). If participants remembered only the group mean intensity, but not the individual target intensities (“ensemble only” outcome), then they should have endorsed group mean intensity probes most frequently. Instead, probes that matched the target intensity received at least as many, if not more, “same” responses as mismatching probes that had the group mean intensity, for all target types. Pairwise t-tests were significant for +20% targets, and marginally significant for +10% and -10% targets, (see Table 1). Importantly, the proportions of “same” responses were never higher for probes matching the group mean expression intensity than for probes matching the target expression intensity. Therefore, participants’ responses were not simply reflecting the group mean expression.

Table 1. *Comparison of proportions "same" responses for probes that match the target intensity with probes that match the group mean intensity.*

Target Position	Probe Intensity			t-statistics			
	Target Match	Group Mean Match					
	Mean (SEM)	*Mean (SEM)*	*Difference*	*df*	*t*	*p*	*d*
+ 20%	.740 (.035)	.648 (.033)	.092	23	2.14	.043	0.55
+ 10%	.766 (.023)	.716 (.024)	.050	23	1.86	.076	0.42
- 20%	.708 (.033)	.659 (.032)	.049	23	1.25	.222	0.31
- 10%	.760 (.030)	.701 (.032)	.059	23	1.78	.089	0.39

In a second set of planned contrasts, we tested whether memory for the target intensity was *systematically* shifted in the direction of the group mean intensity, as would be expected if participants' memory for individual expressions was influenced by the ensemble representation of the group expression. We calculated bias scores by subtracting the proportion of "same" responses for probes shifted away from the mean from the proportion of "same" responses for *equidistant* probes shifted towards the mean, for each target position. As can be seen in Table 2, the average bias for every probe pair was positive, indicating a bias in the direction of the mean group expression in all conditions. One sample t-tests (Bonferroni-corrected) carried out for each probe intensity level were significant, except for probes that were ±10% from target when the target was 10% more or 10% less intense than the group mean, and probes which were ±20 from the target when the target was 10% more intense than the group mean (see Table 2). Note that a bias was expected to be smaller for targets that are closer to the group mean, and also for probes that were closer to the target. Thus, the critical comparisons, for which a stronger bias was expected, were the 20% target comparisons, which were all significant. These findings indicate that the ensemble information had a systematic influence on participants' responses, consistent with the "exemplar biased by ensemble" outcome alternative (Fig 2c).

Table 2. *Results of t-tests for level of bias towards the mean in "same" responses for mismatching probes. Bias is computed by taking proportion of "same" responses for probes shifted away from the mean from proportion of "same" responses for equidistant probes shifted towards the mean, separately for the two possible probe distances. Bonferroni correction for multiple comparisons have been applied to p values.*

Target Position	Probe Distance	Bias *Mean (SEM)*	t-statistics			
			df	*t*	*p*	*d*
+20%	±20	.208 (.034)	23	6.05	.000	1.24
	±10	.128 (.035)	23	3.61	.012	0.74
+10%	±20	.070 (.039)	23	1.80	.686	0.37
	±10	.042 (.032)	23	1.31	1.00	0.28
-20%	±20	.193 (.041)	23	4.72	.000	0.96
	±10	.148 (.025)	23	5.83	.000	1.19
-10%	±20	.141 (.033)	23	4.31	.002	0.88
	±10	.018 (.030)	23	0.60	1.00	0.12

Experiment 2: A response bias to intermediate intensities?

A possible alternative explanation for the bias towards the group mean expression in Experiment 1 is that participants could have a generic bias to endorse intermediate expression intensities (e.g., moderately angry faces of 50% absolute expression intensity) more often than expression intensities at the endpoints of the intensity range (e.g., neutral expressions of 0% absolute intensity or very angry/happy expressions of 100% absolute intensity). ‘Central tendency’ biases, in which participants’ estimates are drawn towards the centre of the range of presented stimuli, are well established (Allred, Crawford, Duffy, & Smith, 2016; Crawford, Huttenlocher, & Engebretson, 2000; Duffy, Huttenlocher, Hedges, & Crawford, 2010; Hollingworth, 1910; Olkkonen, McCarthy, & Allred, 2014). Therefore, we must consider whether the bias to endorse intermediate expression intensities found in Experiment 1 could have been induced by a higher probability of probe and target faces with intermediate than extreme (low or high) expression intensities.

Table 3. *Probability of probes and targets for each level of absolute expression intensity.*

	Absolute expression intensity										
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Target	0.02	0.06	0.08	0.11	0.15	0.15	0.15	0.11	0.08	0.06	0.02
Probe	0.04	0.06	0.09	0.11	0.13	0.13	0.13	0.11	0.09	0.06	0.04

Note. The sum of the shown probabilities may not equal 1 due to rounding errors.

In Experiment 1, participants endorsed probe faces that deviate towards the mean group intensity more frequently than probe faces that deviate away from the mean group intensity. A potential central tendency bias might also lead to this response pattern, because probes that deviated away from the group mean were more likely to be of extreme intensities than probes that deviated towards the group mean (Table 3). For instance, a 0% (neutral) face could only ever occur as a probe that deviated away from the mean of the group, because for it to deviate towards the mean, the group would had to have a mean group intensity *below* 0%, which was outside the range of expressions used in the present study design.

To rule out a central-tendence bias account of our results in Experiment 1, in Experiment 2 we removed non-target faces from the sets (leaving just the target face presented alone) while keeping the procedure otherwise identical to Experiment 1. If the bias observed in Experiment 1 reflects a genuine influence of ensemble coding, then it should no longer be seen, because there is no group to be ensemble coded.

Methods

Participants

Twenty-four students and volunteers from the University of Western Australia participated (mean age = 20.3, SD = 3.4 years; 8 male). Student participants received course credits.

Stimuli

Face stimuli and the experimental protocol from Experiment 1 were used. However, the critical change to Experiment 1 was that only the target face of each study group was presented.

Procedure

The trial procedure was identical to Experiment 1, except that the target face was presented alone. Participants were instructed to remember the expression intensity of the target face as accurately as possible. A single probe face of the same identity as the target was then presented in the centre of the screen, and participants indicated whether they believed the probe face had the same or a different intensity expression as the target face.

Results

Figure 4 shows participants' "same" responses as a function of the probe intensity relative to the group mean intensity. Inspection of the pattern shown suggests that memory for the exemplar expressions was accurate (cf. Fig. 2b expected "exemplar only" pattern), with no evidence of any systematic bias as found in Experiment 1. To formally test accuracy we conducted a repeated measures ANOVA with Emotion (happy, angry), Target Position relative to group mean intensity (-20, -10, +10, +20) and Probe Intensity (0%, +10%, +20%, -

10%, -20% relative to target intensity) as factors. Importantly, there was no interaction of Target Position and Probe Intensity, and no triple interaction of Target Position, Probe Intensity, and Emotion, both $F < 1.1$, both $p > .40$, indicating the absence of a systematic bias towards the mean expression intensity. These results confirm that the bias towards the mean in Experiment 1 reflects the influence of the surrounding set rather than any generic bias to endorse probes with intermediate intensity expressions.

We found several other effects that were not of any theoretical significance. We report them for completeness. There were significant main effects of emotion (angry > happy, $F(1,23) = 35.87$, $p < .001$, $\eta^2 = .609$) and Probe Intensity, $F(1,23) = 155.27$, $p < .001$, $\eta^2 = .871$. Emotion interacted both with Target Position, $F(3,69) = 13.29$, $p < .001$, $\eta^2 = .366$ and with Probe Intensity, $F(4,92) = 15.85$, $p < .001$, $\eta^2 = .408$.[1]

[1] The interaction of Emotion and Probe Intensity is driven by participants' higher rate of "same" responses to mismatching probes (that is, lower accuracy) for angry compared to happy faces, and this effect was particularly pronounced for the ±20 % probes. Post-hoc t-tests confirmed that there were fewer endorsement in happy compared to angry trials when probes were -20 %, $t(23) = 11.09$, $p < .001$, $d = 1.68$; +20 %, $t(23) = 4.84$, $p < .001$, $d = 1.43$; -10 %, $t(23) = 2.30$, $p = .031$, $d = 0.55$ and +10 %, $t(23) = 3.42$, $p = .002$, $d = 0.98$ from the target. In contrast, there was no difference between emotions when the probe matched the target (0 %), $t(23) = 0.32$, $p = .75$, $d = .06$. The Emotion by Target Position interaction was driven by more "same" responses to happy faces of higher intensity than lower intensity (greater "same" responses for probes testing -20 % targets compared to +20 % targets, $t(23) = 5.05$, p < .001, $d = 1.00$). In contrast, responses on trials with angry faces did not appear to depend on target intensity ("same" responses to -20 % targets and to +20 % targets were not significantly different, $t(23) = 1.42$, $p = .169$, $d = 0.31$).

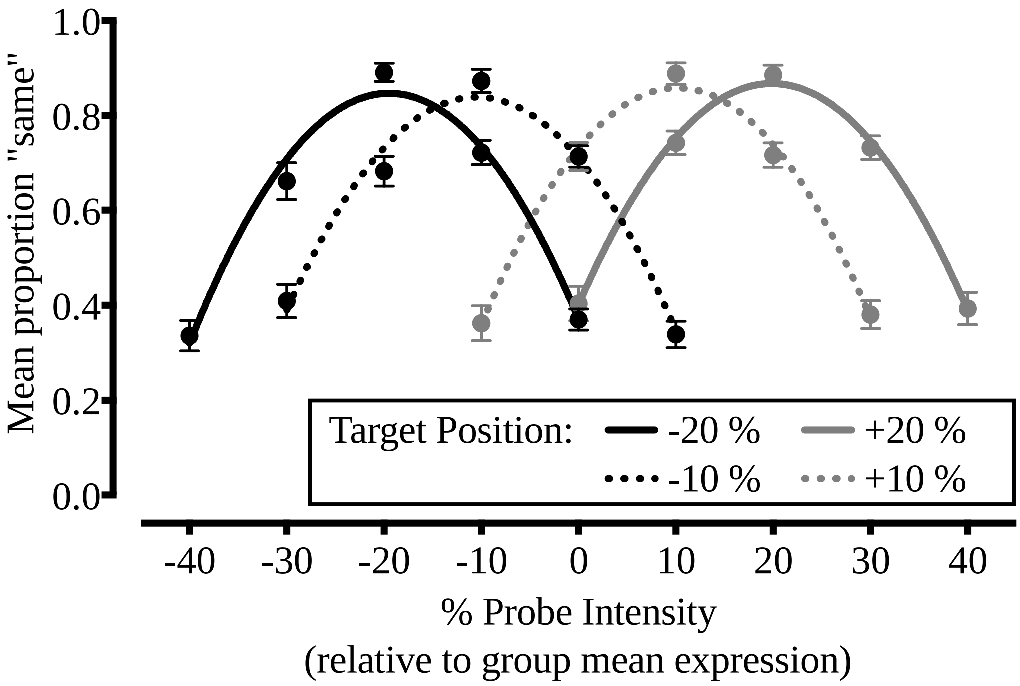

Figure 4. Mean proportion of “same” responses to probes as a function of probe intensity plotted relative to the group mean expression (e.g., for a -20% target (black solid line), the five levels of probe intensity relative to target intensity (-20%, -10%, 0%, +10%. +20%) are depicted as -40%, -30%, -20%, -10%, and 0%, respectively). Separate curves are plotted for each target position (-20%, -10%, +10%, +20%). Note that the target and probe intensities are relative to the mean intensity of a “group” that was not seen. Thus, any potential bias in the data is unrelated to the context of a group. Error bars show standard error.

Discussion

We found that memory for the intensity of a facial expression is biased towards the average intensity of the expressions of a surrounding crowd. Participants overestimated the intensities of individual expressions that were less intense, and underestimated the intensities that were more intense, than the average expression of the group. A control experiment confirmed that this bias could not be explained by a general bias to endorse intermediate intensity expressions. Our results provide an important and novel extension to the studies indicating ensemble coding of expression for groups containing a single face identity (Haberman & Whitney, 2007, 2009) by showing that representations of individual expressions are biased by the ensemble expression. This extends our own previous work that showed that extraction of exemplar and ensemble identity information can co-occur (Neumann et al., 2013), by suggesting an information transfer between these representations (for facial expressions). Finally, by showing that ensemble coding of facial expression occurs for more naturalistic, heterogeneous groups containing distinct identities, our data and those of Yang et al. (2013), provide convincing evidence for the idea that such coding could play an important role in determining the mood of a crowd.

Participants in the current study not only coded ensemble information, but also retained memory for the individual expressions of the group (shown by the fact that "same" responses were given at least as often to probes that matched the target as to probes that deviated towards the mean expression intensity). In previous studies using only one identity face little information about the individual exemplar expressions appeared to be retained (Haberman & Whitney, 2007, 2009). The memory for individual expressions seen here may reflect the increased discriminability of facial expressions displayed on different identities (see Avons, 1999 for evidence that similarity of visual stimuli reduces accuracy of short term memory). Our finding that ensemble expression representations influence memory for the individual expressions in a group parallels findings for ensemble coding of properties of simple objects (Brady & Alvarez, 2011; Brady, Konkle, Oliva, & Alvarez, 2009; Brady & Tenenbaum, 2013). Thus, although there is evidence that ensemble coding for higher level properties such as facial expression is supported by a system that is separate to a system supporting ensemble coding for low-level properties such as size (Haberman, Brady, & Alvarez, 2015), our findings suggest that both types of ensemble representation can have similar effects on memory for the properties of individual group members.

Another line of evidence has suggested that ensemble representations of face properties systematically influence representations' of individual faces in groups. Walker and Vul (2013) showed that individual faces are perceived as more attractive when seen in a group than when seen alone. This finding, known as the "cheerleader effect", has been attributed to the influence of the ensemble identity representation on perception of attractiveness of the individual faces. Average faces (such as an ensemble identity) are generally perceived as attractive (Rhodes, 2006), so it is argued that attractiveness ratings for the individual faces are pulled up by the group ensemble identity. However, others have argued that the "cheerleader effect" could be explained by selective attention to the most attractive individual in a group (van Osch, Blanken, Meijs, & van Wolferen, 2015). Crucially, here we found that memory for individual expressions was biased toward the average expression intensity regardless of whether the individual expression was more or less intense than the average, which rules out selective attention to the most emotionally intense expression in a group as the source of the effect. The demonstration of the cheerleader effect (Walker and Vul, 2013), and the present data provide converging evidence that ensemble representations can influence the coding of information about the individual faces in a group (here, different expression intensities), and that this effect cannot be explained by higher selective attention to the most intense face in a group.

A bias towards the mean property of a group, as seen here, reflects the integration of ensemble information with information about individual members of a group. It has been suggested that the optimal combination of ensemble and individual information could serve to minimize the effects of perceptual errors that result from capacity limitation during encoding or retrieval of individual group members (Brady & Alvarez, 2011). When viewing a crowd, capacity limitations (e.g., too little time to attend to all individual faces) may lead to either inaccurate representations of each individual's expression, or coding only a subset of the group's faces. In contrast, ensemble representations have been shown to accurately represent the mean facial expression of all faces of the group, even when encoding time is very limited (Haberman & Whitney, 2009). Therefore, it is possible that a bias to see individual faces to be slightly more like the group mean reflects an adaptive "optimal integration" process (Brady & Alvarez, 2011) by which the visual system integrates the group information (which contains some information about the individuals) into the representation of the individual items, to increase accuracy of the individual representations, on average. Knowing something about the group, e.g., that the faces were overall very angry, could help us decide that one of the faces that we have perceived as relatively neutral could in fact be more angry than we thought, particularly if our memory for this face is inaccurate.

A strategy of "optimal integration" (Brady & Alvarez, 2011) might predict that the knowledge about the group mean becomes increasingly useful (and could thus cause stronger bias), as our knowledge of an individual becomes less accurate. If we know very little about an individual face of a group, but have encoded the group as overall very angry, it is reasonable to assume that the individual face was about as angry as the group on average. If however we have a very accurate representation of an individual face's expression, we might not need the ensemble group information at all. Future research could establish the precise relationship between representation strength of individual expressions and the strength of the bias towards the mean expression of the group, for instance, by manipulating the presentation duration of the study group. Longer presentation times would be expected to facilitate the coding of individual exemplars, thus potentially reducing the influence of the ensemble, which would be indicated by a weaker bias towards the mean expression intensity of a group.

Finally, in our study we tested participants' visual short-term memory for a previously presented expression. However, it is possible that the bias observed in Experiment 1 is the result of a bias that occurs during perception, rather than a bias that occurs when expressions are stored in short term memory. A perceptual source of the bias would be consistent with evidence from ensemble perception of simpler features, for instance the computations of orientation statistics from Gabor patches (Ross & Burr, 2008). From the present data we are unable to determine whether the observed bias occurs in perception or memory (or both). Whether the there is a bias in perception, which alters how the expressions individuals in crowds are perceived, will be an interesting question for future studies.

Conclusion

This study demonstrates that judgements of the intensity of individual facial expressions in a group are biased towards the group "ensemble" expression intensity. This bias suggests that ensemble coding of expression occurs for groups that consist of different identity faces. It also shows, for the first time, that information about the ensemble expression of a group is integrated with information about the individual group expressions, consistent with an "optimal integration model" (Brady & Alvarez, 2011). More generally, our results add to increasing evidence that emotional expression processing is malleable and generally affected by context (Barrett et al., 2011).

References

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

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in cognitive sciences, 15*(3), 122-131. doi:10.1016/j.tics.2011.01.003

Anobile, G., Cicchini, G. M., & Burr, D. C. (2016). Number As a Primary Perceptual Attribute: A Review. *Perception, 45*(1-2), 5-31. doi:10.1177/0301006615602599

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157-162. doi:10.1111/1467-9280.00327

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., . . . Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science, 19*(7), 724-732. doi:10.1111/j.1467-9280.2008.02148.x

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions. *Science, 338*(6111), 1225-1229.

Avons, S. E. (1999). Effects of Visual Similarity on Serial Report and Item Recognition. *The Quarterly Journal of Experimental Psychology Section A, 52*(1), 217-240. doi:10.1080/713755809

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in Emotion Perception. *Current Directions in Psychological Science, 20*(5), 286-290. doi:10.1177/0963721411422522

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychol Sci, 22*(3), 384-392. doi:10.1177/0956797610397956

Brady, T. F., Konkle, T., Oliva, A., & Alvarez, G. A. (2009). Detecting changes in real-world objects: The relationship between visual long-term memory and change blindness. *Communicative & integrative biology, 2*(1), 1-3.

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological review, 120*(1), 85.

Crawford, L. E., Huttenlocher, J., & Engebretson, P. H. (2000). Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science, 11*(4), 280-284.

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *The Quarterly Journal of Experimental Psychology, 62*(9), 1716-1722. doi:10.1080/17470210902811249

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*(2), 224-230. doi:10.3758/pbr.17.2.224

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*(2), 432-446. doi:10.1037/xge0000053

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*(17), R751-R753.

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance, 35*(3), 718.

Haberman, J., & Whitney, D. (2012). *Ensemble perception: Summarizing the scene and broadening the limits of visual processing*: Oxford University Press.

Halberstadt, J. B., & Niedenthal, P. M. (2001). Effects of emotion concepts on perceptual memory for emotional expressions. *J Pers Soc Psychol, 81*(4), 587-598. doi:10.1037//0022-3514.81.4.587

Hollingworth, H. L. (1910). The central tendency of judgment. *The Journal of Philosophy, Psychology and Scientific Methods, 7*(17), 461-469.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24*(8), 1377-1388. doi:10.1080/02699930903485076

Leib, A. Y., Fischer, J., Liu, Y., Qiu, S., Robertson, L., & Whitney, D. (2014). Ensemble crowd perception: A viewpoint-invariant mechanism to represent average crowd identity. *J Vis, 14*(8). doi:10.1167/14.8.26

Leib, A. Y., Puri, A. M., Fischer, J., Bentin, S., Whitney, D., & Robertson, L. (2012). Crowd perception in prosopagnosia. *Neuropsychologia, 50*(7), 1698-1707. doi:10.1016/j.neuropsychologia.2012.03.026

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition, 128*(1), 56-63. doi:http://dx.doi.org/10.1016/j.cognition.2013.03.006

Ngo, N., & Isaacowitz, D. M. (2015). Use of Context in Emotion Perception: The Role of Top-down Control, Cue Type, and Perceiver's Age. *Emotion, 15*(3), 292-302. doi:10.1037/emo0000062

Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *J Vis, 14*(11), 15. doi:10.1167/14.11.5

Rhodes, G. (2006). The evolutionary psychology of facial beauty *Annual Review of Psychology* (Vol. 57, pp. 199-226). Palo Alto: Annual Reviews.

Righart, R., & de Gelder, B. (2008). Rapid influence of emotional scenes on encoding of facial expressions: an ERP study. *Social cognitive and affective neuroscience, 3*(3), 270-278. doi:10.1093/scan/nsn021

Ross, J., & Burr, D. (2008). The knowing visual self. *Trends in cognitive sciences, 12*(10), 363-364. doi:10.1016/j.tics.2008.06.007

Sweeny, T. D., Grabowecky, M., Paller, K. A., & Suzuki, S. (2009). Within-hemifield perceptual averaging of facial expressions predicted by neural averaging. *J Vis, 9*(3), 2.1-11. doi:10.1167/9.3.2

van Osch, Y., Blanken, I., Meijs, M. H. J., & van Wolferen, J. (2015). A Group's Physical Attractiveness Is Greater Than the Average Attractiveness of Its Members: The Group Attractiveness Effect. *Personality and Social Psychology Bulletin, 41*(4), 559-574. doi:10.1177/0146167215572799

Walker, D., & Vul, E. (2013). Hierarchical encoding makes individuals in a group seem more attractive. *Psychological Science*, 559-574. doi:10.1177/0956797613497969

Yang, J. W., Yoon, K. L., Chong, S. C., & Oh, K. J. (2013). Accurate but Pathological: Social Anxiety and Ensemble Coding of Emotion. *Cognitive Therapy and Research, 37*(3), 572-578. doi:10.1007/s10608-012-9500-5