scientific reports

Check for updates

OPEN

Attention and stimulus structure interact during ensemble encoding of facial expression

Marco A. Sama[1]✉, Moaz Shoura[1], Adrian Nestor[1,2] & Jonathan S. Cant[1,2]

Face ensemble encoding involves synthesizing summary information from groups of faces, providing a mechanism to overcome limitations in visual working memory. Yet, research on the role of attention has revealed mixed findings. Also, the simultaneous processing of summary representations and individual faces within an ensemble remains largely unexplored. Here, across three experiments, participants viewed ensembles without a central face ($n = 32$), or with a central face, while attention was distributed across the entire ensemble ($n = 38$) or focused centrally ($n = 38$). Critically, the consistency of center and surround faces varied as a function of emotional valence (i.e., same versus opposite). Participants completed an expression similarity-rating task between an ensemble and a single face, which was used to recover, via image reconstruction, visual estimates of summary representations. Reconstructions were then assessed against central faces, surrounding faces, and their averages. We show that focused attention enhances central face representation and that consistency benefits the representation of both the center and of the surround. However, central faces outweigh the overall surround representation only when attention is focused on a center face inconsistent with its surround. These findings reveal a flexible relationship between attention and stimulus structure in ensemble perception.

Keywords Ensemble encoding, Image reconstruction, Face perception, Attention, Facial expression

Ensemble encoding involves compressing visual variability across multiple objects into a single, summary metric (i.e., an average[1,2]. Face ensemble encoding—or crowd perception—involves the synthesis of summary information from faces, such as identity[3,4], viewpoint[5,6], or expression[7–9]. This encoding mechanism is robust, as detailed summaries can be derived from partially-occluded faces[10] and outliers are typically discarded from the summary[11,12].

Interestingly, despite the complexity of an ensemble stimulus, integration can take place in the absence of foveal input[8]. However, while foveation is not required, the role of attention in ensemble encoding remains subject to debate[9,13]. On the one hand, some have argued for an attentional requirement, citing inattentional blindness in ensemble processing[14], and summary derivation through subsampling of ensemble items[15]. On the other hand, ensemble encoding has been observed with attention seemingly absent or directed elsewhere. For example, one can track the centroid location of both attended moving targets and unattended moving distractors[16], or accurately report the diversity of stimuli from both cued and un-cued rows in an array of colored letters[17].

These latter findings do not discount the possibility that, upon realization of task demands, participants may still attend to un-cued elements of the ensembles. Recent work addressed this by examining the implicit influence of unattended ensembles. Specifically, one study[18] cued participants to rate whether one of two ensemble stimuli belonged to a target category. Participants could safely ignore the un-cued ensemble as they were never prompted to respond to it. Interestingly, performance on the rating task improved when both attended and unattended ensembles were of the same category, suggesting that summary statistics were derived even from unattended stimuli.

An alternative, unexplored account for the mixed findings on the role of attention may involve the modulation introduced by ensemble structure and experimental task. In particular, the relationship between a target element and the overall ensemble may interact with task-specific attentional demands. Common ensemble paradigms, such as the set-membership identification task, have revealed that participants are more likely to recognize the mean over any individual exemplar from the set, even when the mean value was never explicitly presented within the stimulus display[2]. However, these methods cannot speak to the representation of exemplars within

[1]Department of Psychology, University of Toronto Scarborough, Toronto, ON, Canada. [2]Adrian Nestor and Jonathan S. Cant contributed equally to this work. ✉email: marco.sama@mail.utoronto.ca

the context of a broader ensemble (e.g., speaking to a single student within a crowded lecture theatre). Yet, arguably, examining the representation of a single central face amidst a surrounding group of faces is a more ecologically valid approach, it can assist in understanding the relationship between single and face ensemble domains, and, crucially, it can provide insight into the role of attention in ensemble processing. In this respect, several points must be considered. First, the perception of a single face—when in isolation—is less attentionally demanding than in the context of an ensemble surround. Second, center and surround face information may be competing for visual working memory resources. Third, many ensemble paradigms treat all exemplars in the display equally as, presumably, attention is evenly distributed across items. Fourth, filtering outliers from the ensemble summary[11] may not apply if the outlier is the attended item.

With respect to faces, the representation of single-face expressions can be impacted by flanking emotional stimuli[19], and, conversely, foveating a central face can impact the perception of the surround[20]. Further, distinct neural profiles were observed for central versus surround faces from ensemble stimuli, as revealed by pattern analysis applied to electroencephalography data[21]. However, these studies did not control for attentional scope, thus limiting their conclusions. These studies also did not evaluate the dual representation of exemplars and the ensemble summary derived from the same visual stimulus.

The mixed findings on the role of attention, as well as the common practice of studying separately single faces versus face ensembles, motivate our present study. Here, we explore the relationship between attentional scope and the structure of the ensemble stimulus and we also explore how attention influences summary versus exemplar representations—effectively, we examine single face processing within the context of a broader ensemble. Thus, our study speaks to the joint mechanisms of single and face ensemble encoding.

To this end, we consider ensembles containing a central face, either consistent or inconsistent with its surround (i.e., by having the same or opposite emotional valence), while attention is either distributed across the entire stimulus or focused on the central face. Then, we rely on the perceptual similarity between ensemble stimuli and single face probes to derive summary representations via image reconstruction[22] and we assess their visual content. Specifically, we evaluate the relative extent to which summary representations reflect the visual characteristics of the center and the surrounding faces as a function of attentional deployment.

The image reconstruction procedure is capable of recovering face representations from behavioral[23] and neural data[24], including the summary percept from face ensembles[25]. This makes image reconstruction a particularly powerful tool when applied to face ensembles, as it can be used to visualize properties that are not explicitly part of the original stimulus, such as its mean. Here, we rely on this method to recover summary expression representations of face ensembles. Then, we assess their sensitivity to stimulus structure and attentional scope.

First, in Experiment 1, we apply behavior-based image reconstruction to a typical ensemble stimulus consisting only of surround faces, without a central target, to provide a general validation of our theoretical and methodological approach. Next, across two complementary experiments, we employ ensembles consisting of a central face and six surrounding faces to examine the impact of distributed (attending to an entire ensemble stimulus; Experiment 2) versus focal attention (attending only to the central face; Experiment 3).

Experiment 1

The present experiment aims to validate behavior-based image reconstruction[22] as a viable approach to extracting summary expression representations from face ensembles. To this end, theoretically, we rely on a face space framework, that is, a multidimensional space in which faces are represented as points and their perceptual similarity as pairwise distances[26]. Here, we explore the benefit of projecting ensemble representations in this space based on their similarity with single-face expressions. Given that our face stimuli vary in expression rather than identity, we anticipate that the representational space will capture dimensions relevant for emotion perception, such as valence[27–29], and, more importantly, that ensemble location in this space will reflect the visual characteristics of their perceptual summaries. Then, we capitalize on the structure of this space to recover the visual appearance of both single expressions and ensemble summaries.

Ethics statement

This study protocol was approved by the University of Toronto Research Ethics Board, and was conducted in accordance with the Declaration of Helsinki. All participants in this and subsequent experiments provided electronically-submitted informed consent prior to taking part in the experiment.

Participants

A total of 36 White participants were recruited online via Prolific (prolific.co/) from the USA, the UK, Australia, Canada, or New Zealand. Participants were right-handed, had no history of head injury or mental illness, had normal or corrected-to-normal vision, no color blindness, and maintained a minimum experiment approval rate of 95% (indicative of adequate conduct when completing studies online). Participants provided informed consent via Qualtrics (qualtrics.com/) and were compensated £10.00 or equivalent exchange for those outside the UK. The experiment relied on Psychopy v2021.2.3 (psychopy.org/[30]) and Pavlovia (pavlovia.org/). The data from three participants were lost due to technical difficulties, and data from a fourth participant were discarded due to poor performance (see Results). This resulted in a final sample size of 32 participants (15 females, 17 males, age: 20–35 years).

Stimuli

Facial expression stimuli were selected from the large MPI Facial Expression Database[31]. Stimuli were derived from video recordings of White male actors evincing a variety of facial expressions. Overall, we selected 20 representative expressions (see Fig. 1) from two facial identities (i.e., actors), referred to as ID1 and ID2. For each participant, stimuli were selected from only one of these two identities (ID1 or ID2; counterbalanced across

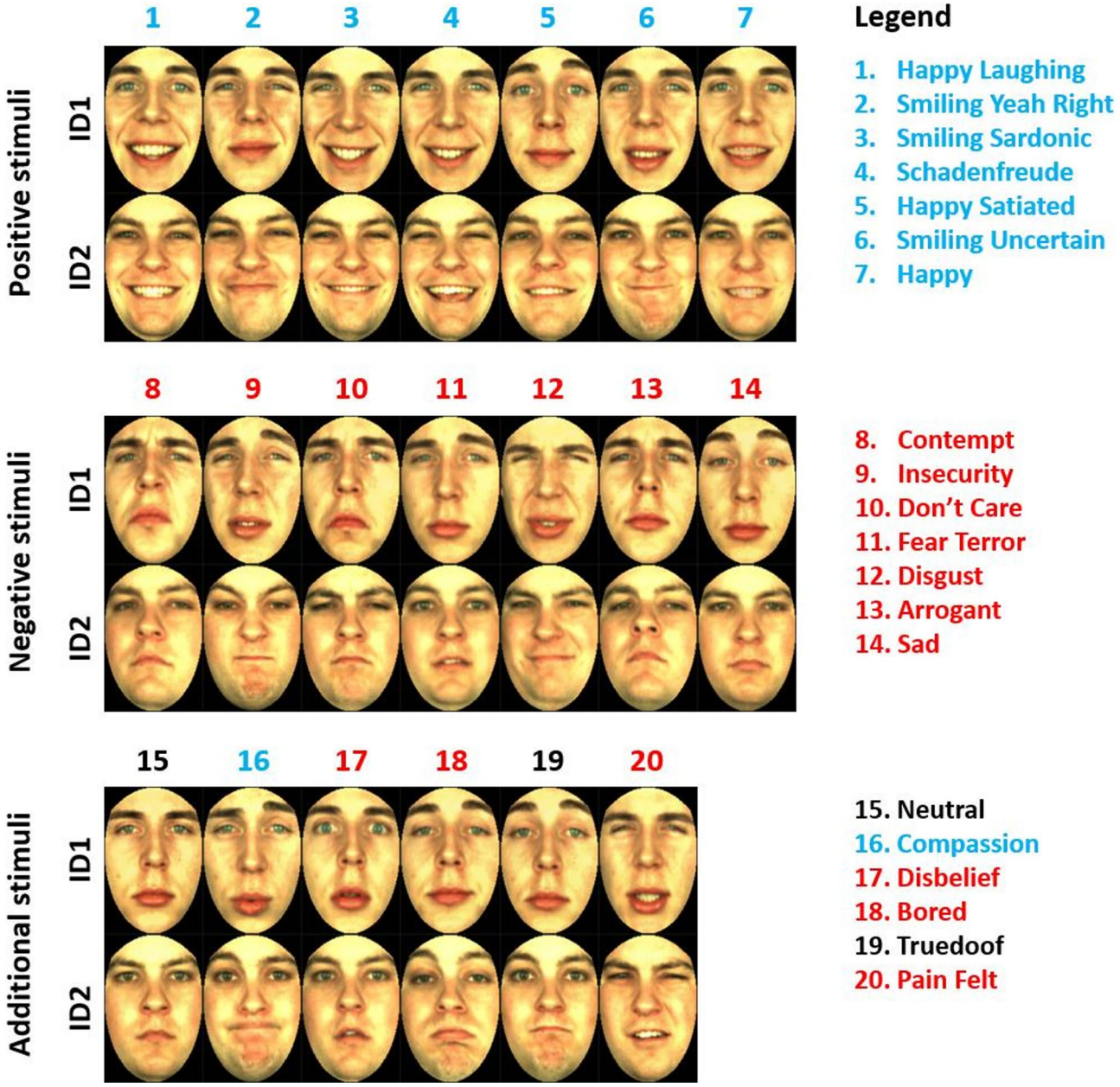

Fig. 1. Twenty facial expression stimuli from two identities used in all experiments. Single-face stimuli relied on 20 different expressions from two facial identities (i.e., actors). Seven expressions with positive valence and seven with negative valence were selected to construct positive ensembles and negative ensembles, respectively (blue = positive; red = negative, black = neutral/ambiguous).

participants). Still images were obtained by selecting the frame displaying peak intensity for a given expression with a frontal viewpoint. This approach aimed to yield prototypical displays of emotion, which were easy to interpret, though differences in the level of arousal are likely still present across our stimuli. Next, faces were cropped, masked to preserve only internal features and geometrically aligned with the position of the main features (e.g., eyes, nose) to control for low-level visual differences.

Single-face stimuli consisted of 20 expressions for each facial identity. Of these, seven, with positive valence, were used to design positive ensemble stimuli, and another seven, with negative valence, were used to design negative ensembles. The remaining six faces included one neutral expression along with a mix of additional positive and negative expressions, which were used only for single-face comparisons (Fig. 1). We note that the position of the faces in an ensemble were randomized from trial to trial to eliminate the potential use of strategies (e.g., anticipating specific images at a given spatial position).

A single-face stimulus subtended 90 × 135 pixels (a 2.1° × 3.3° visual angle from a distance of ~60 cm from the screen). An ensemble stimulus consisted of six faces whose centers were angled 60° from one another relative to a fixation cross at the center of the screen, for a total visual angle of 9.4° × 10.7°.

Procedure

Participants were instructed to complete the experiment in a distraction-free environment and maintain ~ 60 cm from the center of their screen (either a desktop or laptop), which displayed a black background and a center fixation cross. Experimental blocks contained either single-to-single (S-S) or ensemble-to-single (E-S) face pairs. Block order was randomized.

Trials in each block began with a 500 ms intertrial interval (ITI) displaying a fixation cross, followed by stimulus presentation for 300 ms (either a single face or ensemble), a 200 ms interstimulus interval (ISI), a single face probe for 300 ms, followed by another 200 ms ISI, and, finally, the response screen. Participants were prompted to rate the emotional similarity between the stimulus and probe on a scale of 1–7 (very dissimilar–very similar expression) using the keyboard number bar (Fig. 2). For S-S trials, this was the similarity between the two single face probes. For E-S trials in Experiments 1 and 2, this was the similarity between the overall ensemble expression and the single face probe. For E-S trials in Experiment 3, this was the similarity between the centrally-located face of the ensemble and the single face probe—see Fig. 2.

For the S-S block, a single-face stimulus appeared at the center of the screen followed by another single-face probe at the same location. Participants were required to fixate the center cross location at all times, even when absent during the stimulus and probe face displays. For the E-S block, ensemble stimuli consisted of six faces sharing the same valence (i.e., all positive or all negative), arranged in a circle around the fixation cross, followed by a single face probe at the center of the screen.

The S-S block cycled through all single-face pairs, resulting in 190 stimulus-probe combinations (i.e., all possible pairs of the 20 single face-stimuli), repeated twice for a total of 380 S-S trials. During the E-S block, an ensemble of six faces was constructed by leaving out one face at a time, yielding seven positive and seven negative ensembles, for a total of 14 stimuli. These were paired with all 20 single-face probes, resulting in 280 stimulus-probe combinations, out of which 50% were repeated for a total of 420 E-S trials. Trials were randomized within each block. One-minute breaks occurred after every 95 S-S trials and 84 E-S trials. Ten practice trials were included prior to each block to familiarize participants with the task. The average completion time for Experiment 1 was 46 min.

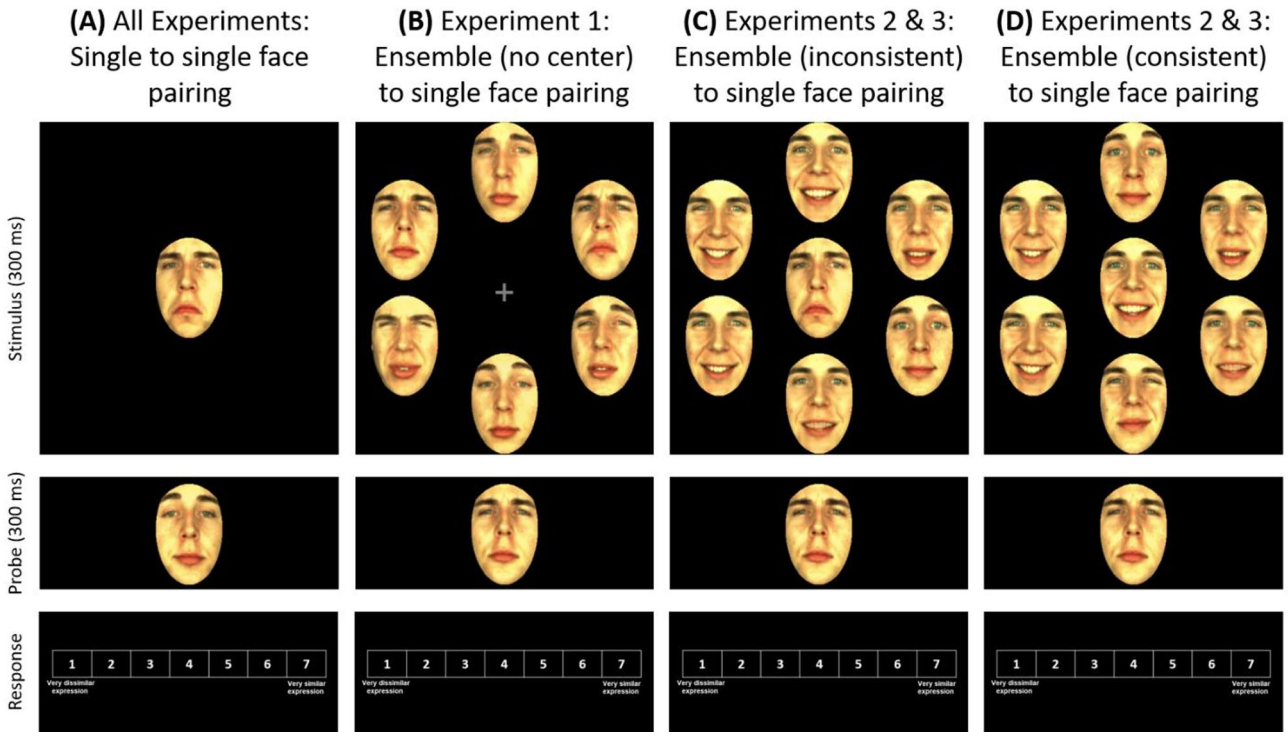

Fig. 2. General trial schematic and examples of stimulus-probe pairs. Participants rated the expression similarity between a stimulus and a subsequent single-face probe: (**A**) single-single (S-S) pairs used in all experiments; (**B**) ensemble-single (E-S) pairs in Experiment 1, with ensembles consisting of same-valence surround faces; (**C**) E-S pairs for Experiments 2 (distributed attention task) and 3 (focused attention task), in which the central face has opposite valence relative to the ensemble surround (i.e., inconsistent center-surround ensemble); and (**D**) E-S pairs for Experiments 2 and 3, in which the central face and surround exhibit the same valence (i.e., consistent center-surround ensemble). For ensembles, participants were instructed to provide ratings based on either the "overall emotional expression" (Experiments 1 and 2), or the emotional expression of only the center face (Experiment 3).

Data analysis

Since participants completed two trials for each S-S pair, corresponding ratings were averaged to provide more robust estimates of single-face similarity, which were used, in turn, to generate a face space (see below). For E-S ratings, though, repeated trials were discarded since only half of the stimulus pairs were repeated. Next, a matrix capturing pairwise similarity ratings was computed for each participant. These matrices were further averaged, separately for the two stimulus identities, and used for visualization purposes—reconstruction was conducted across individual participant data.

Analyses relied on MATLAB 2016b, and JASP 0.17.1 (jasp-stats.org/). The Holm-Bonferroni method was applied for multiple comparison corrections. All tests were supplemented with Bayesian hypothesis testing, adding confidence to both significant and nonsignificant findings. Resulting BF_{10} values are reported, providing weight in favor of either the alternative or null hypothesis using conventional intervals for interpretations[32] (evidence towards the alternative: 1 = no evidence; 1–3 = anecdotal evidence; 3–10 = substantial evidence; 10–30 = strong evidence; 30–100 = very strong evidence; >100 = extreme evidence; reciprocal values are used to interpret confidence towards the null: 1 to 1/2 = anecdotal evidence towards the null, etc.). Modelling was based on the recommended default distributions for unspecified priors: the uniform distribution for ANOVA equivalents[33], and the Cauchy distribution for *t* test equivalents[34]. Note that BF_{10} values are more insightful when evaluating nonsignificant NHST results by providing evidence towards the null rather than reporting a lack of a significant finding. Importantly, we do find general alignment between our NHST and Bayesian results—we do not observe cases where they are in conflict with one another (e.g., a highly significant p value with a BF trending strongly towards the null). On the occasion when these results do not align very well, we take a holistic approach by considering the combined strength of both outcomes (e.g., we would be inclined to interpret a result reporting a $p < .001$ but a BF_{10} between 0.33 and 1.00 as significant).

Image reconstruction procedure

Our reconstruction procedure follows prior work on behavior-based image reconstruction of facial identity[22]. Here, we adapt this method to the reconstruction of summary ensemble representations of facial expressions (see Supplemental Methods and Supplemental Fig. 1 for a detailed description). We used this procedure given its reliance on face space[26], an influential theoretical framework in the study of face processing. Further, our current approach can easily be extended to neuroimaging in the future (e.g., by reconstructing a summary percept from the neural data even when in the absence of an ensemble-specific task)[25].

Briefly, the procedure leverages similarity ratings of single-face stimuli to derive a multidimensional face space via metric multidimensional scaling (MDS), with each dimension capturing a distinct property of facial expressions (e.g., valence). Then, visual features are synthesized via reverse correlation for each dimension from face images populating this space. Next, other single faces or face ensemble representations are projected into this space based on their similarity with existing faces in the same space. Last, the appearance of projected representations is derived through a linear combination of visual features proportional to the coordinates of the projections on corresponding dimensions.

Of note, the procedure above relies on a joint representational space for single faces and summary ensemble representations. However, for simplicity, the structure of the space, which underlies feature derivation prior to projection, is solely based on single-face distances.

The reconstruction procedure was applied to data from each participant separately, for analysis purposes, as well as to data averaged across participants for visualization purposes. A theoretical observer, which assumes maximal access to low-level image-based similarity, was also computed in advance to confirm and assess the ability of our stimuli and of the method to support successful levels of reconstruction (see Supplemental Methods).

Reconstruction accuracy was assessed separately for single faces and face ensembles. Specifically, for single faces, each reconstructed image was compared to its corresponding stimulus, via an L2 pixelwise metric, as well as to all other faces as foils. Accuracy was estimated as the proportion of instances on which the distance to the corresponding stimulus was smaller than to the foils. For clarity, an accuracy of 100% does not indicate a perfect recovery of visual information in a stimulus, but rather a reconstruction closer to its corresponding stimulus than to other images. Of note, this provides a conservative estimate of reconstruction success, since a reconstruction is recovered from visual features derived from the same images which serve as foils.

For ensembles, a similar approach was followed, except that reconstructions were compared to pixelwise averages of all faces in an ensemble, aiming to capture the appearance of the ensemble summary. Then, each reconstruction was compared to its corresponding stimulus average versus all other ensemble averages. While this provides a rough estimate of reconstruction success for ensembles, we note that the thrust of our investigation is to characterize the content of the reconstructions and their sensitivity to different factors, as detailed below. In this sense, significant levels of accuracy mainly serve to ground this investigation for subsequent analyses (i.e., Experiments 2 and 3 in which reconstructions are compared with different stimuli of interest).

Last, the significance of reconstruction accuracy for each type of stimulus (single faces and ensembles) was estimated across participants via a one-sample Wilcoxon Signed Rank (WSR) test against 50% chance. Effect sizes are reported using a rank biserial correlation (RBC).

Results and discussion

To evaluate the consistency of similarity ratings, we correlated repeated trials for each participant, separately for S-S and E-S blocks. Data from one participant were removed due to a nonsignificant correlation across repeated S-S pairings ($r = -.073$, $p = .147$, 95% CI = [−0.170, 0.026], $BF_{10} = 0.18$), demonstrating an inconsistent pattern of ratings. Across remaining participants and facial identities, we found good levels of consistency for both S-S

(average correlation: $r_{188} = 0.647$, $p < .001$, 95% CI = [0.555, 0.722], $BF_{10} = 6.81 \times 10^{20}$) and E-S pairs ($r_{278} = 0.783$, $p < .001$, 95% CI = [0.733, 0.824], $BF_{10} = 5.84 \times 10^{55}$).

Reconstruction accuracy was significantly above chance for single-face expressions from both facial identities (ID1 mean accuracy = 69.4 ± 1.3%, WSR = 105.0, $p < .001$, RBC = 0.294, $BF_{10} = 810.69$; ID2 mean accuracy = 69.1 ± 0.6%, WSR = 171.0, $p < .001$, RBC = 1.000, $BF_{10} = 586.84$). Critically, accuracy was also well above chance for ensemble summary expressions (ID1 mean accuracy = 74.8 ± 1.5%, WSR = 105.0, $p < .001$, RBC = 0.294, $BF_{10} = 1{,}096.43$; ID2 mean accuracy = 70.0 ± 2.9%, WSR = 166.0, $p < .001$, RBC = 0.942, $BF_{10} = 672.83$). An inspection of the generated face space (Fig. 3A) as well as of image reconstructions (see Fig. 3B for representative examples) confirmed that they capture visual characteristics informative of emotional expression (e.g., valence).

Thus, our method was successful for both single face and ensemble stimuli and validated the ability of behavior-based image reconstruction to recover and visualize summary expression representations from face

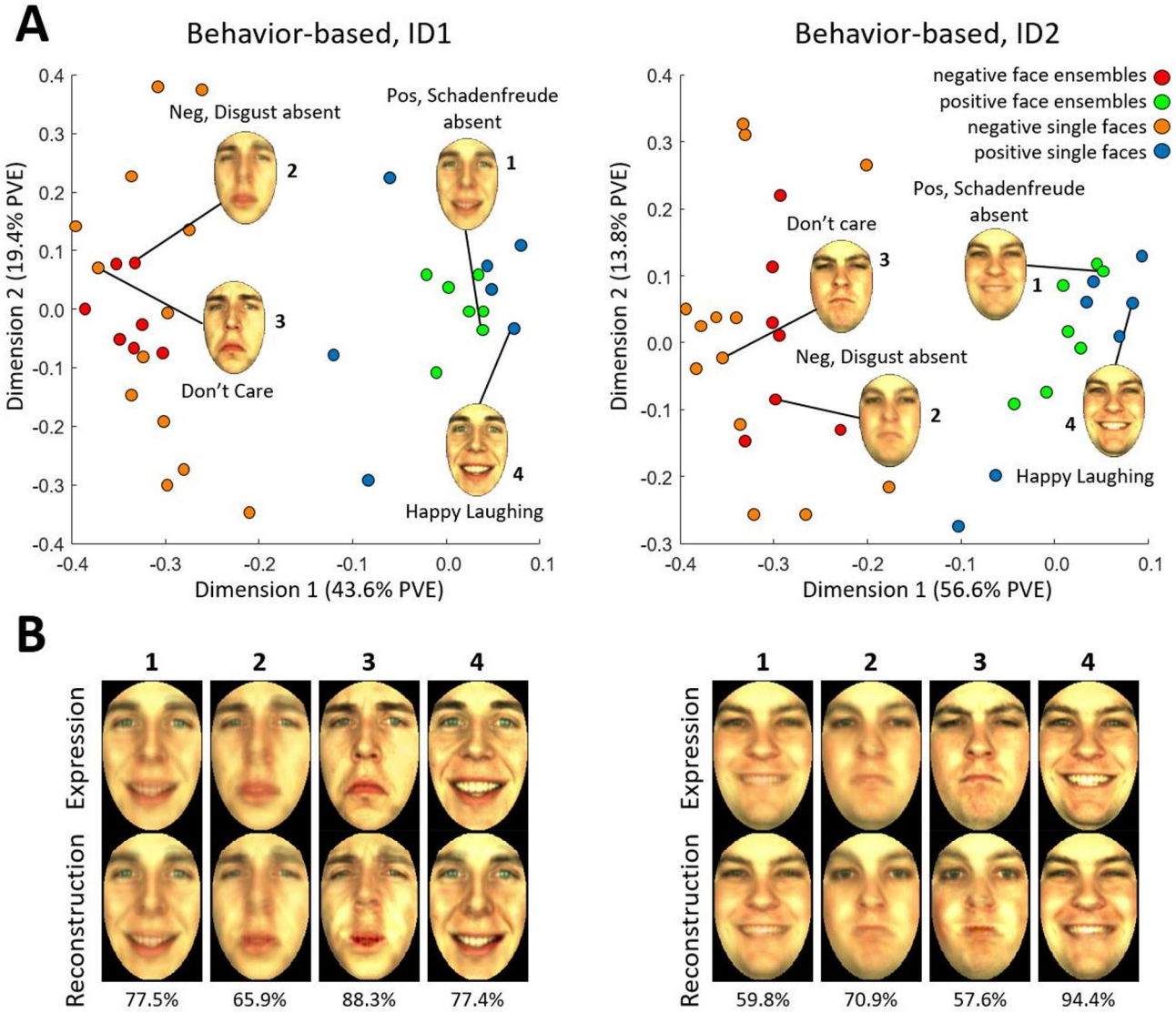

Fig. 3. Joint ensemble and single-face space along with examples of behavior-based image reconstructions from Experiment 1. Ensemble representations were projected in an expression face space and reconstructed from its structure, separately for each facial identity (left: ID1; right: ID2). (**A**) Both single faces and ensembles evinced a marked separation by valence in face space as well as clear within-valence variability: plots show the first two dimensions and percent variance explained (PVE) of a joint single-ensemble face space averaged across participants for visualization purposes (each ensemble, marked by red and green circles, consists of 6 of the 7 within-valence expressions, labelled based on the absent expression; *pos* positive-expression ensemble, *neg* negative-expression ensemble). (**B**) Image reconstructions capture visual representations of ensembles (1 and 2) and single-face stimuli (3 and 4)—ensemble stimuli are replaced in the upper row by the pixelwise average of their individual faces. Examples of positive and negative expressions are displayed along with their reconstruction accuracy (averaged across participants).

ensembles. Further, theoretically, the present results demonstrate the benefit of considering single faces and ensembles in the context of a shared representational face space. In the following experiments, we employ a center-surround paradigm and investigate how the presence of a central face, along with manipulations of distributed (Experiment 2) or focused (Experiment 3) attention, impact the representation of ensemble expressions.

Experiment 2

Experiment 2 uses a center-surround paradigm, which involves placing a centrally-located face within an ensemble stimulus, to investigate the impact of distributed attention on ensemble perception. Here, ensemble stimuli are labeled as consistent if the valence of the central face matches that of the surrounding faces, and inconsistent otherwise.

The results of Experiment 1 conform to previous work[25] and suggest comparable weight for faces in the display when synthesizing a summary representation via image reconstruction. Critically though, here we predict that, even when attention is distributed across all seven faces, the central face will be differently weighted relative to the surround. For inconsistent ensembles, we anticipate one of two outcomes: (1) the central face appears more salient, and, thus, biases the reconstruction towards it; or (2) the central face is discarded from the representation due to its status as an outlier[11]. In light of this, the approach followed below focuses not on the reconstruction accuracy per se, but rather on the similarity of the reconstruction to relevant components of the facial stimuli.

Participants

Forty new participants were recruited via Prolific with the same inclusion criteria as in Experiment 1. Two participants experienced technical difficulties and their data were lost. Of the remaining 38 (19 females, 19 males; age: 20–35 years), all met inclusion criteria based on internal consistency between repeated stimulus pairs.

Stimulus design

Ensemble stimuli included a central face (Fig. 2C, D), which, for consistent ensembles, was the previously left-out facial expression (see Experiment 1, Procedure). Consequently, all consistent ensembles now shared the same seven faces with the same valence, and, by extension, the same pixelwise average. A total of 14 ensembles, seven for each valence, were generated by having each face take a turn at the center position.

For inconsistent ensembles, the face with the weakest valence estimate (i.e., with the lowest absolute value based on pilot ratings) was omitted from the surround. Faces of the opposite valence, in addition to neutral, were swapped in as a central face. This resulted in eight positive inconsistent ensembles (i.e., seven with a negative central face amidst positive surrounds, and one with a neutral central face), and eight negative inconsistent ensembles (i.e., seven with a positive central face amidst negative surrounds, and one with a neutral central face), for a total of 30 ensemble stimuli across valence and consistency. Notably, since the center face differed across inconsistent ensembles with the same surround valence, each pixelwise ensemble average varied slightly by virtue of sharing six out of seven faces. Finally, S-S pairs and all other stimulus parameters were unchanged from the previous experiment.

Procedure

The procedure was similar to that in Experiment 1 (e.g., participants were asked to compare the perceived emotion of a single face or of an entire ensemble relative to that of a single-face probe). However, for the E-S block, the larger number of ensembles, relative to single face stimuli, resulted in 600 E-S pairs. Hence, to mitigate fatigue and to maintain comparable completion time with the previous experiment, only 40 E-S pairs and 57 S-S pairs were repeated.

Participants were tasked with fixating the central face while keeping attention distributed across the entire ensemble (i.e., distribute attention without looking around). To encourage distributed attention, we included 12 oddball trials in which Asian faces were presented in the surround, and an additional 12 trials in which an Asian face was presented in the center. These additional faces, displaying neutral expressions of adult males, were selected from the Chicago Face Database[35], cropped, masked, and contrast-matched to ID1 and ID2 face images. On such trials, participants were instructed to press the spacebar instead of making a similarity rating. These trials were excluded from the main analysis.

Overall, the experiment consisted of 911 trials across both blocks (190 S-S pairs plus 57 repetitions; and 600 E-S pairs plus 40 repetitions and 24 oddball trials). One-minute breaks occurred every 62 trials for S-S, and every 83 trials for E-S. Average completion time for Experiment 2 was 56 min.

Results and discussion

Analysis of oddball trials indicated ceiling-level performance (average sensitivity for center oddballs across participants: d' = 4.49, hit rate = 92.1%; average sensitivity for surround oddballs: d' = 3.54, hit rate = 76.3%; false alarm rate = 1.7%), suggesting that participants were able to distribute their attention over the ensemble stimuli, with an advantage for the center. Importantly, S-S ratings for both identities were highly consistent with their counterparts in the first experiment as indicated by correlations of ratings averaged across participants (ID1: $r_{188} = 0.955$, $p < .001$, 95% CI = [0.941, 0.966], $BF_{10} = 3.11 \times 10^{97}$; ID2: $r_{188} = 0.956$, $p < .001$, 95% CI = [0.941, 0.946], $BF_{10} = 4.70 \times 10^{97}$).

The reconstruction procedure follows the same approach as in Experiment 1. Here, for ensembles, accuracy is estimated relative to the average of the seven faces, including the central one, as opposed to only the six surround faces. Reconstruction accuracy, averaged across ID1 and ID2, was, again, significantly above chance for single faces (mean accuracy = 67.5 ± 0.9%, WSR = 740.0, $p < .001$, RBC = 0.997, $BF_{10} = 14{,}025.10$), for ensembles with consistent center-surround expressions (mean accuracy = 65.1 ± 1.6%, WSR = 719.0, $p < .001$, RBC = 0.941,

BF_{10} = 3,649.52), and, notably, for inconsistent ensembles (mean accuracy = 54.1 ± 1.5%, WSR = 475.5, p = .031, RBC = 0.353, BF_{10} = 4.46), albeit with poorer accuracy when compared to consistent ensembles (mean difference = 11.0 ± 2.3%, WSR = 638.0, p < .001, RBC = 0.778, BF_{10} = 532.26). This suggests that central faces may have biased ensemble perception towards them and away from the surround. Additional analyses will evaluate this possibility in conjunction with the results of the next experiment.

Experiment 3

In the previous experiment, participants attended to the entire ensemble (i.e., both the center and the surround). In Experiment 3, we ask participants to attend only to the central face (i.e., the target), and assess its similarity to the single-face probe. Cross-experiment comparisons will evaluate how changing the scope of attention impacts the perception of the central face versus its surrounding faces. Poor reconstruction accuracy from inconsistent stimuli does point to an unequal integration of center-surround faces. We anticipate that, when comparing the results of Experiment 3 (focused attention) to Experiment 2 (distributed attention), focused attention to the center face will bias the ensemble representation towards it, and away from the surrounding six faces.

A total of 40 participants were recruited via Prolific. The data from two participants were lost due to technical difficulties. The rest met inclusion criteria based on consistent pairwise ratings, resulting in a final sample of 38 participants (18 females, 20 males; age: 20–35 years).

The procedures were the same as for Experiment 2, except that participants were asked to rate similarity based solely on central faces. Also, oddball trials only involved changes in the central face, to encourage focused attention towards the center and away from the surround.

Results

Oddball detection performance reached ceiling (average d' = 5.39, hit rate = 97.7%; false alarm rate = 0.1%), presumably due to the predictability of its location and the narrower scope of attention. Importantly, S-S ratings for both facial identities were highly consistent with those of the prior experiment (ID1: r_{188} = 0.953, 95% CI = [0.938, 0.965], $BF_{10} = 6.34 \times 10^{95}$, p < .001; ID2: r_{188} = 0.956, p < .001, 95% CI = [0.942, 0.967], $BF_{10} = 1.60 \times 10^{98}$). Thus, across all three experiments, participants showed highly similar inter-rater reliability for single faces, which facilitates cross-experiment comparisons in the next section.

Once again, reconstruction was significantly above chance for single faces (mean accuracy = 69.0 ± 0.7%, WSR = 741.0, p < .001, RBC = 1.000, BF_{10} = 191,194.70) and consistent ensembles (mean accuracy = 67.1 ± 1.1%, WSR = 741.0, p < .001, RBC = 1.000, BF_{10} = 46,392.01). However, in line with our expectation, inconsistent ensembles did not yield above-chance accuracy (mean accuracy = 46.7 ± 0.9%, WSR = 109, p = .999, BF_{10} = 0.08). Subsequent results are described in the next section alongside those from Experiment 2.

Collective results and discussion of experiments 2 and 3

Here, we evaluate ensemble reconstructions from Experiments 2 and 3 (n = 76) based on their similarity to relevant stimuli. This approach aims to describe the visual content of the reconstructions and the factors which impact it, beyond mere reconstruction success.

Two main sets of analyses were conducted to investigate the influence of the central face on ensemble representations (see Table 1 for summary): first, relative to exemplar faces in the surround, and, second, relative to the average of surround faces (i.e., the summary representation of the surround). For simplicity, and given the smaller number of trials available for this condition, inconsistent ensembles with a neutral face were excluded from analyses. The distance between ensemble reconstructions and stimulus components was computed via an L2 pixelwise image metric, with lower values denoting higher visual similarity.

When estimating the distance between the reconstruction and the surrounding exemplars, L2 distances were computed between (1) the reconstruction and (2) each of the six surrounding faces separately. These six measurements were then averaged to provide a single metric. In contrast, when estimating the distance between the reconstruction and the ensemble summary, the six surround exemplars underwent pixelwise averaging to derive a summary metric, and the L2 distance was computed between the reconstruction and this summary (see Supplemental Methods for an explanation of how we controlled for spatial smoothing when comparing the reconstruction to a pixelwise average).

The perception of central faces versus surround exemplars

The distance between ensemble reconstructions was computed relative to: (1) the six exemplar faces from the surround; (2) the central face, and (3) other faces not included in the ensemble. This last comparison involved,

Comparison	Reconstruction (Fig. 4)	Reconstruction (Fig. 5)
To surround	The average of L2 distances between the reconstruction and each of the 6 exemplars in the surround	The L2 distance between the reconstruction and the average of the 6 exemplars in the surround
To center face	The L2 distance between the reconstruction and the center face of the ensemble stimulus	The L2 distance between the reconstruction and the center face of the ensemble stimulus
To other faces	The average of L2 distances between the reconstruction and each of the faces not used in the ensemble stimulus	The average of L2 distances between the reconstruction and each of the faces not used in the ensemble stimulus

Table 1. Summary of comparisons between image reconstructions and different components of the ensemble stimuli visualized in Figs. 4 and 5. Reconstructions were compared to either the surrounding faces of the ensemble, the central face, or other faces not included in the ensemble stimulus.

for consistent ensembles, faces of opposite valence (e.g., all negative expressions for a positive ensemble), while, for inconsistent ensembles, it involved faces which shared the same valence with the central face (e.g., all positive expressions other than the central one). Distances to surround faces and distances to other faces were averaged to provide a single estimate for each. These measures were computed separately for inconsistent and consistent stimuli in Experiments 2 and 3.

First, data were analyzed with a 4-way mixed-design ANOVA using 2 between-subject (attention: distributed or focused), 2 within-subject (consistency: consistent or inconsistent ensembles), and 3 within-subject (distance: to surround faces, central face, or other faces) factors, with the fourth factor (identity: ID1 or ID2) considered as a blocking variable. Sphericity was assessed via Mauchly's test and Greenhouse-Geisser corrections were applied to omnibus effects where violations occurred. The attentional task and the identity blocking variable overall had equal variances via Levene's test. Identity was included as a random factor in Bayesian hypothesis testing for follow-up between-experiment comparisons.

Regarding the three main effects, results were significant for distance ($F_{2.1,49.107.07} = 485.40$, $p < .001$, $\eta_p^2 = 0.871$). We only noticed trends for the effect of attention ($F_{1,72} = 3.19$, $p = .078$) and consistency ($F_{1,72} = 2.91$, $p = .093$). Two-way interactions were significant for attention-by-distance ($F_{1.49,107.07} = 21.16$, $p < .001$, $\eta_p^2 = 0.227$) and consistency-by-distance ($F_{1.37,98.35} = 473.45$, $p < .001$, $\eta_p^2 = 0.868$), but not for the consistency-by-attention interaction ($F_{1,72} = 0.55$, $p = .460$). The three-way interaction between consistency, distance, and attention was significant ($F_{1.36,98.35} = 10.92$, $p < .001$, $\eta_p^2 = 0.132$). Bayesian modelling provided support for the three main effects, the two-way interactions of consistency by distance and attention by distance, and the three-way interaction for consistency by distance by attention ($BF_{10} = 2.07 \times 10^{209}$). All results, which also include the identity blocking variable, are detailed in Supplemental Table 1.

Next, considering the interactions above, we computed three sets of multiple comparisons to further investigate how distance and consistency impact ensemble representations in each attentional task, as well as how distributed versus focused attention impacts representations across experiments—see Fig. 4A.

The first set of comparisons examined how the distance to the ensemble representation varies relative to surround faces, the central face, and to other faces, separately for each type of ensemble and attention task. The results show that all three distances are significantly different from one another for both consistent and inconsistent ensembles during both distributed and focused attention (all mean differences $> 1.1 \pm 0.2$, all $ts > 4.97$, all $ps < 0.001$, all $ds > 1.25$, all $BF_{10} > 2.05$). These results indicate that, overall, reconstructions of the summary percept were closer to the central face compared to both surround faces and to other faces. Regarding the latter two, reconstructions were closer to the surround faces compared with other faces for consistent ensembles but closer to other faces compared to surround faces for inconsistent ensembles. This is expected, given that, for inconsistent ensembles, other faces featured expressions sharing the same valence with the center face.

The next set of comparisons assessed the impact of consistency by examining distances for consistent versus inconsistent ensembles, separately for distributed versus focused attention. Overall, we found significant differences for all three distances both for distributed attention (all mean differences $> 1.0 \pm 0.2$, all $ts > 5.18$, all $ps < 0.001$, all $ds > 0.37$, all $BF_{10} > 676.99$) and focused attention (all mean differences $> 0.6 \pm 0.2$, all $ts > 2.96$, all $ps < 0.004$, all $ds > 0.21$, all $BF_{10} > 1{,}932.42$). These results indicate that the distance from the reconstruction to the surround and to the central face were closer for consistent versus inconsistent ensembles, but the distance from the reconstruction to the other faces was closer for inconsistent ensembles.

Last, the third set of comparisons examined corresponding surround, center, and other-face distances between distributed versus focused attention. For consistent stimuli, reconstructions were closer to central faces during focused versus distributed attention (mean difference $= 0.6 \pm 0.2$, $t_{144} = 2.98$, $p = .010$, $d = 0.69$, $BF_{10} = 14.50$). However, attention had no significant effect on the distance to surround or other faces (both mean differences $< 0.2 \pm 0.2$, $ts < 0.86$, $ps > 0.485$, $BF_{10} < 0.49$). For inconsistent stimuli, reconstructions were closer to ensemble surrounds during distributed attention (mean difference $= 1.2 \pm 0.2$, $t_{144} = 5.88$, $p < .001$, $d = 1.35$, $BF_{10} = 484.99$), but focused attention led to closer distances to the center and to other faces (both mean differences $> 0.7 \pm 0.2$, $ts > 3.41$, $ps < 0.003$, $ds > 0.79$, $BF_{10} > 160.79$). These results indicate that focused attention leads to higher similarity between the ensemble representation and the central face, irrespective of ensemble consistency, but to lower similarity relative to surround faces in an inconsistent ensemble.

The current results reflect alterations in face space structure for focused versus distributed attention (see Fig. 4B), in that ensembles that share the central face with single faces are clustered more tightly within this multidimensional space for focused attention. This illustrates how face space can be modulated by attentional mechanisms which impact visual representations.

The perception of central faces versus the surround average

Our results above indicate that ensemble representations are closer to central versus surround face exemplars. In other words, when encoding individual faces from an ensemble, central faces provide more perceptual weight than surround faces, irrespective of whether attention is distributed across the entire ensemble or focused on a central face. Next, we examine the contribution of the average expression of the surround versus the central face. Specifically, we compute the distance between ensemble reconstructions and the pixelwise average of the six surround faces, and, then, we compare this to the distances between the reconstructions of both central and other faces. However, here, image blur introduced by pixelwise averaging may confound the comparison relative to single faces. To this end, we applied a 2D Gaussian filter to match the spatial frequency power spectra between pixelwise averages and single face stimuli (Fig. 5A, see Supplemental Methods). Next, distances were recomputed between the reconstruction and the pixelwise average of the surround, the center, and the other faces. Then, data was subjected to the same analysis of variance as above.

The results revealed significant main effects of distance ($F_{1.42,102.10} = 899.37$, $p < .001$, $\eta_p^2 = 0.926$) and attention ($F_{1,72} = 5.10$, $p = .027$, $\eta_p^2 = 0.066$), and a trend for the effect of consistency ($F_{1,72} = 2.99$, $p = .088$). The

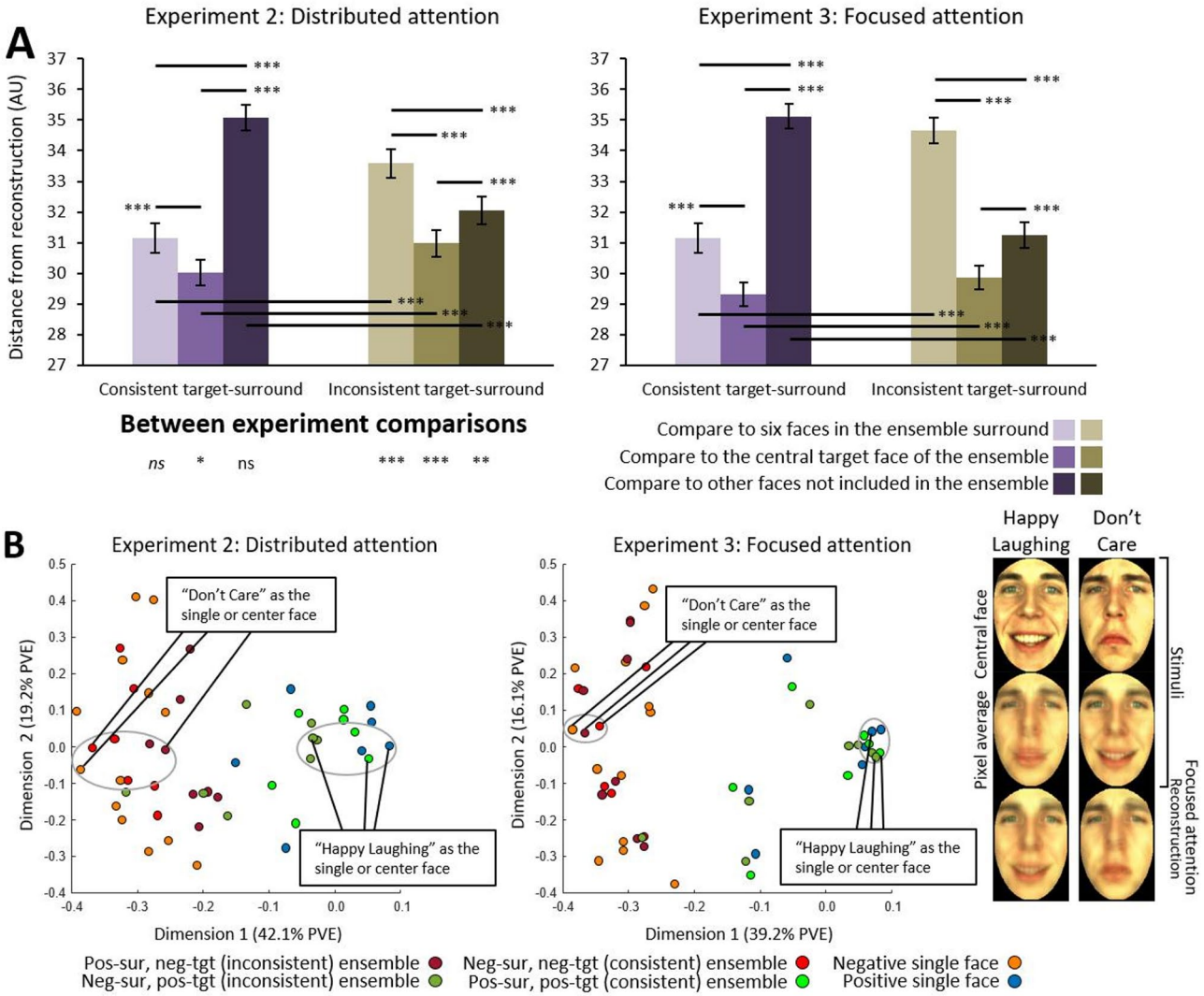

Fig. 4. Image distances between ensemble summary reconstructions and corresponding surround faces, central faces, and other face stimuli across ensemble consistency and attentional task. Summary ensemble representations, as revealed by image reconstruction, are impacted by several factors. (**A**) The central face outweighs the contribution of surround exemplars as the reconstruction is closer to the former (via an L2 pixelwise metric in arbitrary units); consistency reduces the distance between the reconstruction and both central and surround faces; focused attention increases the contribution of the central face and, for inconsistent ensembles, decreases that of the surround (between-experiment comparisons are shown below the left bar plot). (**B**) Ensembles sharing the central face with single faces are less clustered in face space during distributed attention (left) compared to focused attention (right). The plots depict two dimensions of face space, averaged across participants, for one facial identity during Experiments 2 and 3. Corresponding central face stimuli, pixelwise ensemble averages and ensemble reconstructions are shown for focused attention (right)—the central face has minimal impact on the pixelwise average of the stimulus, but it appears to dominate the ensemble representation (*ns*: nonsignificant, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$, Holm-corrected; error bars indicate ± *SE* across participants; *PVE* percent variance explained).

two-way interaction of consistency and attention was not significant ($F_{1,72} = 0.46$, $p = .498$), but consistency-by-distance ($F_{1.27,91.40} = 454.17$, $p < .001$, $\eta_p^2 = 0.863$), and attention-by-distance interactions ($F_{1.42,102.10} = 21.29$, $p < .001$, $\eta_p^2 = 0.228$) were significant. The three-way interaction between consistency, distance, and attention was also significant ($F_{1.27,91.40} = 11.80$, $p < .001$, $\eta_p^2 = 0.141$). Finally, Bayesian hypothesis testing supported a model with the three main effects, a three-way interaction for consistency by distance by attention, and all two-way interactions between the three main effects ($BF_{10} = 1.46 \times 10^{194}$)—see Supplemental Table 1 for all results as well as the effects of the blocking variable. Interestingly, these findings coincide with the previous results with the exception of a significant main effect of attention in this latter analysis.

The same three sets of post-hoc tests described previously were recomputed here (Fig. 5B). First, an examination of distances found, again, significant pairwise differences between surround averages, centers, and other faces for both consistent and inconsistent ensembles during both attention tasks (all mean differences > 0.7 ± 0.3, all

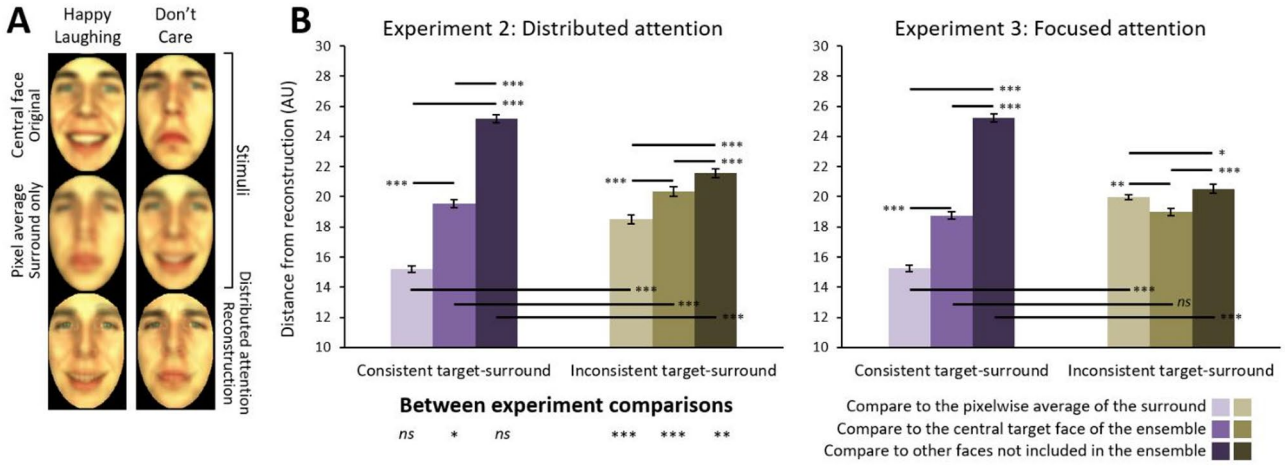

Fig. 5. Image distances between ensemble summary reconstructions and corresponding surround averages, central faces, and other face stimuli across ensemble consistency and attentional task. Summary ensemble representations, as revealed by image reconstruction, evince an interplay between consistency and attentional scope. (**A**) Examples of central face stimuli, matched in image quality with pixelwise surround averages, are shown along with their reconstructions for distributed attention (Experiment 2). (**B**) Face position (central versus surround), consistency and attention impact ensemble representations as found above. However, the reconstruction is closer to the surround average compared with center faces in all cases, except for inconsistent ensembles during focused attention. Between-experiment comparisons are shown below the left bar plot (*ns*: nonsignificant, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$; Holm-corrected; error bars indicate ± *SE* across participants).

ts > 2.45, all *ps* < 0.016, all *ds* > 0.62, all $BF_{10} > 0.56$). However, in contrast to our previous results, reconstructions were now closer to surround averages than to the central face, but still closer to the central face than to the other faces. Critically though, one exception was noticed for inconsistent ensembles during focused attention, in which case the reconstruction was closer to the central face than the surround. This finding has important theoretical implications which we elaborate on in the General Discussion.

Second, the comparison of consistent versus inconsistent ensembles revealed a significant advantage for the former for surround averages, centers, and other faces in the distributed attention task (all mean differences = 0.8 ± 0.2, all *ts* > 3.39, all *ps* < 0.002, all *ds* > 0.73, all $BF_{10} > 13.82$). However, this advantage was only present for surrounds and other faces (both mean differences > 3.3 ± 0.2, both *ts* > 13.71, both *ps* < 0.001, both *ds* > 2.96, both $BF_{10} > 3.08 \times 10^{16}$), not for central faces (mean difference = 0.2 ± 0.2, $t_{144} = 0.99$, $p = .324$, $BF_{10} = 0.95$), during focused attention. In summary, for both distributed and focused attention, distances between reconstructions of the ensemble percept and the surround averages were smaller for consistent, compared with inconsistent, ensembles, but the reverse pattern was observed for other faces (i.e., smaller distances for inconsistent compared with consistent ensembles).

Third, all differences were significant for between-subject comparisons across experiments (mean differences > 0.8 ± 0.3, all *ts* > 2.96, all *ps* < 0.011, all *ds* > 0.68, all $BF_{10} > 25.54$), except those for consistent surround averages and other faces (both mean differences < 0.1 ± 0.3, both *ts* < 0.58, both *ps* > 0.564, both $BF_{10} < 0.36$). Thus, distances were comparable between consistent surrounds across distributed and focused attention. However, reconstructions were closer to central faces during focused attention irrespective of ensemble consistency. Also, they were closer to the surround average for inconsistent ensembles during distributed, compared with focused, attention.

In comparing Figs. 4 and 5, we note two important points. First, the marked dominance of the central exemplar over the surround exemplars (Fig. 4) contrasts that of the surround average over the central exemplar (Fig. 5). In other words, the central face exerts greater influence on the reconstruction than the surrounding exemplars, but the summary representation of the surround exerts a stronger influence than that of the central face. Second, this latter result is reversed during focused attention to inconsistent ensembles (Fig. 5), though the advantage of the central face here is more modest. Thus, even when an exemplar does dominate the perception of a surround summary, the advantage is rather limited.

To assess whether the differences between the two analyses, in Figs. 4 and 5, are due to genuine experimental effects rather than byproducts of the analytical techniques involved we further evaluated within-condition consistency across the two. Specifically, we correlated the distances between reconstructions and surround exemplars across participants with the distances between reconstructions and surround averages. As expected, the correlations were significant in all four conditions (during distributed/focused attention for consistent/inconsistent ensembles; all *rs* > 0.497, all *ps* < 0.006, Bonferroni corrected, all $BF_{10} > 25.58$). Given that consistency is preserved across analysis type, we argue that the differences above are grounded in experimental effects.

Last, we note that our investigation above takes aim at differences between experimental conditions averaged across participants. Yet, it is possible that our data also evince systematic differences across participants (e.g., due

to attentional deployment), which cut across experimental conditions. Additional analyses and results, detailed in the Supplemental Results and Supplemental Fig. 2, provide preliminary evidence for this hypothesis.

General discussion

Overall, our work reveals a flexible relationship between attentional scope and the structure of ensemble stimuli. Our results address the mixed findings regarding the role of attention in the ensemble literature. Additionally, they shed light, with the aid of image reconstruction, on the visual representations for individual faces and summary percepts as well as on their interaction. Arguably, these findings yield new insights into the cognitive mechanisms mediating crowd perception.

Briefly, in Experiment 1, we showed the benefit of examining ensemble summary representations in an expression space framework, which has typically been reserved for single faces. First, we show a clear demarcation between negative and positive expressions. Beyond this though, we find that ensemble reconstructions evince adequate spatial spread in face space (Fig. 3). This is particularly interesting given their homogeneity (i.e., recall that each ensemble shared five out of six exemplars with other ensembles of the same valence). To be clear, this shows that the structure of face space is sensitive even to slight differences in ensemble percepts. Capitalizing on this structure, we used behavior-based image reconstruction to recover the visual appearance of summary representations. Importantly, this validates the present methodology in the study of expression ensembles.

Building upon these results, Experiments 2 and 3 used a center-surround paradigm to assess the impact of attention on expression for consistent versus inconsistent ensembles. Our results revealed two notable findings: (1) focused attention enhances central face representation, and (2) the summary surround (i.e., the average of surround faces) outweighs the central face in ensemble encoding in all cases except when attention is focused on an inconsistent central face. We elaborate on these findings and their implications below.

First, focused attention biases the summary representation towards the central face. At the same time, the influence of the surround remains unchanged for consistent stimuli despite changes in attentional scope—for both surround face exemplars (Fig. 4A) and for their average (Fig. 5B). Overall, these results illustrate the potency of ensemble statistics, in agreement with claims that ensemble encoding can occur in the absence of direct visual attention[16–18,36]. However, they also showcase the modulatory role of attention, which can enhance the contribution of the central face without a necessary cost for the surround faces.

Further, the dominance of the surround summary during focused attention can be characterized as a failure to filter and as evidence for the strength of implicit processing. Specifically, our manipulation of attention across experiments relates to the distinction between implicit and explicit ensemble processing[37]. Prior work[38] indicates that implicit judgements, which do not target directly the ensemble summary, rely on rich visual information. Interestingly, the availability of this information may be limited for explicit judgements, which do target directly the summary representation. However, other work found that explicit processing with non-face objects, such as simple geometrical figures, tends to be more precise than implicit processing[37]. Our results with face ensembles agree with the richness of visual information in implicit processing, given the impact of the surround on reconstructions even during center-focused attention. However, we also find its impact to be amplified by distributed attention, thus, evincing a benefit for explicit processing.

Second, the summary representation of the surround outweighs that of the central face in all cases except during focused attention to inconsistent ensembles (Experiment 3). This is a novel repulsion effect, running contrary to the discounting of outlier expressions in face ensembles[11]. Theoretically, this supports our claim for a flexible role of attention in ensemble processing, offering a new perspective on this topic. Specifically, the role of attention varies as a function of stimulus properties (e.g., consistency).

Further, we note the facilitatory role of consistency for ensemble stimuli[18], applied, in this case, to center faces and surround faces. Specifically, our results show that the consistency between central and surround faces can benefit the processing of both. This occurs irrespective of attentional scope. Prior work using EEG also suggests that central and surround faces are processed separately by the visual system[21]. Specifically, for facial identity, both center and surrounding faces are decoded separately starting at around 100 ms and 220 ms from stimulus onset, respectively. Moreover, information about center-surround consistency is decoded as early as 150 ms. Whether the target and surround faces are processed by the same underlying anatomical structures remains subject to future research.

Thus, collectively, our findings reveal the impact of ensemble structure and attention on face ensemble encoding. While direct attention may not be required to extract summary ensemble information, it can modulate its representation. This is consistent with prior work on both emotional expression[39] and lower-level visual processing[40–42]. Also, despite claims regarding the separability of focused attention and ensemble encoding[43], the present results show that they can work in unison to dynamically alter ensemble representations.

To illustrate our findings with a real-world example, imagine a lecturer focusing their attention towards a student during class. Information about the surrounding crowd would be well captured by the visual system insofar as the collective expression remains consistent with that of the attended student. However, if that student were to exhibit a negative expression (e.g., frustration with the material) amidst a positively-expressive class, our findings suggest that the visual system would commit more resources to encoding the student's face, at the expense of information about the crowd. If the negatively expressing student was not directly attended to, their inconsistent expression – now an outlier – may well be discounted from the summary representation[11].

Additionally, our results also address conflicting findings from other center-surround paradigms in the study of face ensembles. For instance, on the one hand, the presence of a central face within a 3×3 display was found to impact the emotional summary representation of the overall ensemble[20]. On the other hand, recent research has emphasized the impact of surrounding faces on the perceived expression of a central face[19]. Critically, our results explain such discrepancies by noting the interaction between attention and the properties of ensembles, such as their center-surround consistency.

Last, we note the benefit of image reconstruction, which provided us with the opportunity to visualize and rigorously assess the content of ensemble representations. In particular, it allowed us to evaluate the reliance of reconstructed percepts not only on individual faces from an ensemble stimulus, but also on surround averages. Thus, we were able to directly assess the quality of the summary percept despite participants not encountering it as a stimulus per se.

In summary, across three experiments we evaluated the representational content of ensemble facial expression using a center-surround paradigm and image reconstruction. Our results reveal that attentional requirements in ensemble processing may interact with the structure of the stimulus itself. Given the possibility of a domain-general mechanism for ensemble processing[44] our findings may well generalize to visual categories other than faces (though, to be clear, our present results speak only to face ensembles). Last, methodologically, our findings demonstrate the utility of image reconstruction as a tool for assessing the nuanced role of attention in ensemble processing.

Data availability

Data availability, including code and scripts used to conduct and analyze the experiments, as well as statistical results and reconstructed images, are available online via OSF (https://osf.io/kgzbh/?view_only=e3c416ec3ab1411182b958af2ac13a9e).

Received: 28 November 2024; Accepted: 20 May 2025
Published online: 28 May 2025

References

1. Whitney, D. & Yamanashi Leib, A. Ensemble perception. *Rev. Psy*. **69**, 105–129 (2018).
2. Corbett, J., Utochkin, I. & Hochstein, S. *The Pervasiveness of Ensemble Perception: Not Just Your Average Review (Elements in Perception)* (Cambridge University Press, 2023).
3. de Fockert, J. W. & Wolfenstein, C. Rapid extraction of mean identity from sets of faces. *Q. J. Exp. Psy*. **62** (9), 1716–1722 (2009).
4. Neumann, M. F., Schweinberger, S. R. & Burton, A. M. Viewers extract mean and individual identity from sets of famous faces. *Cognition* **128** (1), 56–63 (2013).
5. Florey, J., Clifford, C. W. G., Dakin, S. & Mareschal, I. Spatial limitations in averaging social cues. *Sci. Rep.* **6**, 32210 (2016).
6. Sama, M. A., Nestor, A. & Cant, J. S. Independence of viewpoint and identity in face ensemble processing. *J. Vis.* **19** (5), 2. https://doi.org/10.1167/19.5.2 (2019).
7. Haberman, J. M. & Whitney, D. Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* **17** (17), 39. https://doi.org/10.1016/j.cub.2007.06.039 (2007).
8. Wolfe, B. A., Kosovicheva, A. A., Yamanashi Leib, A., Wood, K. & Whitney, D. Foveal input is not required for perception of crowd facial expression. *J. Vis.* **15** (4), 1–13 (2015).
9. Ji, L., Rossi, V. & Pourtois, G. Mean emotion from multiple facial expressions can be extracted with limited attention: evidence from visual erps. *Neuropsychologia* **111**, 92–102 (2018).
10. Haberman, J. M. & Ulrich, L. Precise ensemble face representation given incomplete visual input. *I-Perc* **10** (1), 14. https://doi.org/10.1177/2041669518819014 (2019).
11. Haberman, J. M. & Whitney, D. The visual system discounts emotional deviants when extracting average expression. *Att Perc Psychophys*. **72** (7), 1825–1838 (2010).
12. Avci, B. & Boduroglu, A. Contributions of ensemble perception to outlier representation precision. *Att Perc Psychophys*. **83**, 1141–1151 (2021).
13. Elias, E., Padama, L. & Sweeny, T. D. Perceptual averaging of facial expressions requires visual awareness and attention. *Consc Cogn*. **62**, 110–126 (2018).
14. Jackson-Nielsen, M., Cohen, M. A. & Pitts, M. A. Perception of ensemble statistics requires attention. *Consc Cogn.* **48**, 149–160 (2017).
15. Myczek, K. & Simons, D. J. Better than average: alternatives to statistical summary representations for rapid judgments of average size. *Perc Psychophys*. **70** (5), 772–788 (2008).
16. Alvarez, G. A. & Oliva, A. The representation of simple ensemble visual features outside the focus of attention. *Psychol. Sci.* **19** (4), 392–398 (2008).
17. Bronfman, Z. Z., Brezis, N., Jacombson, H. & Usher, M. We see more than we can report: cost free color phenomenality outside focal attention. *Psychol. Sci.* **25** (7), 56. https://doi.org/10.1177/0956797614532656 (2014).
18. Chen, Z., Ran, Z., Xiaolin, W., Ren, Y. & Abrams, R. A. Ensemble perception without attention depends upon attentional control settings. *Att Perc Psychophys*. **83** (3), 1240–1250 (2020).
19. Wu, Y. & Ying, H. The background assimilation effect: facial emotional perception is affected by surrounding stimuli. *I-Perc* **14** (4), 54. https://doi.org/10.1177/20416695231190254 (2023).
20. Dandan, Y. R., Ji, L., Song, Y. & Sayim, B. Foveal vision determines the perceived emotion of face ensembles. *Att Perc Psychophys*. **85**, 209–221 (2022).
21. Sama, M. A., Nestor, A. & Cant, J. S. The neural dynamics of face ensemble and central face processing. *J. Neuro*. **44** (7), 23. https://doi.org/10.1523/JNEUROSCI.1027-23.2023 (2024). (2024).
22. Nestor, A., Lee, A. C. H., Plaut, D. C. & Behrmann, M. The face of image reconstruction: progress, pitfalls, prospects. *Tr. Cogn. Sci.* **24** (9), 747–759 (2020).
23. Chang, C-H., Drobotenko, N., Ruocco, A. C., Lee, A. C. H. & Nestor, A. Perceptual and memory-based representations of facial emotions: associations with personality functioning, state affect and recognition abilities. *Cognition* **245**, 105724 (2024).
24. Nemrodov, D., Behrmann, M., Niemeier, M., Drobotenko, N. & Nestor, A. Multimodal evidence on shape and surface information in individual face processing. *NeuroImage* **184** (1), 813–825 (2019).
25. Roberts, T., Cant, J. S. & Nestor, A. Elucidating the neural representation and the processing dynamics of face ensembles. *J. Neuro*. **39** (39), 7737–7747 (2019).
26. Valentine, T., Lewis, M. B. & Hills, P. J. Face space: A unifying concept in face recognition research. *Q. J. Exp. Psy*. **69** (10), 1996–2019 (2015).
27. Barrett, L. F. & Russell, J. A. The structure of current affect: controversies and emerging consensus. *Curr. Dir. Psychol. Sci.* **8** (1), 10–14 (1999).
28. Calvo, M. G. & Nummenmaa, L. Perceptual and affective mechanisms in facial expression recognition: an integrative review. *Cogn. Emot.* **30** (6), 1081–1106 (2016).
29. Liu, M. et al. Facial expressions elicit multiplexed perceptions of emotion categories and dimensions. *Curr. Biol.* **32** (1), 200–209 (2022).

30. Peirce, J. et al. PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods*. **51**, 195–203 (2019).
31. Kaulard, K., Cunningham, D. W., Bülthoff, H. H. & Wallraven, C. The MPI facial expression Database—a validated database of emotional and conversational facial expressions. *PLoS ONE*. **7** (3), e32321. https://doi.org/10.1371/journal.pone.0032321 (2012).
32. Wagenmakers, E. J., Wetzels, R., Borsboom, D. & van der Maas H.L.J. Why psychologists must change the way they analyze their data: the case of Psi: comment on bem (2011). *J. Person Soc. Psychol.* **100** (3), 426–432 (2011).
33. Rouder, J. N., Morey, R. D., Speckman, P. L. & Province, J. M. Default Bayes factors for ANOVA designs. *J. Math. Psychol.* **56** (5), 356–374 (2012).
34. Ly, A., Verhagen, J. & Wagenmakers, E. J. Harold Jeffrey's default Bayes factor hypothesis tests: explanation, extension, and application in psychology. *J. Math. Psychol.* **72**, 19–32 (2016).
35. Ma, D. S., Correll, J. & Wittenbrink, B. The Chicago face database: A free stimulus set of faces and norming data. *Behav. Res. Methods*. **47**, 1122–1135 (2015).
36. Alvarez, G. A. & Oliva, A. Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proc. Nat. Acad. Sci.* **106** (18), 7345–7350 (2009).
37. Khayat, N., Pavlovskaya, M. & Hochstein, S. Comparing explicit and implicit ensemble perception: 3 stimulus variables and 3 presentation modes. *Atten. Percept. Psychophys*. **86** (2), 482–502 (2024).
38. Hansmann-Roth, S., Kristjánsson, Á., Whitney, D. & Chetverikov, A. Dissociating implicit and explicit ensemble representations reveals the limits of visual perception and the richness of behavior. *Sci. Rep.* **11**, 3899. https://doi.org/10.1038/s41598-021-83358-y (2021).
39. Ying, H. Attention modulates the ensemble coding of facial expression. *Perc* **51** (4), 276–285 (2022).
40. Chong, S. C. & Triesman, A. Attentional spread in the statistical processing of visual displays. *Perc Psychophys*. **67**, 1–13 (2005).
41. de Fockert, J. W. & Marchant, A. P. Attention modulates set representation by statistical properties. *Perc Psychophys*. **70**, 789–794 (2008).
42. Lin, Z., Gong, M. & Li, X. On the relation between crowding and ensemble perception: examining the role of attention. *PsyCh J.* **11** (6), 804–813 (2022).
43. Baek, J. & Chong, S. C. Ensemble perception and focused attention: two different modes of visual processing to Cope with limited capacity. *Psychol. Bull. Rev.* **27**, 602–606 (2020).
44. Chang, T-Y., Cha, O. & Gauthier, I. A general ability for judging simple and complex ensemble. *J. Exp. Psy Gen*. **153** (6), 1517–1537 (2024).

Author contributions

M.A.S designed the project, conducted analyses, and wrote the main manuscript. M.S provided code and conducted analyses, assisted in editing manuscript. J.S.C and A.N contributed equally to this work as co-senior authors, assisting in designing the project, provided funding, and edited manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-03289-w.

Correspondence and requests for materials should be addressed to M.A.S.