Journal of Vision (2019) 19(5):2, 1–17

Independence of viewpoint and identity in face ensemble processing

Marco A. Sama
Department of Psychology,
University of Toronto Scarborough, Toronto, Canada

Adrian Nestor*
Department of Psychology,
University of Toronto Scarborough, Toronto, Canada

Jonathan S. Cant*
Department of Psychology,
University of Toronto Scarborough, Toronto, Canada

Ensemble encoding refers to the visual system's ability to extract a summary representation from multiple items in a set—such as the mean identity of faces in a crowd—circumventing capacity limitations in visual working memory. In the present study we investigated face ensemble representations of higher level identity and lower level viewpoint with the aim of elucidating the extent of their overlap or independence. To this end, we used ensemble displays consisting of six face stimuli which could vary in identity, viewpoint, or both. Across three experiments, participants were asked to report an average identity, a single identity, an average viewpoint, or a single viewpoint, as cued by a central probe face following a stimulus display. In Experiment 1, we observed a dissociation between the processing of identity and viewpoint, as well as between average- and single-viewpoint extraction. Experiment 2 showed viewpoint-invariant identity processing across a wide range of mean viewpoints, spanning 120°. In Experiment 3, accuracy in reporting a response-relevant attribute was unaffected by changes in an irrelevant attribute. Participants were also capable of extracting both attributes simultaneously with little change in accuracy. Taken together, these results argue for the independence of identity and viewpoint in face ensemble processing.

Introduction

When investigating the visual processing of everyday items (e.g., faces or objects), most psychophysical work has focused on single items. In real-world settings, though, we often encounter multiple, simultaneously visible items exhibiting statistical redundancy, here referred to as an ensemble. For example, an ensemble of faces, like those in a crowd, can be characterized by many attributes, such as expression, identity, or viewpoint. Any of these attributes can be conveniently encoded by their summary statistics (e.g., the average expression, identity, or viewpoint of a face ensemble).

Encountering a large number of simultaneous stimuli introduces a challenge for the visual system (Cowan, 2010), since the ability to hold detailed representations of multiple individual stimuli in visual working memory (VWM) appears limited in capacity. Generally, it has been thought that VWM has a capacity limitation of about three or four items (Luck & Vogel, 1997; Raffone & Wolters, 2001), with ongoing debates regarding the actual capacity (Brady, Störmer, & Alvarez, 2016). Ensemble processing circumvents VWM limitations by encoding a summary representation from a set of stimuli, such as a mean identity from a crowd of faces, effectively pooling these statistical redundancies across the visual field into a single estimate per feature. The hallmark of ensemble processing is the ability to extract a summary representation from large sets of stimuli more accurately than the feature value of any individual item (for review, see Alvarez, 2011; Whitney & Leib, 2018). For instance, in an early demonstration, Ariely (2001) showed that observers could report the average size of a set of circles fairly accurately but were poor at reporting the size of any one particular circle from the set. This summary representation may partially explain the subjective experience of having a rich and detailed perceptual world despite the inability to accurately process all of it (Cohen, Dennett, & Kanwisher, 2016; Leib, Kosovicheva, & Whitney, 2016), and conforms to the visual system utilizing inference-based mechanisms for fast and efficient processing (Kersten, Mamassian, & Yuille, 2004; Purves, Monson, Sundararajan, & Wojtach, 2014).

Citation: Sama, M. A., Nestor, A., & Cant, J. S. (2019). Independence of viewpoint and identity in face ensemble processing. *Journal of Vision*, *19*(5):2, 1–17, https://doi.org/10.1167/19.5.2.

https://doi.org/10.1167/19.5.2 Received August 9, 2018; published May 6, 2019 ISSN 1534-7362

Since Ariely's (2001) study, there has been an increased interest in ensemble processing, with faces being one type of stimulus of particular interest. For humans, the ability to effectively process faces plays a significant role in daily activities. A range of information such as facial expression, gender, and, importantly, identity can be extracted from faces as supported, at the neural level, by an extensive network of cortical regions (Gobbini & Haxby, 2007; Ishai, 2008; Zhen, Fang, & Liu, 2013; Nestor, Plaut, & Behrmann, 2016). Viewpoint, a lower level facial attribute, can also be extracted from single faces by appealing to a multitude of cortical regions (Axelrod & Yovel, 2012; Kietzmann, Gert, Tong, & König, 2017; Kuo, Chen, & Chen, 2018; Ramírez, 2018).

Regarding the relationship between identity and viewpoint, for single-face perception identity recognition is more sensitive to viewpoint for unfamiliar than familiar faces (Bruce & Valentine, 1987; Logie, Baddeley, & Woodhead, 1987; Hill, Schyns, & Akamatsu, 1997; Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005), with a nonfrontal (three-quarter) view generally conferring the largest advantage. Moreover, early visual processing of single-face identity appears to be more sensitive to changes in viewpoint (Ewbank, Smith, Hancock, & Andrews, 2008; Caharel, Collet, & Rossion, 2015; Ramírez, 2018). With regard to neural processing, facial identity recruits a network of cortical regions at higher levels of the visual hierarchy (Gobbini & Haxby, 2007), and regions that subserve both attributes, such as the fusiform face area, exhibit at least some separability between the processing of identity and viewpoint (Guntupalli, Wheeler, & Gobbini, 2017). Further, viewpoint and identity appear to rely on different neural-encoding schemas with regard to sparseness or clustering (Dubois, de Berker, & Tsao, 2015), likely reflecting the encoding needs of visual attributes with different levels of complexity.

While an extensive body of work has investigated the processing of identity and viewpoint with single faces, much less is known in this respect about face ensembles, especially regarding the overlap or independence of such representations. To date, several studies have demonstrated efficient extraction of summary identity (Haberman & Whitney, 2007; de Fockert & Wolfenstein, 2009; Neumann, Schweinberger, & Burton, 2013) and viewpoint (Florey, Clifford, Dakin, & Mareschal, 2016), as well as other relevant features such as gaze direction (Sweeny & Whitney, 2014) and gender (Haberman & Whitney, 2007). Interestingly, summary feature extraction does not seem to rely exclusively on foveal input, as adequate extraction of expression can occur outside central fixation (Wolfe, Kosovicheva, Leib, Wood, & Whitney, 2015; Ji, Rossi, & Pourtois, 2018).

Yet despite a growing understanding of the types of summary attributes which can be extracted from face ensembles, a good understanding of the precise mechanisms which underlie these representations is still missing. In particular, investigating the extent of independence versus interaction in the processing of these attributes may help to elucidate the perceptual mechanisms mediating ensemble face processing. Here, we outline three behavioral studies in the ensemble literature which speak to this issue.

First, Haberman, Brady, and Alvarez (2015) provided significant evidence pointing to independent ensemble processing of higher level face identity and lower level object shape. However, these results also raised the question of whether such independent ensemble mechanisms are associated with different stimulus domains (i.e., faces vs. objects) or different levels of the visual hierarchy (i.e., low vs. high). Resolving this question is important, as it will inform behavioral, neural, and computational models of ensemble processing. Accordingly, here we investigate whether such independent mechanisms operate with stimuli from the same visual domain, specifically higher level identity and lower level viewpoint processing of face ensembles. Second, Cant, Sun, and Xu (2015) demonstrated distinct cognitive mechanisms involved in processing single objects and object ensembles. Using a Garner interference task (Garner, 1974), they found that participants could ignore changes in an unattended property when classifying single objects (e.g., changes in shape when reporting texture) but could not ignore changes in an unattended property when classifying object ensembles. Here we seek to evaluate the extension of these findings to face ensembles. Third, Leib et al. (2014) reported viewpoint-invariant average-identity processing where the average viewpoint is close to frontal. However, a more conclusive claim for independence would need to consider such invariance across a wider range of average viewpoints, deviating significantly from a frontal orientation.

Accordingly, across three experiments we systematically investigated viewpoint-invariant identity processing and identity-invariant viewpoint processing. Additionally, we examined potential dissociations between the processing of single and ensemble facial attributes. Our results argue for the independence of identity and viewpoint processing and for a dissociation between single-face and ensemble-face processing.

General methods

Participants

Participants from the University of Toronto community volunteered in exchange for monetary compensation or course credit. All procedures were carried

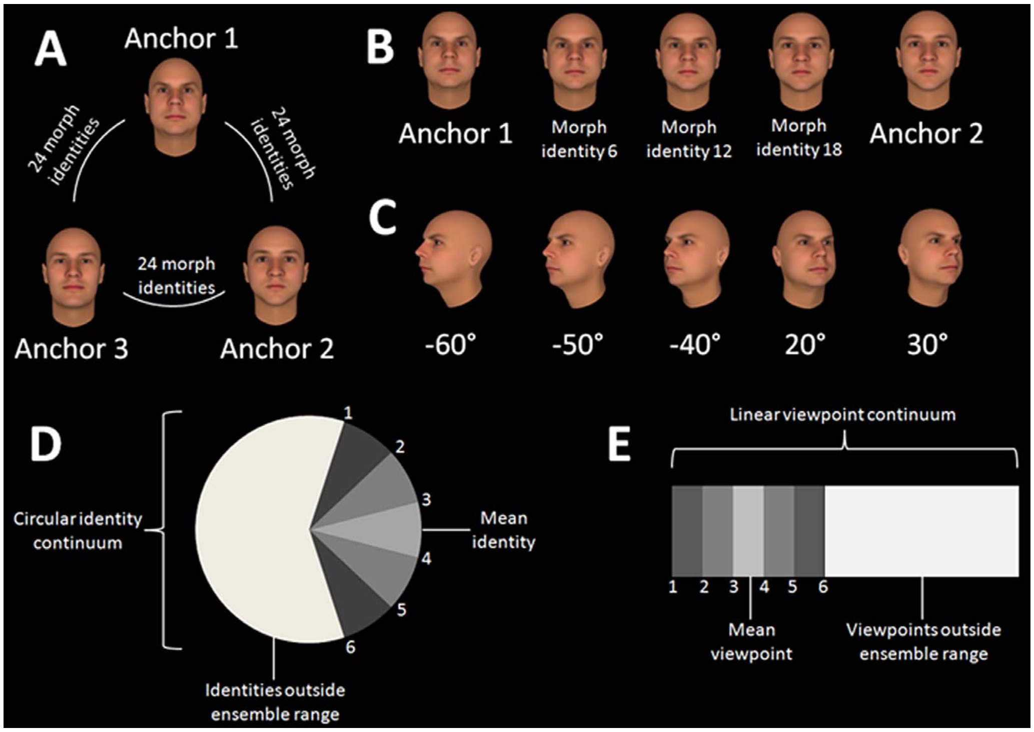

Figure 1. Stimulus generation and selection. (A) Circular arrangement of 75 facial identity morphs relative to three anchor identities. (B) Examples of identity morphs between Anchors 1 and 2. (C) Examples of different viewpoints for Anchor 1. (D–E) Schematic illustration of identity and viewpoint selection for the design of ensemble stimuli: 1–6 correspond to single face stimuli which collectively can be summarized by a mean identity or viewpoint (not shown).

out in accordance with the Declaration of Helsinki and were approved by the University of Toronto Research Ethics Board, including obtaining informed consent from all participants. All participants scored above 60% on the Cambridge Face Memory Test (Duchaine & Nakayama, 2006), ensuring that their face-recognition abilities fall within the norms for healthy young adults (Bowles et al., 2009).

A total of 87 participants were recruited across three experiments: 23 (12 women, 11 men; age range: 18–23 years) in Experiment 1, 43 (21 women, 22 men; age range: 18–29 years) in Experiment 2, and 21 (16 women, five men; age range: 18–21) in Experiment 3.

Stimuli and apparatus

Color face stimuli were generated using FaceGen Modeller Pro 3.5 (Singular Inversions Inc., Toronto, Canada; http://facegen.com/modeller.htm). We selected three randomly generated 3-D face meshes which varied considerably in visual appearance to act as anchors. Then we linearly interpolated between anchor pairs to obtain a face-shape continuum. This procedure yielded a total of 24 equally spaced identities between each pairing of the three anchors (Figure 1A, 1B) spanning a circular continuum of 75 arbitrary identity units (IUs). Each identity was rendered across different viewpoints in steps of 2°, for a total of 91 viewpoints ranging from −90° to +90° (Figure 1C), resulting in a total of 6,825 unique face images. Of note: By generating faces through 3-D mesh interpolation, we aimed to avoid perceptual inhomogeneities arising from traditional image-morphing techniques (for a discussion of this issue, see ZeeAbrahamsen & Haberman, 2018).

Specifically, morphing was achieved by weighing the 3-D mesh between anchor faces. This maintained a sharp image for each identity morph between anchors compared with other face-generation techniques (e.g., averaging pixel intensities). Facial texture (e.g., luminance, hue) was kept constant across different identities. Thus, identity information was provided only by shape, in agreement with its importance for recognition of unfamiliar faces (F. Jiang, Blanz, & Rossion, 2011; Lai, Oruc, & Barton, 2013). We also note that the granularity of the steps, namely 75 IUs for identity and 2° for viewpoint, was designed so that there was a noticeable change between any two neighboring pairs on each continuum. To be more rigorous, additional psychophysical testing was conducted to ensure that any observed results were not confounded by differences in perceptual units across the identity and

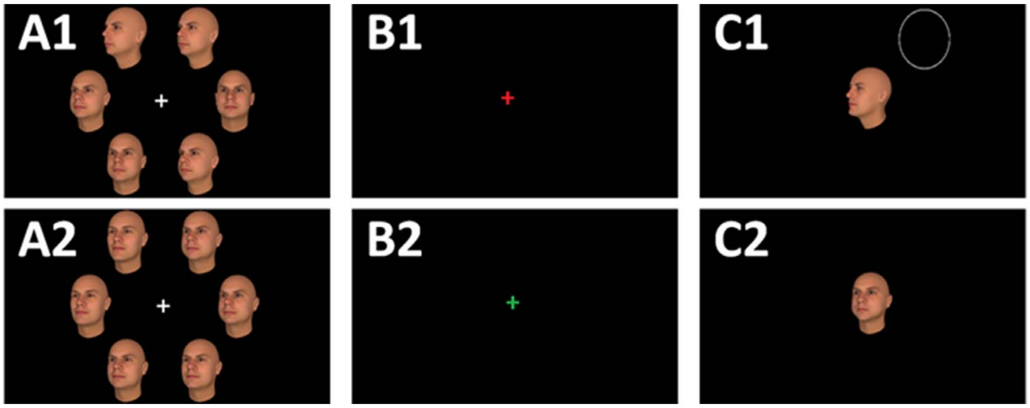

Figure 2. Examples of ensemble stimuli varying in (A1) viewpoint or (A2) identity along with the respective response cues (B1: red for report viewpoint; B2: green for report identity). Participants were probed to report either a single face at a randomly chosen location (C1; white oval) or the mean of the ensemble (C2; absence of an oval).

viewpoint continua (see Supplemental methods and Results in Supplementary File S1).

In designing ensemble stimuli, six single-face images were evenly spaced in a circular arrangement on screen, with three faces on the left of fixation and three on the right (Figure 2A). For an inhomogeneous identity ensemble, six facial identities were selected in steps of −15, −9, −3, 3, 9, and 15 IUs relative to a given mean identity (Figure 1D). Analogously, for an inhomogeneous viewpoint ensemble, six distinct viewpoints were selected in steps of −20°, −12°, −4°, 4°, 12°, and 20° from the chosen mean viewpoint (Figure 1E). A single face image subtended 2.9° × 4.2° of visual angle from a distance of 60 cm, while an entire face ensemble subtended 12.6° × 13.9°.

Importantly, in inhomogeneous displays the value of the ensemble mean for any attribute was not assigned to any single face of the ensemble. For example, if the mean identity of an ensemble consisting of different facial identities was Morph 35, then single-face identities would consist of Morphs 20, 26, 32, 38, 44, and 50. Similarly, if the mean viewpoint of an ensemble containing variable viewpoints was 10°, then single-face viewpoints would consist of −10°, −2°, 6°, 14°, 22°, and 30°. For homogeneous displays, where a given attribute does not vary, all faces in the display are consistent with the mean attribute.

Procedure

Prior to each experiment, participants completed the Cambridge Face Memory Test. They were then familiarized with the stimuli and with the keyboard controls. For five trials per attribute, two faces appeared on-screen and participants cycled through identity or viewpoint of the right face to match the appearance of the target face on the left. Participants then completed a series of practice trials which followed the design of the main experiment (see later) with the addition of feedback given at the end of each trial, where the correct stimulus was shown next to the image that participants selected as a response.

Each trial began with a white fixation cross appearing at the center of the screen for a variable 300–800-ms interval, followed by an ensemble consisting of six faces which varied in identity, viewpoint, or both, shown on the screen for 400 ms. During this interval the cross remained on-screen and participants were instructed to maintain fixation and not divert their eyes to any of the individual faces. Next, a response cue appeared in the form of a colored fixation cross (green for identity, red for viewpoint, see Figure 2B) for 500 ms followed by a single face probe (Figure 2C). The task of the participants was to navigate a continuum of face identities or viewpoints by using the up and down or left and right arrow keys, respectively. The initial value of the attribute of interest for that trial was chosen randomly from among off-range values (see Figure 1D, 1E). For the task-irrelevant attribute (e.g., viewpoint when only identity was reported), its value was locked to the ensemble mean or to the specific target face when participants reported ensemble mean or single target-face values, respectively. Participants were cued to report either the attribute value of a single face at a designated location, signaled by the presence of an oval, or the average value of the ensemble, signaled by the absence of an oval cue (Figure 2C). When finished, participants pressed the space bar to initiate the next trial.

For each response, accuracy was recorded as mean percent distance (MPD), which is the closest distance between the selected attribute value and the correct one as a percentage relative to the total number of units for that attribute (75 for identity, 91 for viewpoint). In addition, reaction time (RT) was recorded as the duration from probe onset to the first key response.

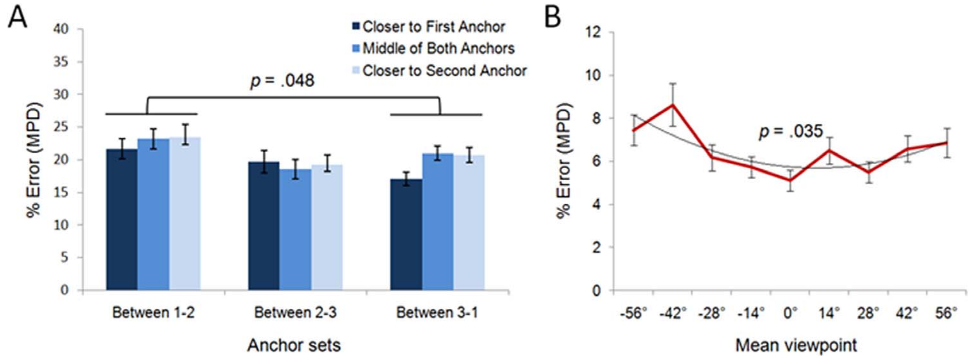

Figure 3. Error rates for mean identity and viewpoint along each continuum in mean percent distance. (A) Mean identity was reported more accurately for morphs between Anchors 3 and 1 than between Anchors 1 and 2 ($p = 0.048$). (B) Participants were also more accurate at reporting mean viewpoints closer to a frontal orientation ($p = 0.035$). Error bars indicate ± 1 standard error.

All stimuli were presented on a monitor with a resolution of 1,920 × 1,080 pixels and a 60-Hz refresh rate. The head of each participant was stabilized with the aid of a chin rest placed 60 cm away from the screen. Stimulus presentation as well as data collection and analysis relied on MATLAB (Version 2015a; The MathWorks, Natick, MA) using Psychtoolbox 3.0.13 (Brainard, 1997). When necessary, multiple comparisons where corrected using the Bonferroni procedure. Modified Greenhouse–Geisser p values and degrees of freedom are reported in cases where sphericity was violated. Effect sizes for significant effects are given via Cohen's d or partial η^2.

Experiment 1

Procedure

Experiment 1 investigated the relationship between average and single-face attribute processing for identity and viewpoint. Participants first completed 10 familiarization trials followed by 24 practice trials. Next, in the main experiment, they completed 216 trials equally divided across four conditions (average and single reports for identity and viewpoint) presented in random order. A 20-s break took place every 36 trials.

For identity conditions, nine mean identities were preselected, with three between each anchor pair. They were combined with three mean viewpoints, resulting in 27 face-ensemble stimuli. Similarly, for viewpoint conditions, nine mean viewpoints were preselected and then combined with three mean identities, giving an additional 27 ensemble stimuli for a total of 54 unique ensemble stimuli. Each unique ensemble stimulus was repeated twice in its respective condition. Testing took approximately 1 hr to complete for each participant in a single experimental session.

Results

Evaluation of identity and viewpoint continuum

To assess the perceptual homogeneity of each continuum, we evaluated performance of mean reports across different ranges separately for identity and viewpoint. For instance, if some viewpoints are more distinguishable than others, then performance along the viewpoint continuum is expected to vary systematically.

First, for identity, the nine preselected identity morphs were divided between the three anchor pairs and labeled as being closer to one anchor or equidistant to both. We then analyzed MPD error using a 3 (anchor pair: 1–2, 2–3, or 3–1) × 3 (anchor distance: closer to first anchor, equidistant to both, or closer to second anchor) within-subject analysis of variance (ANOVA). The results (Figure 3A) showed a significant main effect of anchor pair, $F(2, 44) = 4.36$, $p = 0.019$, $\eta^2 = 0.165$) but not of anchor distance, $F(2, 44) = 2.09$, $p = 0.136$, and no interaction, $F(3, 66.03) = 0.92$, $p = 0.435$. Post hoc tests found that reports of mean identity for anchors 3–1 were more accurate than reports for anchors 1–2, mean difference = 3.2% MPD, $t(22) = 2.67$, $p = 0.048$, $d = 0.56$. No other comparisons reached significance (mean difference of both = 2% MPD, both ts < 2.57, both ps > 0.056). These results likely reflect small but systematic perceptual differences along the identity continuum.

Second, mean viewpoints were organized in ascending order (–56°, –42°, –28°, –14°, 0°, 14°, 28°, 42°, and 56°) and analyzed with a one-way within-subject ANOVA. Results (Figure 3B) showed a significant effect of viewpoint, $F(4.49, 98.88) = 3.32$, $p = 0.011$, $\eta^2 = 0.131$. Curve fitting revealed a significant quadratic fit, $F(1, 22) = 5.02$, $p = 0.035$, $\eta^2 = 0.186$, indicating that participants were more accurate at reporting mean viewpoints that were closer to a frontal orientation.

For identity, some perceptual inhomogeneities may exist between faces from different anchor combina-

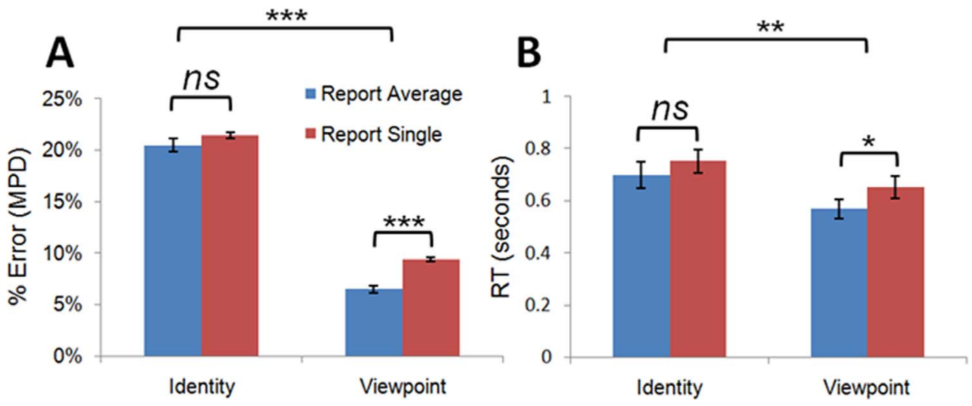

Figure 4. Results for Experiment 1, quantified by (A) error and (B) response time for average and single reports of identity and viewpoint. (A) Viewpoint was reported more accurately than identity, and average-viewpoint attributes were reported more accurately than single-viewpoint attributes. (B) Viewpoints were reported faster than identity, and average-viewpoint reports were reported faster than single-viewpoint reports. Error bars indicate ±1 standard error; *ns* = not significant. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

tions. However, since all ensemble means are repeated for each identity condition, these differences would not affect our main results. For mean viewpoint reports, the sensitivity to changes in global viewpoint likely reflects cognitive-processing differences (e.g., better judgment of orientation for frontal viewpoints).

On a related note, we examined the possibility that different members of the identity or viewpoint continua would bias judgments. Specifically, while an MPD metric provides a convenient way of assessing performance and facilitates comparison with prior findings in the field (Haberman et al., 2015; Wolfe et al., 2015), we note that it is sensitive to both bias and precision. Hence, we assessed bias (i.e., offset with respect to relevant reference stimuli) as well as precision (i.e., standard deviation of responses) separately for each condition (see Supplemental methods and Results in Supplementary File S1). This analysis found significant effects for both bias and precision, though their contribution was different for identity versus viewpoint judgments. Concretely, bias affected viewpoint more clearly, while precision was poorer for identity relative to viewpoint.

Mean and single-item processing of identity and viewpoint

Participant performance across the four conditions was analyzed using a 2 (reported attribute: identity, viewpoint) × 2 (report: average, single face) within-subject ANOVA. The results (Figure 4A) showed significant main effects of attribute, $F(1, 22) = 1{,}282.02$, $p < 0.001$, $\eta^2 = 0.983$, and report, $F(1, 22) = 20.69$, $p < 0.001$, $\eta^2 = 0.485$, as well as a significant interaction, $F(1, 22) = 6.36$, $p = 0.019$, $\eta^2 = 0.224$. Further analyses found a significant difference between single and average viewpoint, mean difference = 2.9% MPD, $t(22) = 9.67$, $p < 0.001$, $d = 2.02$, but not between single and average identity, mean difference = 1.0% MPD, $t(22) = 1.43$, $p = 0.190$. These results were replicated using an adjusted scale equating perceptual differences across identity and viewpoint stimulus steps (see Supplementary Figures S1 and S2).

An ANOVA of RT (Figure 4B) was consistent with the accuracy results—namely, the main effects of attribute, $F(1, 22) = 11.21$, $p = 0.003$, $\eta^2 = 0.338$, and task, $F(1, 22) = 7.46$, $p = 0.012$, $\eta^2 = 0.253$, were significant, but the interaction was not, $F(1, 22) = 0.86$, $p = 0.363$. Post hoc tests revealed a significant mean difference between reports of average and single viewpoint—85 ms, $t(22) = 2.74$, $p = 0.012$, $d = 0.57$—but only a marginally significant mean difference between average and single identity: 53 ms, $t(22) = 1.76$, $p = 0.095$.

To assess potential independence in the cognitive mechanisms mediating identity and viewpoint extraction for both face ensembles and single faces, we performed six Pearson correlations comparing participant performance in each of the four conditions to one another. No correlations among the accuracy data were significant (all ps > 0.162 prior to Bonferroni correction). To ensure that these null results were not due to a lack of statistical power, we conducted a split-half analysis by correlating performance on odd and even trials for each of our four conditions separately (i.e., within an attribute). We observed significant correlations for average identity ($r = 0.55$, Bonferroni-corrected $p = 0.027$) and average viewpoint ($r = 0.744$, corrected $p < 0.001$), but not for single identity or viewpoint (both rs < 0.51, both corrected ps > 0.053). Thus, the present results suggest that at least as far as reports of average attributes are concerned, the accuracy estimates across the two attributes are independent.

Interestingly, all correlations across the same four conditions for RT were significant (all *r*s between 0.626 and 0.801, all Bonferroni-corrected *p*s < 0.019), save for the correlation between average identity and single viewpoint ($r = 0.49$, Bonferroni-corrected $p = 0.102$). Consistent with the analysis of accuracy, though, the largest correlations were within a facial attribute (average and single identity: $r = 0.801$; average and single viewpoint: $r = 0.709$) rather than across facial attributes (average identity and average viewpoint: $r = 0.629$; single identity and single viewpoint: $r = 0.626$). The smallest correlations were found when both attribute and report varied (average identity and single viewpoint: $r = 0.492$; single identity and average viewpoint: $r = 0.587$).

Ensemble mean bias when reporting single attributes

While the trend for an advantage of average reports over single reports of identity did not reach significance (Figure 4A), it is possible that mean identity biased the extraction of single identity, consistent with the influence of summary statistics on the perception and memory of single items (de Fockert & Wolfenstein, 2009; Brady & Alvarez, 2011).

To assess this possibility, MPD estimates for single reports were referenced to the mean value of an ensemble in addition to the correct value of the target single item. The data were then analyzed with a 2 (attribute: identity, viewpoint) × 2 (reference: ensemble mean value, target item value) within-subject ANOVA. The results (Figure 5) revealed significant main effects of attribute, $F(1, 22) = 664.51$, $p < 0.001$, $\eta^2 = 0.968$, and reference, $F(1, 22) = 20.00$, $p < 0.001$, $\eta^2 = 0.476$, as well as a significant interaction, $F(1, 22) = 4.88$, $p = 0.038$, $\eta^2 = 0.182$. Further tests revealed that single reports of both identity, mean difference = 1.4% MPD, $t(22) = 4.67$, $p < 0.001$, $d = 0.97$, and viewpoint, mean difference = 0.5% MPD, $t(22) = 2.50$, $p = 0.039$, $d = 0.52$, were closer to the average value of the ensemble than to the correct value of the single target.

Discussion

The present experiment finds that participants reported viewpoint more accurately than identity even when the perceptual difficulty of discriminating across units of different attributes was equated. This is not unexpected, given that the processing of face identity presumably requires a higher level of abstraction than does viewpoint (Grill-Spector & Kanwisher, 2005; but see Or & Wilson, 2010).

More importantly, the average/single dissociation typically reported in the ensemble literature (Alvarez, 2011; Whitney & Leib, 2018) was significant only for viewpoint. This seems to point to different perceptual mechanisms for identity versus viewpoint processing (Haberman et al., 2015). Regarding the absence of a clear dissociation for identity, this is convergent with previous findings (Neumann, Ng, Rhodes, & Palermo, 2017) suggesting that processing a summary identity does not preclude or negate the extraction of individual identity. However, we did find that reports of single facial attributes were closer to the mean attribute than to the single-face targets which participants were instructed to report. This finding is representative of a mean bias in ensemble processing (de Fockert & Wolfenstein, 2009; Brady & Alvarez, 2011), and conforms to the general theory of a global bias in the visual system (Hochstein, Pavlovskaya, Bonneh, & Soroker, 2015).

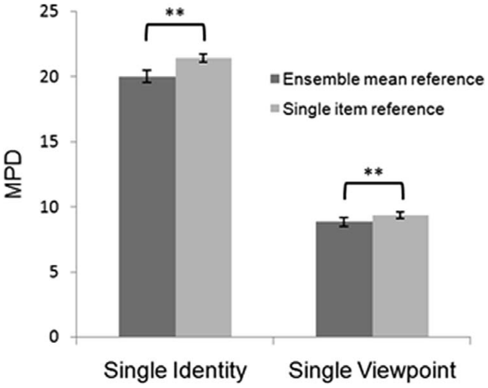

Figure 5. Distance (mean percent distance) of single reported values relative to the mean ensemble value and to the target single-item value. Reports are closer to the ensemble mean attribute for both identity and viewpoint. Error bars indicate ±1 standard error. $^{**}p < 0.01$.

Further, we found no significant correlations of accuracy across attributes, and in the case of RT we found larger correlation values within rather than across attributes, consistent with the idea of independent mechanisms for viewpoint and identity processing.

Finally, participants were more accurate at reporting mean viewpoints that were closer to a frontal orientation, presumably by virtue of the privileged neural processing of frontal orientations (Ramírez, Chichy, Allefeld, & Haynes, 2014).

Experiment 2

Here we investigated whether identity reports are affected by changes in mean viewpoint. If mean identity reports show viewpoint sensitivity, this would be evidence of an interaction between the processing of identity and viewpoint in face ensembles, reflecting shared underlying cognitive mechanisms. If, however,

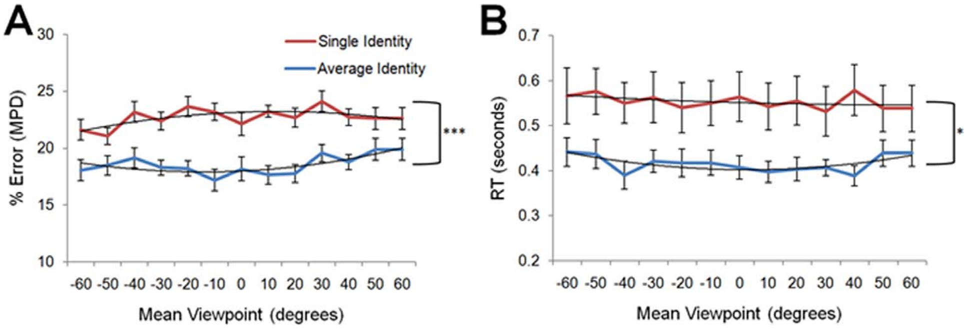

Figure 6. (A) Accuracy (mean percent distance) and (B) response time of identity reports for mean- (blue line) or single- (red line) identity reports across varying mean viewpoints in Experiment 2. We find no effect of varying mean viewpoint on the accuracy of either mean- or single-identity extraction, and only a marginally significant effect of viewpoint on response time ($p = 0.056$) during the reporting of mean identity. Participants were (A) more accurate and (B) faster when reporting an average identity. Error bars indicate ±1 standard error. *$p < 0.05$; ***$p < 0.001$.

identity reports are not affected by changes in mean viewpoint, this would be consistent with independent cognitive mechanisms mediating the processing of these facial attributes. Given the results of Experiment 1, we hypothesized that viewpoint changes would not interact with the processing of identity.

Procedure

In this experiment, participants reported either an average identity ($n = 20$) or a single identity ($n = 23$). The identity and viewpoint of single faces in the ensemble varied on each trial, rendering them distinct from their respective mean values—recall that in Experiment 1 the attribute that was not reported on a trial was homogeneous. As in Experiment 1, the mean feature was never visually presented to participants. There were 13 mean viewpoints spanning –60° to 60° in steps of 10°. Viewpoints were combined with 15 mean identities (one every 5 IUs), giving a total of 195 unique stimuli. Participants reporting single identity were cued in the same way as in Experiment 1.

Prior to the main experiment, five familiarization trials were completed followed by 24 practice trials. In all other respects the procedure followed that of Experiment 1.

Results

Accuracy (MPD) and RT were each analyzed by a 2 (between-subjects factor: report mean identity, report single identity) × 13 (within-subject factor: viewpoint degree) mixed-design ANOVA. Accuracy results are plotted in Figure 6A. We found no significant effect of varying viewpoint on identity extraction, $F(12, 492) = 1.18$, $p = 0.294$. Quadratic fitting also showed no significant curve for reporting either average, $F(1, 19) = 2.28$, $p = 0.147$, or single identities, $F(1, 22) = 1.52$, $p = 0.231$. Of note here, participants who reported average identity were more accurate than those who reported single identity, $F(1, 41) = 67.65$, $p < 0.001$, $\eta^2 = 0.623$. The Report × Viewpoint degree interaction was not significant, $F(12, 492) = 0.94$, $p = 0.510$.

We report similar findings for RT (Figure 6B). Specifically, there was no significant change in RT across viewpoints, $F(12, 492) = 0.87$, $p = 0.583$, there was a significant difference in processing speed between average- and single-identity reports, $F(1, 41) = 5.59$, $p = 0.023$, $\eta^2 = 0.120$, and there was no interaction, $F(12, 492) = 0.96$, $p = 0.484$. For mean identity (Figure 6B, blue line), there was a marginally significant quadratic fit, $F(1, 19) = 4.15$, $p = 0.056$, but this was not found for single-identity reports (Figure 6B, red line), $F(1, 22) = 0.12$, $p = 0.729$.

Finally, in Experiment 1 single-attribute reports were biased to their respective means. We also explored whether or not this was the case with single-identity reports in Experiment 2. Using a 2 (single face, average face) × 13 (mean viewpoint value) within-subject ANOVA to analyze the data, we observed no significant results—main effect of reference: $F(1, 22) = 1.12$, $p = 0.302$; main effect of mean viewpoint: $F(12, 264) = 0.85$, $p = 0.599$; interaction: $F(7.43, 163.45) = 1.34$, $p = 0.230$. This indicates that, unlike the results from Experiment 1, reports of single identity were not closer to the mean identity of the ensemble compared with the correct value of the to-be-reported single identity.

Discussion

In Experiment 2, participants reported mean identity more accurately and more quickly than single identity. At the same time, we note that the mean bias for single-identity reports, as found in Experiment 1, did not occur here; single-identity reports were not closer to the mean face identity than to the single face target. This suggests a trade-off between the two effects, possibly driven by differences in experimental design—for instance, Experiment 1 required participants to report both the average and a single face, whereas Experiment 2 required only one of the two, hence reducing the complexity of the task. Thus, while the influence of the mean is apparent in both experiments, its impact with respect to different measures may vary depending on aspects of the experimental design that remain to be further explored and clarified.

More importantly, ensemble identity reports were not affected by changes in mean viewpoint, consistent with the idea of independence (Leib et al., 2014), despite the use of a wide range of mean viewpoints. Somewhat surprisingly, though, we did not find viewpoint sensitivity for single faces either, as documented by previous work (Bruce & Valentine, 1987; Logie et al., 1987; Pourtois et al., 2005; but see Liu & Chaudhuri, 2002).

Experiment 3

In order to better evaluate the potential independence of face identity and viewpoint processing in ensemble perception, we introduced two important changes in Experiment 3, with the goal of maximizing the possibility of an interaction between these two facial attributes. First, we examined the influence of holding both summary attributes in VWM while being tasked to report only one of those attributes at a time. Second, we examined the influence of variation in an irrelevant attribute when participants attended to and reported a relevant summary attribute, similar to a Garner interference task. If summary viewpoint and identity are processed by independent cognitive mechanisms, then we would expect to see little to no interference in the two scenarios. We did not examine the processing of identity and viewpoint for single faces in this experiment, given our focus on the potential independence of ensemble identity and viewpoint.

Procedure

In each trial, participants reported either one or both ensemble face attributes. There were three ways individual faces varied from their respective mean attributes of the set on a given trial: Only individual identity varied (IV), only individual viewpoint varied (VV), or both attributes varied (BV). We used six mean identities (specifically, Morphs 7, 19, 32, 44, 57, and 69, with two IUs between each of the three anchor faces) and six mean viewpoints (specifically, −50°, −30°, −10°,10°, 30°, and 50°), yielding 36 unique combinations which were balanced across the different experimental conditions. In total, there were 324 trials across the three report conditions (report only viewpoint, report only identity, report both) and three attribute-variation conditions (IV, VV, BV).

All display parameters were consistent with previous experiments. In addition to a green fixation cross indicating identity reports and a red fixation cross indicating viewpoint reports during the task-indication phase, a blue fixation cross was used to prompt participants to report both attributes. To control for possible order effects in this condition, on half of the trials participants reported identity first (signaled by $\wedge$ and $\vee$ flanking the blue fixation cross vertically) and viewpoint second, and on the other half they reported viewpoint first (signaled by $<$ and $>$ flanking the blue fixation cross horizontally). The ability to cycle through variations in the second attribute (e.g., viewpoint in a report-identity-first trial) was locked until participants made their response to the first attribute, which was done by pressing the space bar. This then unlocked the second attribute and allowed participants to make their second response. Pressing the space bar a second time ended the current trial, and initiated the next trial in the sequence.

Before the main experiment, participants were given 15 familiarization trials (five each for reporting identity, viewpoint, and both) and 27 practice trials (nine each for reporting identity, viewpoint, and both). In the main experiment, participants were given 20-s breaks after every 54 trials.

Results

Order effects on summary attribute reports

When participants reported both attributes in a sequence, it is possible that the second attribute was reported less accurately due to VWM capacity limitations. To assess this possibility we analyzed ensemble identity and viewpoint from trials in which participants reported both attributes. Accuracy was analyzed with a 2 (reported attribute: identity, viewpoint) × 2 (order: identity first, viewpoint first) × 3 (varying attribute: IV, VV, BV) within-subject ANOVA. Results are plotted in Figure 7A and 7B.

There was a significant effect of reported attribute, $F(1, 20) = 858.12$, $p < 0.001$, $\eta^2 = 0.977$, indicating that participants were more accurate at reporting mean viewpoint compared with mean identity, consistent with the results of Experiment 1. There was also a significant main effect of order, $F(1, 20) = 6.54$, $p = 0.019$, $\eta^2 = 0.246$; but the main effect of varying attribute was not significant, $F(1.65, 33.10) = 1.35$, $p =$

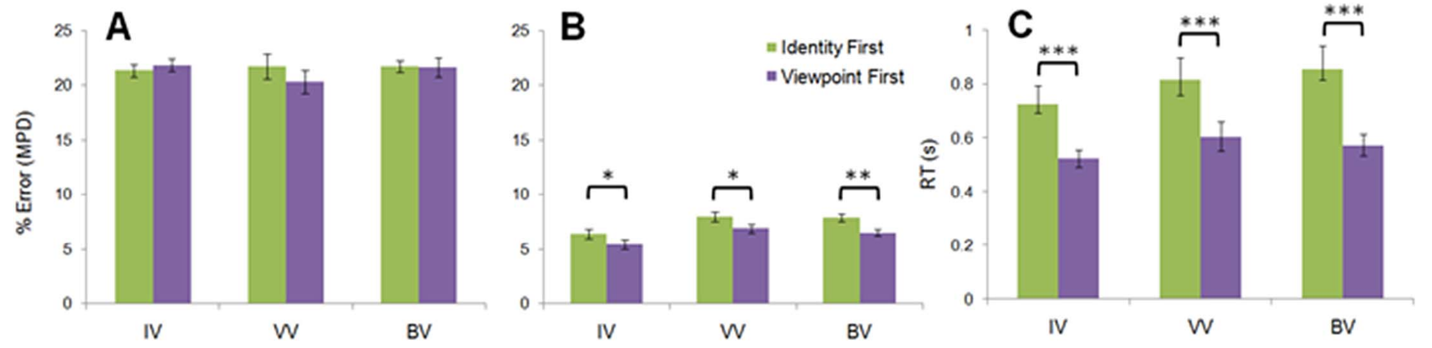

Figure 7. Percent error (mean percent distance) of (A) identity and (B) viewpoint reports, depending on whether they were reported first or second during trials where both attributes were reported. (C) Response-time estimates are shown as a function of attribute order (identity first or viewpoint first). For accuracy (mean percent distance), there was no difference when identity was reported either first or second, but viewpoint had decreased accuracy when it was reported second. Consistently faster response times were also noted when the first attribute to be reported was viewpoint. IV = only identity varies; VV = only viewpoint varies; BV = both vary. Error bars indicate ±1 standard error. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

0.270. No interactions reached significance (all Fs < 3.04, all ps > 0.075).

Further pairwise testing revealed that the significant effect of order was restricted to instances in which viewpoint was reported. That is, when participants reported identity, there was no difference in accuracy regardless of whether identity was reported first or second, mean difference = 0.4% MPD, $t(20) = 0.63$, $p > 0.500$ (Figure 7A), but participants were less accurate at reporting viewpoint when it was reported after identity, mean difference = 1.1% MPD, $t(20) = 5.5$, $p < 0.001$, $d = 1.20$ (Figure 7B).

Next, we analyzed RT with a 2 (order: identity first, viewpoint first) × 3 (varying attribute: IV, VV, BV) within-subject ANOVA. RT was analyzed differently than MPD owing to the fact that on each trial, RT was recorded only once, at the first response. Results are plotted in Figure 7C. We report effects of varying attribute, $F(1.64, 32.81) = 3.569$, $p = 0.048$, $\eta^2 = 0.151$, and order, $F(1, 20) = 22.55$, $p < 0.001$, $\eta^2 = 0.530$. There was no Order × Varying attribute interaction, $F(2, 40) = 1.91$, $p = 0.161$. For varying attribute, post hoc testing revealed a difference between IV and BV conditions, $t(20) = 2.97$, $p = 0.021$, $d = 0.65$, but not between VV and either IV or BV (both ts < 2.35, both ps > 0.094). Further pairwise comparisons for report order showed faster RTs when viewpoint was reported first across all three attribute-varying conditions (mean difference across all three conditions = 231 ms, all ts > 3.96, all ps < 0.001, all ds > 0.86).

Further, we found that the MPD spread in participant reports was greater when viewpoint was reported after identity, mean difference = 0.80% SD_{MPD}, $t(20) = 3.85$, $p = 0.001$, $d = 0.87$, but not in the reverse order, mean difference = 0.30% SD_{MPD}, $t(20) = 0.99$, $p = 0.344$. This larger spread is consistent with increased error due to guessing in the former case (Zhang & Luck, 2009; see discussion later).

Independence of identity and viewpoint

Accuracy of summary reports for all attributes was analyzed with a 2 (number reported: one attribute, both) × 2 (reported attribute: identity, viewpoint) × 3 (varying attribute: IV, VV, BV) within-subject ANOVA and is plotted in Figure 8. Participants were more accurate when reporting a single summary attribute compared with both attributes, $F(1, 20) = 10.12$, $p = 0.005$, $\eta^2 = 0.336$, as well as when reporting viewpoint compared with identity, $F(1, 20) = 987.27$, $p < 0.001$, $\eta^2 = 0.980$. There was also a marginally significant main effect of varying attribute, $F(1.41, 28.25) = 3.21$, $p = 0.071$. The only significant interaction observed was between reported attribute and varying attribute, $F(1.33, 26.62) = 12.50$, $p = 0.001$, $\eta^2 = 0.385$; for all other interactions, all Fs < 2.66, all ps > 0.082.

The question of why reporting both attributes had a larger MPD than reporting one attribute may be answered by the findings of the previous section. Specifically, this was due to an increase in MPD for viewpoint when reported after identity. Here we conducted post hoc tests examining how changes in an irrelevant attribute affected reports of the relevant attribute. When participants reported only identity, we found a significant difference in accuracy between IV and VV conditions, mean difference = 2.5% MPD, $t(20) = 2.69$, $p = 0.014$, $d = 0.68$, a marginally significant difference between VV and BV conditions, mean difference = 2.7% MPD, $t(20) = 2.10$, $p = 0.052$, and no difference between IV and BV conditions, mean difference = 0.2% MPD, $t(20) = 0.33$, $p > 0.500$. Accuracy was highest when there was no variation in the relevant attribute. Variation in viewpoint, an irrelevant attribute, did not affect reports of mean identity.

We observed similar findings for reports of mean viewpoint. There were significant differences in accu-

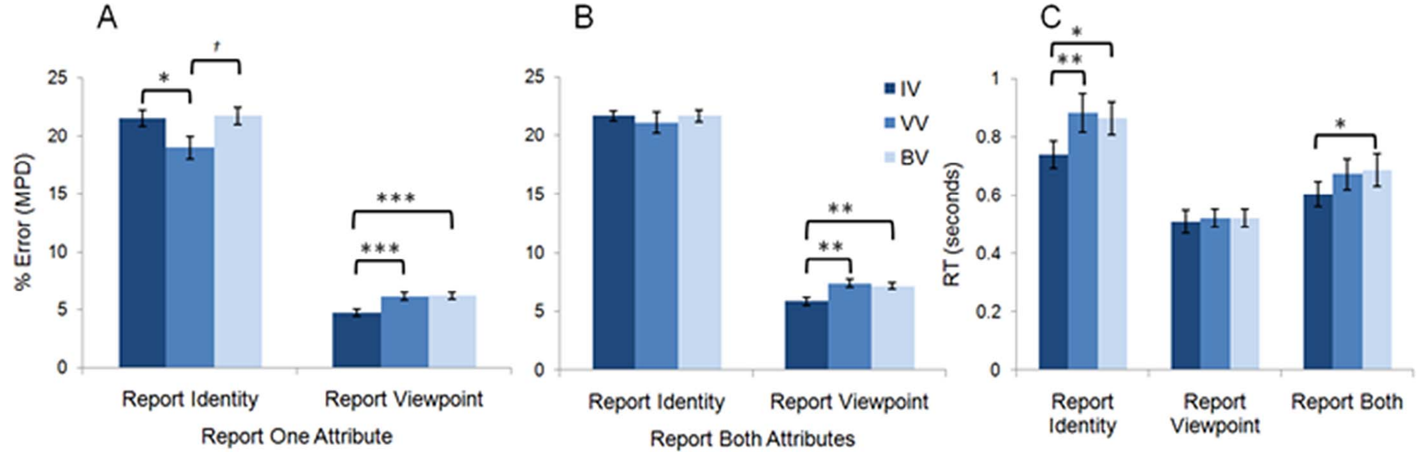

Figure 8. (A–B) Mean percent distance and (C) response time for reports of identity and viewpoint across conditions of varying attributes. Participants reported either (A) one or (B) both attributes on a given trial. For mean percent distance we find that accuracy improves when relevant attributes do not vary, regardless of variation in irrelevant attributes. For response time, we find that viewpoint reports are made faster than identity reports, and that when identity is reported, there is a facilitation effect when identity varies. IV = only identity varies; VV = only viewpoint varies; BV = both vary. Error bars indicate ±1 standard error. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; $^{+}p < 0.10$.

racy between VV and IV conditions, mean difference = 1.4% MPD, $t(20) = 3.85$, $p < 0.001$, $d = 1.02$, and between IV and BV conditions, mean difference = 1.4% MPD, $t(20) = 3.85$, $p < 0.001$, $d = 1.02$. There was no difference between VV and BV conditions, mean difference = 0.0%, $t(20) = 0.13$, $p > 0.500$. Consistent with the results for mean identity, variations in irrelevant identity did not affect mean-viewpoint report accuracy.

When participants reported both attributes, we noted a similar result, but only in reporting of mean viewpoint. There were significant differences in accuracy between IV and VV conditions, mean difference = 1.5% MPD, $t(20) = 3.55$, $p = 0.002$, $d = 0.82$, and between IV and BV conditions, mean difference = 1.3% MPD, $t(20) = 2.89$, $p = 0.009$, $d = 0.71$, but not between VV and BV conditions, $t(20) = 0.67$, $p > 0.500$. For reports of mean identity, no conditions differed (all ts < 0.75, all ps > 0.500).

RT was analyzed differently than accuracy, owing to the fact that when both conditions are reported, RT is recorded only once (i.e., the time when the first response is initiated). Here we used a 3 (reported attribute: identity, viewpoint, or both) × 3 (varying attribute: IV, VV, BV) within-subject ANOVA, and the results are plotted in Figure 8C. We report significant main effects of reported attribute, $F(2, 40) = 42.18$, $p < 0.001$, $\eta^2 = 0.678$, and varying attribute, $F(1.66, 33.12) = 7.27$, $p = 0.004$, $\eta^2 = 0.267$, as well as a marginally significant interaction, $F(4, 80) = 2.27$, $p = 0.069$. Post hoc testing revealed that all three reported-attribute conditions differed significantly from one another—identity versus viewpoint: mean difference = 0.31 s, $t(20) = 8.00$, $p < 0.001$, $d = 1.74$; identity versus both: mean difference = 0.17 s, $t(20) = 5.87$, $p < 0.001$, $d = 1.27$; viewpoint versus both: mean difference = –0.14 s, $t(20) = 4.19$, $p < 0.001$, $d = 0.90$.

Additional pairwise comparisons investigating the significant main effect of varying attribute revealed that, when only identity was reported, there was a significant difference in RT between IV and VV conditions, mean difference = –0.14 s, $t(20) = 4.21$, $p = 0.002$, $d = 0.90$, and IV and BV conditions, mean difference = –0.12 s, $t(20) = 2.95$, $p = 0.024$, $d = 0.64$; additionally, when both features were reported, there was a significant difference between IV and BV conditions, mean difference = –0.08 s, $t(20) = 2.75$, $p = 0.037$, $d = 0.60$.

Effect of position in space

The larger number of trials in Experiment 3 allowed us to assess how the position of single faces in the ensemble influenced average-feature reports. We obtained, for every trial, the difference in accuracy (MPD) between the participant's reported mean attribute and the value for each of the six individual face positions in the ensemble, and compared each of these difference scores to the MPD derived from the difference between the participant's report of the mean attribute and the true mean value. This analysis tells us how close the reported face attribute was to each individual face on the screen relative to the unseen mean facial attribute, and thus has the potential to reveal which individual face positions were more or less likely to be integrated into the mean percept (i.e., the closer in MPD a probe is to a specific location on-screen, the more strongly that

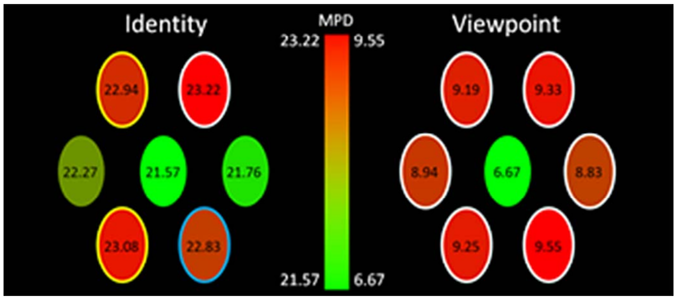

Figure 9. Each oval represents a face position in the ensemble display, with the unseen mean face in the center. The relevant value of mean percent distance (MPD), coded as an MPD difference score between the participant's report and each single face, is displayed within each oval. The color gradient is weighted relative to the smallest (green colors) and largest (red colors) MPD for that attribute (left = identity, right = viewpoint). Multiple pairwise tests were conducted comparing MPD to each face with MPD to the mean. We observe that the identities directly to the left and right are strongly incorporated into the participant's mean-identity report; viewpoint seems to be more evenly distributed. Oval frames indicate the significance of distance from the mean: yellow = $p < 0.01$; white = $p < 0.001$; blue = $p < 0.10$; no frame = not significant.

specific face location was likely incorporated into the participant's mean report).

Data for identity positions and viewpoint positions were each analyzed using one-way within-subject ANOVAs and are displayed in Figure 9. For identity, there was a significant effect of MPD, $F(3.59, 71.84) = 5.46$, $p = 0.001$, $\eta^2 = 0.214$. MPD for the left and right face did not differ from that for the mean (average difference of both faces = 0.45% MPD, both $ps > 0.500$). MPD for the top left, top right, and bottom left were all significantly different from the mean face (average difference of all three faces = 1.51% MPD, all $ps \leq 0.001$, all $ds > 1.00$). The bottom right face was marginally significantly different from the mean face, difference = 1.26% MPD, $t(20) = 3.19$, $p = 0.097$. For viewpoint, there was also a significant difference in MPD, $F(4.12, 82.38) = 47.42$, $p < 0.001$, $\eta^2 = 0.703$. Post hoc pairwise analyses demonstrated that MPD for the mean viewpoint was significantly different from those for all six face positions (average difference from all six faces = 2.51% MPD, all $ps < 0.001$, all $ds > 2.36$).

Discussion

The current experiment reports three main findings. First, we found that variations in a response-irrelevant attribute do not affect reports of a relevant attribute—for example, reports of mean identity were generally not affected by whether viewpoint varied or not.

Second, when participants reported both attributes, accuracy for mean identity was not affected by whether it was reported before or after viewpoint, but mean viewpoint was reported less accurately, and with increased spread, if participants first responded to mean identity. While this may be construed as interference of identity on viewpoint, it may instead be explained by the decay or sudden loss of mean viewpoint representations in VWM due to the increased amount of time it takes to report mean identity. Indeed, such loss of information in VWM has been reported by Zhang and Luck (2009), who found increased guess rates for reporting color and shape features after 4 s, which is consistent with the 3.56 s average time it took participants to report mean identity in our experiment. It is also possible that the capacity limitation of VWM may not apply to processing face identity (Curby & Gauthier, 2007; Wong, Peterson, & Thompson, 2008), unlike lower level visual features such as orientation (Y. V. Jiang, Shim, & Makovski, 2008). This larger capacity advantage in short-term memory is also found for other categories of acquired expertise, such as with car experts (Curby, Glazek, & Gauthier, 2009). Of course, these possibilities are speculative, and it is possible that something other than VWM load may account for these findings. It is the goal of future research to disentangle possible explanations for the decrease in viewpoint reports following identity reports, when both ensemble facial attributes are held in VWM.

Third, the faces to the left and right of fixation were more likely to be integrated into the representation of ensemble identity, but the evidence for this with viewpoint processing was less clear. In a review of ensemble processing, Whitney and Leib (2018) point out that ensemble encoding mechanisms do not integrate all items in a display when forming a summary statistic, but instead integrate the square root of the set size (Dakin, 2001). Here we find that two faces, those to the left and the right of fixation, contributed the most to the encoding of mean identity. While it seemed that all faces were equally factored for computing average viewpoint, the slightly lower MPD values to the left and right of fixation are broadly consistent with the results for mean identity. To be clear, it is not necessarily the case that discrete individual items are incorporated and the rest dismissed; more likely, different weights are assigned to certain regions of the visual field depending on the nature of the stimulus domain and of the attribute over which summary statistics are computed.

General discussion

The goal of the present study was to investigate the potential independence of higher level identity and lower level viewpoint processing in face ensemble perception. Across three experiments, our data are consistent, overall, with independent identity and viewpoint processing of face ensembles. Our first experiment found a dissociation between average and single reports for viewpoint but not for identity, as well as more precise responses for viewpoint than for identity. Adding to that, Experiment 2 found that reports of identity were not sensitive to changes in mean viewpoint. Critically, Experiment 3 found that report accuracy for a relevant attribute, identity or viewpoint, was not affected by changes in the irrelevant attribute, irrespective of whether only the relevant attribute or both attributes varied.

Across all three experiments, participants were generally more accurate at reporting an average face attribute over a single-face attribute, especially for viewpoint. This effect did not reach significance for identity in Experiment 1, but even in this case, a mean bias was confirmed for single-face reports via additional analyses. More conclusively, a mean identity advantage was reliably found in Experiment 2 across a wide range of viewpoints. These results are in broad agreement with accounts of ensemble processing emphasizing summary statistics at the expense of single-item encoding (Alvarez, 2011). More importantly, they serve as a platform for evaluating the relationship between viewpoint and identity in ensemble processing. Several results are notable in this respect.

First, ensemble processing exhibited different patterns for viewpoint and identity in Experiment 1. Specifically, a mean advantage, as indexed by report accuracy, was found to be significant only for viewpoint, as mentioned before. Yet a mean bias for single reports was more pronounced in the case of identity than for viewpoint. Further, correlations of behavioral performance, as indexed by RT, were maximized within attributes (i.e., when relating average and single-identity conditions) rather than across attributes (e.g., average identity and viewpoint).

Second, average viewpoints were reported more accurately when the mean was closer to a frontal orientation in Experiment 1, while identity reports were roughly invariant to changes in mean viewpoint in Experiment 2. Somewhat surprisingly, the latter result held not only for average identity reports but also for single-face identity reports. This is in contrast to viewpoint sensitivity documented for single face, as illustrated by the three-quarters-viewpoint advantage for face identification (Bruce & Valentine, 1987; Hill et al., 1997; O'Toole, Edelman & Bülthoff, 1998). However, this sensitivity is a source of debate (Liu & Chaudhuri, 2002; Burke, Taubert, & Higman, 2007), and methodological aspects related to response selection (e.g., navigating a stimulus continuum) may also limit the impact of viewpoint sensitivity here. Additionally, the representation of single faces that are processed in isolation may differ from that of faces presented in the context of an ensemble, a topic that awaits investigation in the ensemble literature.

Third, to address the potential independence between ensemble identity and viewpoint processing, in Experiment 3 we attempted to maximize the possibility of interaction. Specifically, we asked participants to report, on any given trial, either one or both attributes, and we cued them regarding the type of report (identity, viewpoint, or both) only after the face ensemble was removed from the screen. Yet even in this case, we found that variations in the response-irrelevant attribute did not affect reports of the relevant attribute—that is, accuracy was not affected by whether the irrelevant attribute was set to the same value across the six faces of an ensemble or varied across them.

Fourth, faces to the left and right of fixation were more effectively integrated into the representation for ensemble identity, while a similar result did not reach significance for viewpoint. Overall, this is consistent with the idea that ensemble processing is dominated by a subset of ensemble elements (Whitney & Leib, 2018), and it further suggests that patterns of spatial integration for ensembles might vary across different attributes even within the same visual domain (e.g., faces).

Thus, the present results highlight systematic differences between the processing of ensemble face identity and viewpoint. These findings extend and complement a series of recent findings. For instance, they confirm the hypothesis of viewpoint-invariant processing for face ensembles, as initially reported by Leib et al. (2014), and provide evidence for its validity across a wide range of mean viewpoints. Further, they demonstrate that, conversely, viewpoint judgments are unaffected by variation in identity.

Importantly, our results complement previous findings which show domain-specific independence in ensemble encoding (Haberman et al., 2015), by demonstrating independence within the same visual domain (i.e., faces). This independence suggests that ensemble perception is highly adaptable to properties of increasing visual complexity and that it is pervasive throughout the visual hierarchy dedicated to visual recognition, convergent with the idea that ensemble processing defines a fundamental and versatile aspect of visual perception.

Interestingly, recent work (Cant et al., 2015) documents the presence of an interaction between the processing of shape and surface properties for object ensembles. We note, though, that both shape and surface in this case are identity-related properties, hence these results are not at odds with the present findings. Specifically, our experiments manipulate viewpoint as an attribute unrelated to identity. Thus, they do not preclude the possibility of interaction in ensemble processing between attributes more tightly linked with each other in a given visual domain.

Collectively, the current findings will assist in framing suitable models of ensemble encoding with regard to its computational and neural underpinnings. For instance, the differential weighing of ensemble elements into summary representations has been evaluated in terms of its efficiency and optimality in different contexts (Haberman & Whitney, 2010; de Gardelle & Summerfield, 2011; Solomon, Morgan, & Chubb, 2011). Our results point to the possibility of different weighing schemas within the same domain (i.e., faces) across different attributes, and thus provide a new testing ground for computational accounts of optimal averaging.

On a related topic, the current findings raise a number of questions regarding the integration of multiple types of visual information into ensemble representations. For instance, face identity was determined here by shape. However, recent work has emphasized the importance of surface properties for face identification (Burton, Schweinberger, Jenkins, & Kaufmann, 2015; Andrews, Baseler, Jenkins, Burton, &Young, 2016). Hence, it is important to clarify whether and how shape and surface interact in the processing of face ensembles, as well as how efficiently different visual features (e.g., the shape of the eyes, the texture of the cheeks) are integrated across face ensembles.

Last, little is currently known regarding the neural basis of ensemble processing. Recent work (Im et al., 2017) appears to implicate the dorsal visual stream in expression ensemble perception. At the same time, though, object ensembles seem to recruit the parahippocampal place area (Cant & Xu, 2012, 2017), a region classically associated with scene processing (Epstein & Kanwisher, 1998; Park, Brady, Greene, & Oliva, 2011). While the parahippocampal place area is not robustly activated by viewing single faces, it does appear to be involved in the processing of shape and surface for object ensembles. Hence, it will be informative to examine its involvement along with that of dorsal areas in the processing of face ensembles. More generally, future work will be needed to uncover the representational basis and spatiotemporal dynamics of neural ensemble processing both within and beyond the domain of face perception.

Conclusions

Across a sequence of three experiments, we provide evidence and argue for independent processing of identity and viewpoint for face ensembles. More generally, these findings argue for independence in the processing of higher and lower level ensemble features not only across visual domains (e.g., faces versus objects) but within them as well.

Keywords: ensemble encoding, face processing, visual working memory

Acknowledgments

This work was supported by the Natural Science and Engineering Research Council to JSC and AN. The authors would like to thank members of the Visual Perception lab for their assistance in collecting and analyzing data.

*AN and JSC contributed equally to this work.

Commercial relationships: none.

Corresponding author: Marco A. Sama.
Email: marco.sama@mail.utoronto.ca.
Address: Department of Psychology, University of Toronto Scarborough, Toronto, Canada.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *16*(3), 122–131, https://doi.org/10.1016/j.tics.2011.01.003.

Andrews, T. J., Baseler, H., Jenkins, R., Burton, M. A., & Young, A. W. (2016). Contributions of feature shapes and surface cues to the recognition and neural representation of facial identity. *Cortex*, *83*, 280–291, https://doi.org/10.1016/j.cortex.2016.08.008.

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

Axelrod, V., & Yovel, G. (2012). Hierarchical processing of face viewpoint in human visual cortex. *The Journal of Neuroscience*, *32*(7), 2442–2452, https://doi.org/10.1523/JNEUROSCI.4770-11.2012.

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., . . . Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant-stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology*, *26*(5), 423–455, https://doi.org/10.1080/02643290903343149.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392, https://doi.org/10.1177/0956797610397956.

Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences, USA*, *113*(27), 7459–7464, https://doi.org/10.1073/pnas.1520027113.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Bruce, V., & Valentine, T. (1987). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology*, *1*(2), 109–120, https://doi.org/10.1002/acp.2350010204.

Burke, D., Taubert, J., & Higman, T. (2007). Are face representations viewpoint dependent? A stereo advantage for generalizing across different views of faces. *Vision Research*, *47*(16), 2164–2169, https://doi.org/10.1016/j.visres.2007.04.018.

Burton, M. A., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments against a configural processing account of familiar face recognition. *Perspectives on Psychological Science*, *10*(4), 482–496, https://doi.org/10.1177/1745691615583129.

Caharel, S., Collet, K., & Rossion, B. (2015). The early visual encoding of a face (N170) is viewpoint-dependent: A parametric ERP-adaptation study. *Biological Psychology*, *106*, 18–27, https://doi.org/10.1016/j.biopsycho.2015.01.010.

Cant, J. S., Sun, S. Z., & Xu, Y. (2015). Distinct cognitive mechanisms involved in the processing of single objects and object ensembles. *Journal of Vision*, *15*(4):12, 1–12, https://doi.org/10.1167/15.4.12. [PubMed] [Article]

Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. *The Journal of Neuroscience*, *32*(22), 7685–7700, https://doi.org/10.1523/JNEUROSCI.3325-11.2012.

Cant, J. S., & Xu, Y. (2017). The contribution of object shape and surface properties to object ensemble representation in anterior-medial ventral visual cortex. *Journal of Cognitive Neuroscience*, *29*(2), 398–412, https://doi.org/10.1162/jocn_a_01050.

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences*, *20*(5), 324–335, https://doi.org/10.1016/j.tics.2016.03.006.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57, https://doi.org/10.1177/0963721409359277.

Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic Bulletin & Review*, *14*(4), 620–628, https://doi.org/10.3758/BF03196811.

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 94–107, https://doi.org/10.1037/0096-1523.35.1.94.

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A*, *18*(5), 1016–1026, https://doi.org/10.1364/JOSAA.18.001016.

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, https://doi.org/10.1080/17470210902811249.

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences, USA*, *108*(32), 13341–13346, https://doi.org/10.1073/pnas.1104517108.

Dubois, J., de Berker, O., & Tsao, D. Y. (2015). Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *The Journal of Neuroscience*, *35*(6), 2791–2802, https://doi.org/10.1523/JNEUROSCI.4037-14.2015.

Duchaine, B. C., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585, https://doi.org/10.1016/j.neuropsychologia.2005.07.001.

Epstein, R., & Kanwisher, N. (1998, April 9). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601, https://doi.org/10.1038/33402.

Ewbank, M. P., Smith, W. A. P., Hancock, E. R., & Andrews, T. J. (2008). The M170 reflects a viewpoint-dependent representation for both familiar and unfamiliar faces. *Cerebral Cortex*, *2*(18), 364–370, https://doi.org/10.1093/cercor/bhm060.

Florey, J., Clifford, C. W. G., Dakin, S., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Report*, *6*(32210), 1–12, https://doi.org/10.1038/srep32210.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Lawrence Erlbaum Associates.

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, *45*(1), 32–41, https://doi.org/10.1016/j.neuropsychologia.2006.04.015.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science*, *16*(2), 152–160, https://doi.org/10.1111/j.0956-7976.2005.00796.x.

Guntupalli, J. S., Wheeler, K. G., & Gobbini, M. I. (2017). Disentangling the representation of identity from head view along the human face processing pathway. *Cerebral Cortex*, *27*(1), 46–53, https://doi.org/10.1093/cercor/bhw344.

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, https://doi.org/10.1037/xge0000053.

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

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics*, *72*(7), 1825–1838, https://doi.org/10.3758/APP.72.7.1825.

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition*, *62*(2), 201–222, https://doi.org/10.1016/S0010-0277(96)00785-8.

Hochstein, S., Pavlovskaya, M., Bonneh, Y. S., & Soroker, N. (2015). Global statistics are not neglected. *Journal of Vision*, *15*(4):7, 1–17, https://doi.org/10.1167/15.4.7. [PubMed] [Article]

Im, H. Y., Albohn, D. N., Steiner, T. G., Cushing, C. A., Adams, R. B., & Kveraga, K. (2017). Differential hemispheric and visual stream contributions to ensemble encoding of crowd emotion. *Nature Human Behaviour*, *1*, 828–842, https://doi.org/10.1038/s41562-017-0225-z.

Ishai, A. (2008). Let's face it: It's a cortical network. *NeuroImage*, *40*(2), 415–419, https://doi.org/10.1016/j.neuroimage.2007.10.040.

Ji, L., Rossi, V., & Pourtois, G. (2018). Mean emotion from multiple facial expressions can be extracted with limited attention: Evidence from visual ERPs. *Neuropsychologia*, *111*, 92–102, https://doi.org/10.1016/j.neuropsychologia.2018.01.022.

Jiang, F., Blanz, V., & Rossion, B. (2011). Holistic processing of shape cues in face identification: Evidence from face inversion, composite faces, and acquired prosopagnosia. *Visual Cognition*, *19*(8), 1003–1034, https://doi.org/10.1080/13506285.2011.604360.

Jiang, Y. V., Shim, W. M., & Makovski, T. (2008). Visual working memory for line orientations and face identities. *Perception & Psychophysics*, *70*(8), 1581–1591, https://doi.org/10.3758/PP.70.8.1581.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304, https://doi.org/10.1146/annurev.psych.55.090902.142005.

Kietzmann, T. C., Gert, A. L., Tong, F., & König, P. (2017). Representational dynamics of facial viewpoint encoding. *Journal of Cognitive Neuroscience*, *29*(4), 637–651, https://doi.org/10.1162/jocn_a_01070.

Kuo, P., Chen, Y., & Chen, L. (2018). Manifold decoding for neural representations of face viewpoint and gaze direction using magnetoencephalographic data. *Human Brain Mapping*, *39*, 2191–2209, https://doi.org/10.1002/hbm.23998.

Lai, M., Oruc, I., & Barton, J. J. (2013). The role of skin texture and facial shape in representations of age and identity. *Cortex*, *49*(1), 252–265, https://doi.org/10.1016/j.cortex.2011.09.010.

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. *Journal of Vision*, *14*(8):26, 1–13, https://doi.org/10.1167/14.8.26. [PubMed] [Article]

Leib, A. Y., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, *7*, 13186, https://doi.org/10.1038/ncomms13186.

Liu, C. H., & Chaudhuri, A. (2002). Reassessing the 3/4 view effect in face recognition. *Cognition*, *83*(1), 31–48, https://doi.org/10.1016/S0010-0277(01)00164-0.

Logie, R. H., Baddeley, A. D., & Woodhead, M. M. (1987). Face recognition, pose and ecological validity. *Applied Cognitive Psychology*, *1*(1), 53–69, https://doi.org/10.1002/acp.2350010108.

Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281, https://doi.org/10.1038/36846.

Nestor, A., Plaut, D. C., & Behrmann, M. (2016). Feature-based face representations and image reconstruction from behavioral and neural data. *Proceedings of the National Academy of Sciences, USA*, *113*(2), 416–421, https://doi.org/10.1073/pnas.1514551112.

Neumann, M. F., Ng, R., Rhodes, G., & Palermo, R. (2017). Ensemble coding of face identity is not independent of the coding of individual identity. *The Quarterly Journal of Experimental Psychology*, *71*(6), 1–27, https://doi.org/10.1080/17470218.2017.1318409.

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

Or, C. C., & Wilson, H. R. (2010). Face recognition: Are viewpoint and identity processed after face detection? *Vision Research*, *50*(16), 1581–1589, https://doi.org/10.1016/j.visres.2010.05.016.

O'Toole, A. J., Edelman, S., & Bülthoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research*, *38*(15–16), 2351–2353, https://doi.org/10.1016/S0042-6989(98)00042-X.

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: Complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *The Journal of Neuroscience*, *31*(4), 1333–1340, https://doi.org/10.1523/JNEUROSCI.3885-10.2011.

Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F., & Vuilleumier, P. (2005). Portraits or people? Distinct representations of face identity in the human visual cortex. *Journal of Cognitive Neuroscience*, *17*(7), 1043–1057, https://doi.org/10.1162/0898929054475181.

Purves, D., Monson, B. B., Sundararajan, J., & Wojtach, W. T. (2014). How biological vision succeeds in the physical world. *Proceedings of the National Academy of Sciences, USA*, *111*(13), 4750–4755, https://doi.org/10.1073/pnas.1311309111.

Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience*, *13*(6), 766–785, https://doi.org/10.1162/08989290152541430.

Ramírez, F. M. (2018). Orientation encoding and viewpoint invariance in face recognition: Inferring the neural properties from large-scale signals. *The Neuroscientist*, *24*(6), 1–7, https://doi.org/10.1177/1073858418769554.

Ramírez, F. M., Chichy, R. M., Allefeld, C., & Haynes, J.-D. (2014). The neural code for face orientation in the human fusiform face area. *The Journal of Neuroscience*, *34*(36), 12155–12167, https://doi.org/10.1523/JNEUROSCI.3156-13.2014.

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision*, *11*(12):13, 1–11, https://doi.org/10.1167/11.12.13. [PubMed] [Article]

Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychological Science*, *25*(10), 1903–1913, https://doi.org/10.1177/0956797614544510.

Whitney, D., & Leib, A. Y. (2018). Ensemble perception. *Annual Review of Psychology*, *69*, 105–129, https://doi.org/10.1146/annurev-psych-010416-044232.

Wolfe, B. A., Kosovicheva, A. A., Leib, A. Y., Wood, K., & Whitney, D. (2015). Foveal input is not required for perception of crowd facial expression. *Journal of Vision*, *15*(4):11, 1–13, https://doi.org/10.1167/15.4.11. [PubMed] [Article]

Wong, J. H., Peterson, M. S., & Thompson, J. C. (2008). Visual working memory capacity for objects from different categories: A face-specific maintenance effect. *Cognition*, *108*(3), 719–731, https://doi.org/10.1016/j.cognition.2008.06.006.

ZeeAbrahamsen, E., & Haberman, J. (2018). Correcting "confusability regions" in face morphs. *Behavior Research Methods*, *50*(4), 1686–1693, https://doi.org/10.3758/s13428-018-1039-2.

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*(4), 423–428, https://doi.org/10.1111/j.1467-9280.2009.02322.x.

Zhen, Z., Fang, H., & Liu, J. (2013). The hierarchical brain network for face recognition. *PLoS One*, *8*(3), e59886, https://doi.org/10.1371/journal.pone.0059886.