Behavioral/Cognitive

The Neural Dynamics of Face Ensemble and Central Face Processing

Marco Agazio Sama, Adrian Nestor,* and Jonathan Samuel Cant*
Department of Psychology, University of Toronto Scarborough, Toronto, Ontario M1C 1A4, Canada

Extensive work has investigated the neural processing of single faces, including the role of shape and surface properties. However, much less is known about the neural basis of face ensemble perception (e.g., simultaneously viewing several faces in a crowd). Importantly, the contribution of shape and surface properties have not been elucidated in face ensemble processing. Furthermore, how single central faces are processed within the context of an ensemble remains unclear. Here, we probe the neural dynamics of ensemble representation using pattern analyses as applied to electrophysiology data in healthy adults (seven males, nine females). Our investigation relies on a unique set of stimuli, depicting different facial identities, which vary parametrically and independently along their shape and surface properties. These stimuli were organized into ensemble displays consisting of six surround faces arranged in a circle around one central face. Overall, our results indicate that both shape and surface properties play a significant role in face ensemble encoding, with the latter demonstrating a more pronounced contribution. Importantly, we find that the neural processing of the center face precedes that of the surround faces in an ensemble. Further, the temporal profile of center face decoding is similar to that of single faces, while those of single faces and face ensembles diverge extensively from each other. Thus, our work capitalizes on a new center-surround paradigm to elucidate the neural dynamics of ensemble processing and the information that underpins it. Critically, our results serve to bridge the study of single and ensemble face perception.

Key words: **EEG; face ensembles; neural dynamics; pattern analysis; shape cues; surface cues**

Significance Statement

We often view groups of faces, called a "face ensemble," whenever we encounter multiple people (e.g., on a crowded subway). Yet, little is known about how the brain represents ensembles and how attention to a single individual (e.g., during a face-to-face conversation in a crowd) impacts ensemble representations. Here, we show that shape and surface properties contribute to ensemble processing, with shape exhibiting a different neural profile for single versus ensemble face representations. We also show that the visual system processes central faces within an ensemble differently from unattended faces. These results indicate that the encoding of ensemble and single faces is mediated by distinct neural mechanisms, highlighting the importance of studying both as well as their relationship.

Introduction

Visual face processing leverages information about shape (i.e., the configuration of face parts and their spacing; Piepers and Robbins, 2012) and surface properties (i.e., facial pigmentation and textural properties; Russell et al., 2006). Extensive research has documented the neural mechanisms of face processing (Tsao and Livingstone, 2008; Little et al., 2011; Duchaine and Yovel, 2015), including the representation of shape and surface properties (O'Toole et al., 1999; Russell et al., 2007, 2012; Jiang et al., 2009; Burton et al., 2015; Andrews et al., 2016). While these properties play partly different roles in single-face processing (e.g., in the representation of identity; Bruce and Young, 1998), they work together to benefit face recognition (Dzhelyova and Rossion, 2014). Of Relevance here, the neural dynamics of face processing (Dobs et al., 2019; Smith and Smith, 2019; Bae, 2020, Muukkonen et al., 2020) evince early encoding of perceptual information (e.g., around 100 ms after stimulus onset; Nemrodov et al., 2016), with partial overlap between shape and surface encoding (e.g., the latter being encoded prior to the former; Nemrodov et al., 2019a).

Despite these advances, a broader picture of human face processing is limited by the focus on single faces. Often, individuals are simultaneously exposed to multiple faces in everyday life (e.g.,

Received June 9, 2023; revised Nov. 21, 2023; accepted Dec. 11, 2023.
Author contributions: M.A.S., A.N., and J.S.C. designed research; M.A.S. performed research; M.A.S., A.N., and J.S.C. contributed unpublished reagents/analytic tools; M.A.S. analyzed data; M.A.S. wrote the paper.
This work was funded by an NSERC discovery grant to A.N. and J.S.C.
*A.N. and J.S.C. contributed equally to this work.
The authors declare no competing financial interests.
Correspondence should be addressed to Marco Agazio Sama at marco.sama@mail.utoronto.ca.
https://doi.org/10.1523/JNEUROSCI.1027-23.2023

on a crowded subway train), referred to as a "face ensemble." Extracting information from the multitude of items in an ensemble poses a challenge due to limitations in visual working memory capacity (Luck and Vogel, 1997; Raffone and Wolters, 2001; Cowan, 2010). To circumvent this, the visual system compresses redundant statistical information into a single summary metric (e.g., an average face identity; Alvarez, 2011; Whitney and Yamanashi Leib, 2018; for review, see Corbett et al., 2023; but see Ji et al., 2020). Ensembles of stimuli benefiting from perceptual expertise, such as faces, are processed highly efficiently, even though ensemble processing may not rely on exemplar-level knowledge (Cha et al., 2020).

Research on the neural underpinnings of face ensemble processing is relatively sparse. One study found a positive relationship between N170 amplitude and the number of faces in a display (Puce et al., 2013). Another study, utilizing a multivariate approach, found distinct neural profiles for single and ensemble faces (Roberts et al., 2019). A third study showed minimal attention is required to extract face ensemble summaries for emotion (Ji et al., 2018). However, to date, the contribution of shape and surface information in face ensemble encoding has not been addressed.

Even less understood is the representation of single faces within the context of an ensemble. Most studies on ensemble perception employ an exemplar-cueing paradigm, in which participants are instructed to identify if a particular item, such as an object or a face, was a member of a previously seen display (e.g., de Fockert and Wolfenstein, 2009; Haberman and Whitney, 2009). However, this does not address how a particular target central face is represented relative to the overall ensemble (e.g., an individual engaging with you in a face-to-face conversation on a crowded subway train). For instance, it is not known whether the visual system simply includes the central face in the ensemble summary, or whether it processes it differently.

Given these gaps in our current understanding, here, we aim to elucidate the neural representation of face ensembles, appealing to pattern analysis as applied to electroencephalography (EEG) data. Specifically, we address the relative contribution of shape and surface properties to face ensemble perception, their accompanying temporal profiles, and the processing of single faces within the context of ensembles. To this aim, we use a unique stimulus design that incorporates a central face in the ensemble, and we independently vary the shape and surface properties of face stimuli. In summary, this work provides, to the best of our knowledge, the first in-depth investigation into the underlying contribution of shape and surface properties in face ensemble encoding and into the contribution of central face perception to neural ensemble processing.

Materials and Methods

Participants. Seventeen healthy adult participants were recruited. The data of one participant were discarded due to excessively noisy recordings. Of the remaining 16, 7 were males, and 9 were females, with an age range of 18–28 years. All participants were right-handed, had normal or corrected-to-normal visual acuity, and had adequate face processing abilities as assessed with the aid of the Cambridge Face Memory Test (Duchaine and Nakayama, 2006). Participants gave informed consent and received monetary compensation upon completion of the experiment. The University of Toronto Research Ethics Board provided ethical approval to conduct this study.

Stimulus design. We sought to manipulate shape and surface properties independently to evaluate their contribution to facial identity processing in face ensembles. Critically, we aimed to ensure that faces within a given set could be discriminated from faces in different sets, and, by extension, that these ensembles yield different summary identities, which is an important aspect when considering comparisons of central and surround faces (see below). This involved three general steps: (1) separating relevant shape and surface properties from a set of faces; (2) computing cardinal anchor points for shape, surface, and both combined; and (3) interpolating between the anchor points to generate a matrix of parametrically varying faces.

First, we selected 60 male faces from the Radboud database (Langner et al., 2010) and cropped them with an oval mask to retain only their internal features (Fig. 1*A*). A total of 82 fiducial points were marked using

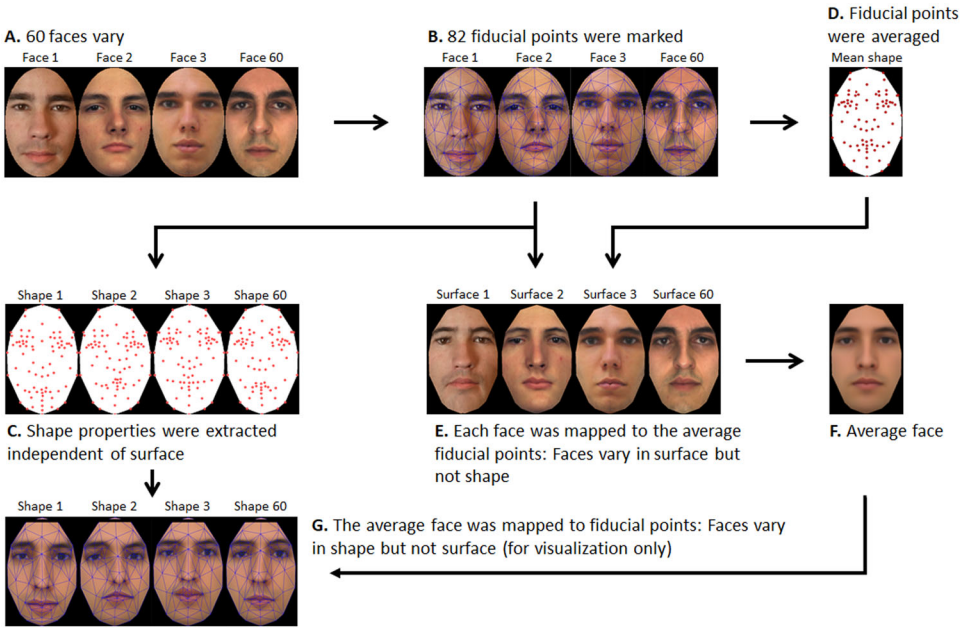
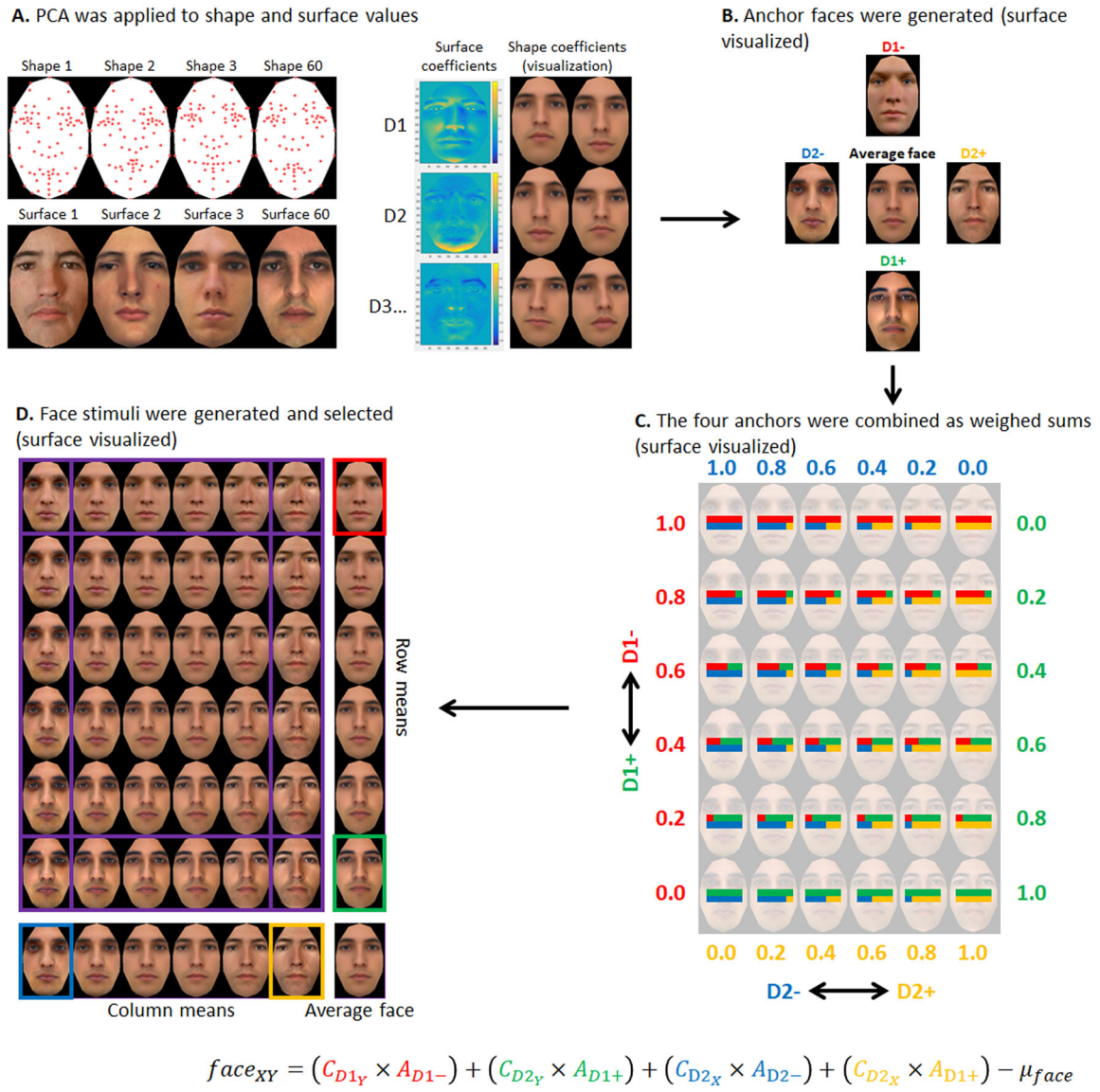


Figure 1. Disentangling face shape and surface information. ***A***, Sixty faces were marked with (***B***) 82 fiducial points accounting for facial shape. ***C***, The coordinates of these points were extracted and (***D***) averaged as the mean shape. ***E***, All faces were subsequently remapped to this mean shape, resulting in faces that varied only in surface properties. ***F***, A pixelwise average of these faces yielded an average face. ***G***, The average face could then be mapped to the fiducial points of each of the 60 faces to yield faces that vary in shape but not surface.

the InterFace tool (Fig. 1*B*; Kramer et al., 2017) to extract shape information. Fiducial points were averaged (Fig. 1*C*) across the 60 faces to generate an average shape (Fig. 1*D*). All 60 face images were subsequently warped to this shape average to provide a set of faces that varied in surface, with shape held constant (Fig. 1*E*). The pixelwise average of these faces yielded a face average across variation in both shape and surface (Fig. 1*F*). Finally, mapping this face average to the 60 sets of original shape coordinates generated faces that changed in shape, with surface properties held constant (Fig. 1*F*)—these were used only as a visual guide for shape variations as subsequent manipulations were conducted on shape coordinates and not the images themselves.

Second, to design face stimuli whose shape and surface could be independently controlled, as well as discriminated based on set membership, we sought to parametrize the stimuli, ensuring similarity between faces within an ensemble set, but not across sets. To this end, we applied principal component analysis (PCA) separately to shape coordinates and surface images (Fig. 2*A*). Two dimensions for both shape and surface were derived from resulting PCs that contained information related to facial identity. For surface, Dimension 1 (D1) was provided by PC 2, which explained 15.7% variance while Dimension 2 (D2) relied on a combination of PCs 4 and 8, explaining a total of 6.4% variance—PCs 1, 3, 5, and 7 were not selected as they captured information unrelated to facial identity (e.g., lighting). For shape, D1 and D2 were provided by PC1 and PC2, which explained 48.9% and 12.4% of the total variance, respectively. New face images were generated along the two dimensions relative to the average face yielding four cardinal anchors: D1−, D1+, D2−, and D2+ (Fig. 2*B*, surface anchors). Of note, anchors for both attributes were equidistant from the average face as measured via an L2 pixelwise metric to ensure that variation in low-level image properties is comparable for shape and surface across their corresponding dimensions.

Third, a face matrix that varied only in surface properties was generated by a weighted sum of the four anchor faces derived for surface



$$face_{XY} = \left(C_{D1_Y} \times A_{D1-}\right) + \left(C_{D2_Y} \times A_{D1+}\right) + \left(C_{D2_X} \times A_{D2-}\right) + \left(C_{D2_X} \times A_{D1+}\right) - \mu_{face}$$

Figure 2. Generating a stimulus matrix of faces that vary parametrically in shape and surface properties. ***A***, PCA was separately applied to shape coordinates and to surface images previously derived from a set of faces (Fig. 1). ***B***, Four cardinal anchors were generated as the endpoints of two dimensions, equidistantly from the average face in image space. ***C***, The anchors were linearly interpolated (color-coded bars indicate the proportional contribution of each dimension and polarity to any given cell in the matrix). ***D***, A matrix of stimuli varying in surface properties based on the stimulus design schema from (***C***). In the bottom formula, A are anchors, C are coefficients based on the *x* and *y* position of a face in the matrix, and μ is the average face. Four stimulus sets were created by extracting faces along the edges of the matrix (purple boxes). Only faces bounded by these boxes were used as experimental stimuli. Row and column averages are displayed next to the face matrix. This procedure was similarly applied to shape coordinates to yield stimuli that only varied in shape, which were subsequently combined with surface information to produce stimuli varying in both shape and surface properties.

dimensions (Fig. 2*C*). To be clear, all faces in this matrix had the same average shape. This process was repeated for shape, using fiducial points instead of surface, and the resulting coordinates were then used to warp the face average into faces that varied in shape properties only. In addition, a third set of faces was constructed to vary in both shape and surface properties by mapping coordinates from the shape matrix onto the corresponding faces in the surface matrix.

To maximize between-set discriminability, stimulus sets were extracted from the edge of the face matrix (Fig. 2*D*). Sets were labeled in accordance with their closest anchor face. Thus, the final set of stimuli included four unique sets of faces (labeled D1−, D1+, D2−, and D2+), which varied in shape, surface, or both (Fig. 3).

The ability to control within- and between-set resemblance is a benefit of the current procedure. Individual identities within a set were relatively similar compared to faces in a different set, ensuring that the resulting ensembles were discriminable from each other. This also allowed us to manipulate consistency between an ensemble's center and surround: in consistent ensembles, center and surround faces were selected from the same set, whereas in inconsistent ensembles, center and surround faces were selected from different sets and, hence, exhibited a noticeably different appearance.

However, we note that, despite the four sets appearing perceptually dissimilar, faces still exhibit high levels of low-level image similarity across sets (Pearson's r for pairwise image correlations across pixel RGB values range between 0.83 and 0.99, with most >0.90). Thus, we seek to ensure that differences in neural patterns are not simply driven by low-level variability across face images.

To design ensemble stimuli, all six identities from a given set and attribute group were placed in a circle (i.e., as surround faces) around a face at the center (Fig. 4). The consistency of the center-surround arrangement was manipulated as follows: a consistent ensemble would contain faces from the same set (e.g., D1+ surround paired with the mean of D1+ as the central face), whereas an inconsistent ensemble would contain faces from sets with opposite polarities (e.g., D1+ surround paired with the mean of D1− as the central face). The placement of the six surround faces in the circular arrangement was randomized.

Of note, using a set mean as the central face, instead of specific set members, presumably maximizes the consistency of the center to the surround. This is because the representation of face ensembles is reduced to the mean rather than any exemplar face, as suggested by both behavioral (de Fockert and Wolfenstein, 2009; Haberman and Whitney, 2009) and neural results (Roberts et al., 2019).

Experimental design. Participants completed two EEG recording sessions, each lasting around 2 h including equipment setup, during which they viewed single-face stimuli or face ensembles. Each session consisted of 6 single-face and 8-face ensemble runs, totaling 12 and 16 runs across both sessions, respectively. Ensemble runs were repeated four times consecutively, followed by three consecutive single runs. The order of ensemble and single-face runs was counterbalanced based on session and participant number.

Participants sat in a quiet, dark room, 60 cm away from a 1,920 × 1,080 pixel resolution monitor (60 Hz) displaying a black screen and fixation cross in the center. Stimuli were displayed using Psychtoolbox for MATLAB (Brainard, 1997). All trials, irrespective of run type, began with a 100 ms cue consisting of a bright fixation cross, which was followed by a 300 ms stimulus display (i.e., a single face or a face ensemble), and ended with a variable 700–800 ms blank screen with a dimmer central fixation cross. Participants were tasked with responding to oddball trials during this blank period by pressing the space bar if a female face was present in the center of the screen, either as a single face presented in isolation or the central face of the ensemble. All oddball trials were discarded from data analysis. Prior to each session, participants also completed a short ensemble and single-face practice run, consisting of half the trials of an experimental run. Feedback on accuracy was given only during practice.

Importantly, regarding the oddball task, both shape and surface properties contribute significantly to sex perception in human observers (Bruce et al., 1993; Bruce and Langton, 1994), and they provide complementary information in sex recognition tasks (Nestor and Tarr, 2008). Hence, we believe that the use of our oddball task (i.e., sex recognition) was unlikely to radically shift attention to one type of facial property over the other.

For each single-face run, all face stimuli were displayed in random order to participants. We note that, due to our stimulus design procedure (Fig. 2), Identities 1 and 6 are shared between dimensions, and so these faces were only displayed once. This yielded a total of 24 faces (20 unique single faces plus 4 set mean faces) for each of the three attribute groups (shape, surface, and both). Each face was repeated three times for a total of 240 trials per run, which included an additional 24 oddball trials consisting of six female faces repeated four times (i.e., 24 identities × 3 attribute groups × 3 repetitions + 24 oddball trials). Each ensemble run had a total of 312 trials, containing the same 24 female oddballs as central target faces (i.e., 2 consistency conditions × 3 attribute groups × 4 sets × 12 repetitions + 24 oddball trials). Individual faces in ensembles as well as

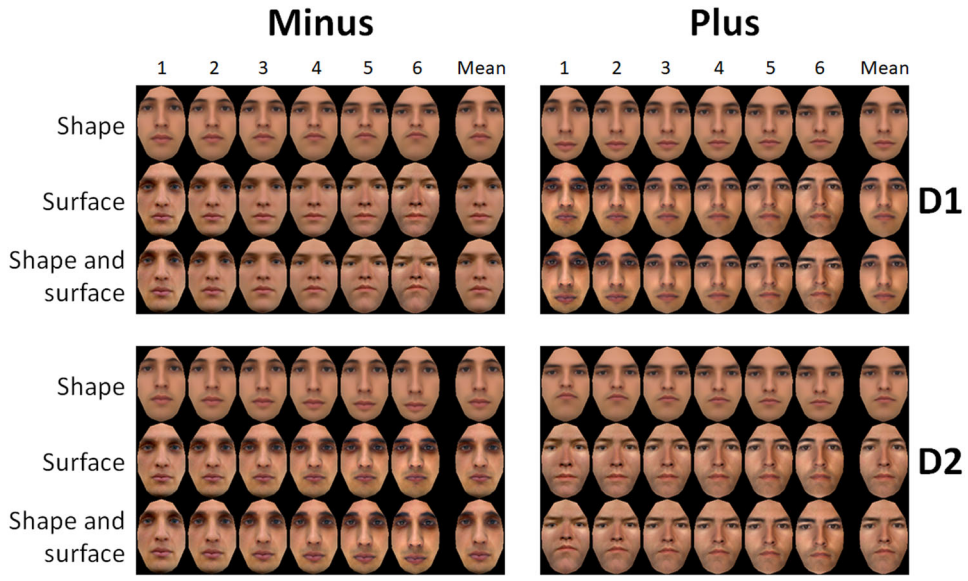


Figure 3. All single-face stimuli. There were four sets of faces, split across dimensions D1 and D2, and across plus and minus polarities. Six faces varied parametrically in shape, surface, or both. The mean is shown in the rightmost column of each panel (note that Faces 1 and 6 are shared between neighboring stimulus sets as a result of how they were extracted from the face matrix; Fig. 2*D*).

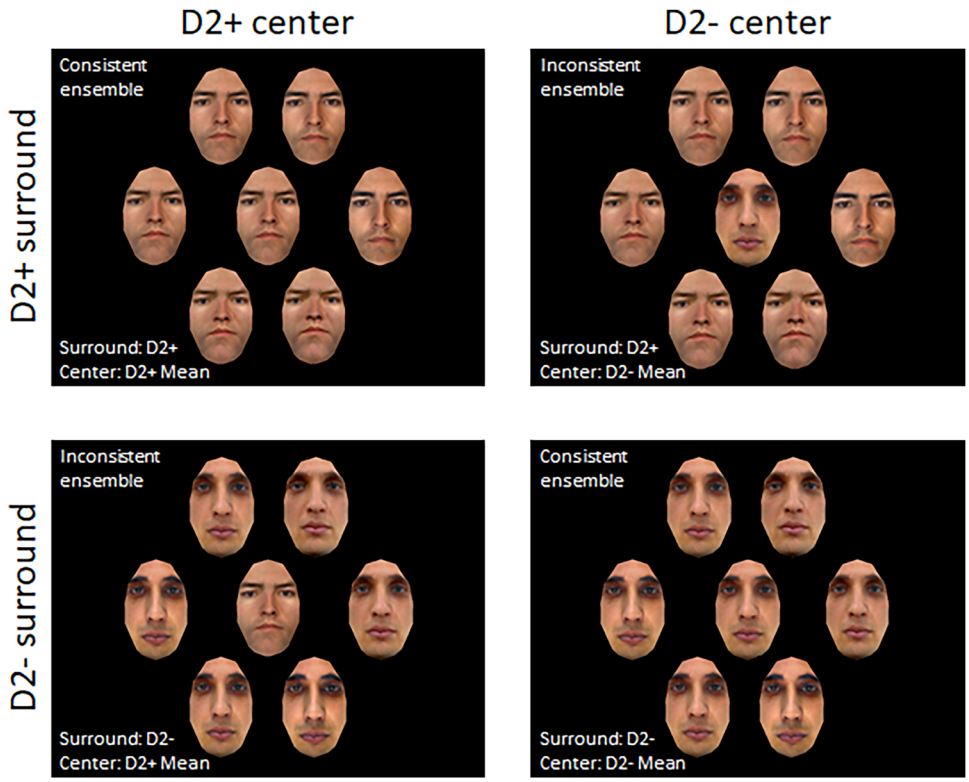


Figure 4. Examples of ensemble stimuli from set D2 that vary in both shape and surface properties. Six faces form the ensemble surround, and the mean of the set of the same or opposite polarity provides the ensemble's central face. Consistent ensembles have matching centers and surrounds (e.g., D2+ surround, and the mean of D2+ as center). Inconsistent ensembles have opposing centers and surrounds (e.g., D2+ surround, and the mean of D2− as center).

single-face stimuli subtended a visual angle of 2.1° × 3.3° from the 60 cm viewing distance. Entire ensemble stimuli subtended 9.4° × 10.7° visual angle.

EEG preprocessing. Data were recorded using a Biosemi ActiveTwo EEG system with 64 electrodes arranged according to the international 10/20 system. The electrode offset was kept under 40 mV. EEG data were low-pass filtered using a fifth-order sync filter with a half-power cutoff at 204.8 Hz and digitized at 512 Hz (~1.95 ms per time bin) with 24 bits of resolution.

EEG data were preprocessed offline. Data were digitally filtered (zero-phase 24 dB/octave Butterworth filter) with a bandpass of 0.01–40 Hz, along with DC removal and linear detrending for each epoch, which included −100 ms to +900 ms from stimulus onset. Epochs for false alarms and all oddball trials were discarded from analyses. Epochs were visually inspected across trials for each electrode and run, and ocular artifacts such as eye blinks were removed using Infomax ICA (Delorme et al., 2007). Overly noisy epochs were also discarded. Electrodes that were consistently noisy during a given run were interpolated between three neighboring channels. Only 2.1% of trials were removed due to false alarms, and another 0.2% were removed because of artifacts, yielding 97.7% of trials retained on average.

Statistical analysis. A subset of 12 bilateral electrodes located over homolog occipitotemporal (OT) areas (left, P5, P7, P9, PO3, PO7, and O1; right, P6, P8, P10, PO4, PO8, and O2). These channels exhibited large amplitudes for event-related potential (ERP) components associated with face processing (e.g., N170). Also, prior work has documented their ability to support identity decoding for single faces (Nemrodov et al., 2018) as well as for face ensemble stimuli (Roberts et al., 2019) at levels comparable to or better than those supported by all channels (Nemrodov et al., 2019b). For thoroughness, we considered data based on all 64 electrodes here as well. This led to a systematic decrement in decoding performance relative to the results reported below based on only 12 OT electrodes (collapsing across the pairwise comparisons reported in our multivariate analyses: $F_{(1,15)} = 25.41$, $p < 0.001$; see Multivariate results). Further, a comparison between left and right OT electrodes yielded no significant difference for any type of decoding described below ($F_{(1,15)} = 1.53$, $p = 0.235$). Taken together, these findings motivated our decision to focus here on EEG signals from the subset of OT electrodes noted above.

For univariate analyses, ERPs for all unique identities within an attribute group were averaged. We analyzed whether the variability of stimulus types (consistent ensembles, inconsistent ensembles, or single faces) modulated the amplitude and onset of the N170, P1, and P2 ERP components. The latter two, similar to N170, have been implicated in face processing (Halit et al., 2000; Mercure et al., 2008; Kuefner et al., 2010; Wang et al., 2015; Tanaka, 2018).

For multivariate pattern analyses, repetitions of all unique stimuli within a given run (i.e., repetitions of the same ensemble or single stimulus), timepoint, and OT channel were averaged and *z* scored, and outliers were winsorized to three *SD*s. The entire dataset was subsequently normalized within the range of 0–1. We computed the binary classification of stimulus pairs for each participant by training a linear support vector machine ($c = 1$) using leave-one-block-out cross-validation (i.e., 12 blocks for single-face decoding and 16 blocks for face ensemble decoding). Classification performance was assessed by comparing each participant's accuracy against permutation-based chance-level accuracy (yielding values between 49.92% and 50.05% based on 1,000 permutations) via Wilcoxon signed-rank test (*WSR*, standardized *z* values reported) and by applying appropriate corrections for multiple comparisons (see below). Analyses were conducted both jointly across time points in a large temporal window to maximize decoding success (i.e., temporally cumulative analysis), as well as for sequences of smaller temporal windows to capture the time course of decoding.

Patterns for classification analysis were constructed based on stimulus pairs that targeted specific properties of the stimuli as follows (see

Table 1. List of stimulus classes for decoding purposes

Decoding	Approach	Instances of class pairs
Single faces	Faces are decoded within dimension but across polarity (e.g., all faces from D1− vs all from D1+; Fig. 3)	74 per attribute: any 7 D1− versus any 7 D1+; any 5 D2− versus any 5 D2+*
Face ensembles	Ensembles are decoded across polarity but within set and consistency	4 per attribute: D1− C versus D1+ C; D2− C versus D2+ C; D1− I versus D1+ I; D2− I versus D2+ I
Ensemble consistency	Consistent ensembles are decoded from both inconsistent ensembles for each dimension	2 per attribute: averaged signals for D1− C and D1+ C versus D1− I and D1+ I; same for D2
Ensemble surround	Surrounds are decoded while controlling for central face and consistency information	2 per attribute: averaged signals for D1− C and D1− I versus D1+ I and D1− C; same for D2
Ensemble center	Center faces are decoded while controlling for surround and consistency information	2 per attribute: averaged signal for D1− C and D1+ I versus D1− I and D1− C; same for D2

Decoding is repeated for each of three attribute groups: shape, surface, or both. Ensemble stimuli are designated by the identity of surround faces (e.g., D1+ I designates an ensemble with surround faces from group D1+ and the mean of faces from group D1− as a center face; D, dimension; C, consistent ensemble; I, inconsistent ensemble). *Two single faces are shared between D1 and D2 for each polarity and are counted here only as members of D1.

Table 1 for examples). First, for single faces, each single face in a particular set was compared to each single face from the same dimension but opposing polarity—this also included the comparison of mean faces across polarity. Second, for face ensembles, stimuli were similarly compared across dimension polarity. Of note, they were compared within consistency (i.e., consistent stimuli were compared only to consistent stimuli, and inconsistent stimuli only to inconsistent stimuli). Third, for the purpose of consistency decoding, consistent stimuli were jointly compared to inconsistent stimuli to avoid any confounds based on the facial identity of the center or surround (e.g., the ERP profiles of D1− consistent and D1+ consistent were averaged and, then, compared to the average of ERP profiles elicited by D1− inconsistent and D1+ inconsistent). Fourth, for surround decoding, different surrounds sharing the same center faces were compared to each other. To avoid a consistency-based confound, the ERP profiles of certain consistent and inconsistent stimuli were averaged and, then, compared to their averaged counterpart (e.g., the average of D1− consistent and D1− inconsistent was decoded from the average of D1+ consistent and D1+ inconsistent). This ensured that consistent information was matched across decoding classes. Last, for target decoding, we followed a similar approach while controlling for surround information and consistency.

For temporally cumulative analyses, decoding of different stimulus classes (Table 1) was conducted across spatiotemporal patterns consisting of 3,684 features (12 OT electrodes × 307 time bins within a temporal interval of 50–650 ms). This 600 ms window was chosen to capture both early- and high-level visual information for single faces (Ghuman et al., 2014; Nemrodov et al., 2016, 2018; Vida et al., 2016) and face ensembles (Haberman et al., 2009; Roberts et al., 2019). This window was further validated by our time course analyses, revealing that most above-chance decoding took place during this interval. Differences between decoding accuracy across combined shape and surface, as well as shape and surface alone, were assessed via the nonparametric Friedman test (χ^2), with multiple comparison *WSR* tests evaluating significant omnibus findings. χ^2 effect size is reported via Kendall's *W* and the rank-biserial correlation (abbreviated as *RC*) for *WSR*.

For time course analyses, classification used a sliding ∼10 ms window, corresponding to five ∼1.95 ms time bins, between −100 ms to 700 ms relative to stimulus onset. Given the number of time bin comparisons (i.e., 358), significant levels of above-chance accuracy were determined by adopting a more conservative approach (i.e., a nonparametric sign test against 50% accuracy). To further control for false positives, we discarded significant time intervals that did not include at least two consecutive significant time bins.

All analyses were conducted using Letswave 6 (Mouraux and Iannetti, 2008), MATLAB 2016b, and JASP 0.17.1 (jasp-stats.org/). Multiple comparison correction was implemented using Bonferroni (for <10 comparisons) and false discovery rate (FDR; for ≥10 comparisons). Finally, to add confidence in significant findings and to aid the interpretation of statistically nonsignificant results, we employed Bayesian hypothesis testing (JASP). Resulting BF_{10} values are reported, providing weight in favor of the alternative ($BF_{10} > 1$) or null ($BF_{10} < 1$) hypotheses. Modeling was based on the recommended default distributions for unspecified priors: the uniform distribution for ANOVA equivalents (Rouder et al., 2012) and Cauchy's distribution for *t* test equivalents (Ly et al., 2016). Multiple corrections for omnibus results relied on a Bonferroni-like adjustment to model posteriors (Westfall et al., 1997). Below, we adopt conventional interpretations (BF in favor of H_1 over H_0, 1–3 = anecdotal, 3–10 = substantial, 10–30 = strong, 30–100 = very strong, and >100 = extreme evidence; BF in favor of H_0 over H_1, 0.33–1 = anecdotal, 0.10–0.33 = substantial, 0.01–0.03 = very strong, and <0.01 = extreme evidence; Wagenmakers et al., 2011). Overly large BF_{10} are converted to scientific notation.

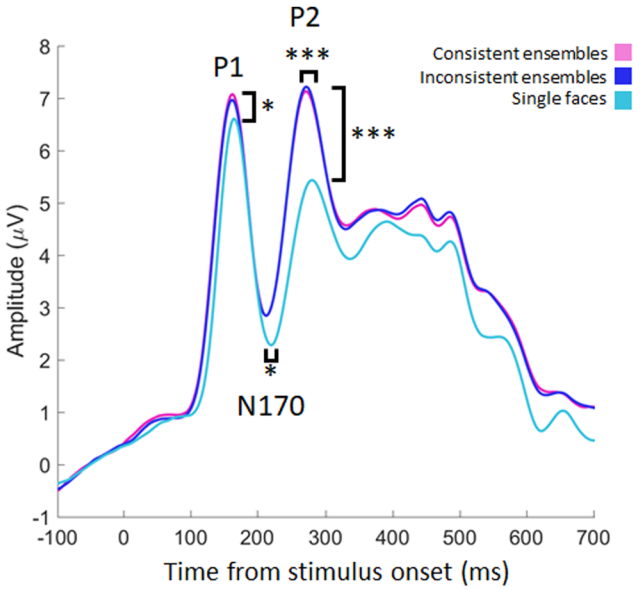


Figure 5. ERPs averaged across 12 OT channels for consistent ensemble, inconsistent ensemble, and single-face stimuli. Univariate differences were found between single faces and ensembles, but not between different types of ensembles. *$p < 0.05$, ***$p < 0.001$; Bonferroni-corrected.

Results

Univariate results

Ensemble stimuli elicited ERP signals with larger amplitude than single faces at P1 (mean difference = 0.62 μV, $t_{(30)} = 2.10$, $p = 0.044$, $d = 0.53$, $BF_{10} = 3.04$) and P2 (mean difference = 1.68 μV, $t_{(30)} = 8.67$, $p < 0.001$, $d = 2.17$, $BF_{10} = 3.14 \times 10^{10}$), as well as an earlier peak for N170 (mean difference = 4.66 ms, $t_{(30)} = 2.68$, $p = 0.024$, $d = 0.67$, $BF_{10} = 86.18$) and P2 (mean difference = 12.71 ms, $t_{(30)} = 3.85$, $p < 0.001$, $d = 0.96$, $BF_{10} = 318.32$; Fig. 5). Interestingly, there were no differences between consistent and inconsistent ensembles, as reflected by the overlap of their ERP profiles (all $p > 0.999$, all $BF_{10} < 0.73$).

These results reveal that modulation of ERP profiles is driven by differences between single and ensemble faces, which may be explained in part by differences in the number of faces in the stimulus display and visual angle as well as by differences in visual complexity. The present lack of univariate differences for center-surround consistency further motivates our multivariate analysis.

Multivariate results

Single faces

Overall, differences in single-face decoding accuracy were observed across attribute groups ($\chi^2_{(2)} = 23.63$, $p < 0.001$, $W = 0.738$, $BF_{10} = 3.61 \times 10^9$). Significant decoding was found for face pairs in each of the three attribute groups (shape and surface combined, accuracy = 62.8 ± 1.2%, $z = 3.52$, $p < 0.001$, RC = 1.000, $BF_{10} = 1.02 \times 10^6$; shape, accuracy = 52.7 ± 0.6%, $z = 3.26$, $p = 0.002$, RC = 0.926, $BF_{10} = 165.90$; surface, accuracy = 64.9 ± 1.2%, $z = 3.52$, $p < 0.001$, RC = 1.000, $BF_{10} = 5.27 \times 10^6$; Fig. 6*A*). Surface information, both by itself or when combined with shape, contributed more to decoding relative to shape by itself (in both cases: mean difference >10.1 ± 1.3%, both $z > 3.46$, both ps < 0.002, both RC > 0.999, both $BF_{10} > 13{,}818.64$). No difference in decoding accuracy was observed between the surface by itself or when combined with shape (mean difference = 2.2 ± 0.9%, $z = 2.22$, $p = 0.079$, $BF_{10} = 2.80$).

The time course of single-face processing (Fig. 7*A*) revealed reliable pairwise decoding of identity across attribute groups. Here, surface properties contributed more extensively to identity decoding compared with shape, consistent with our temporally cumulative findings. Interestingly, reliable decoding for combined

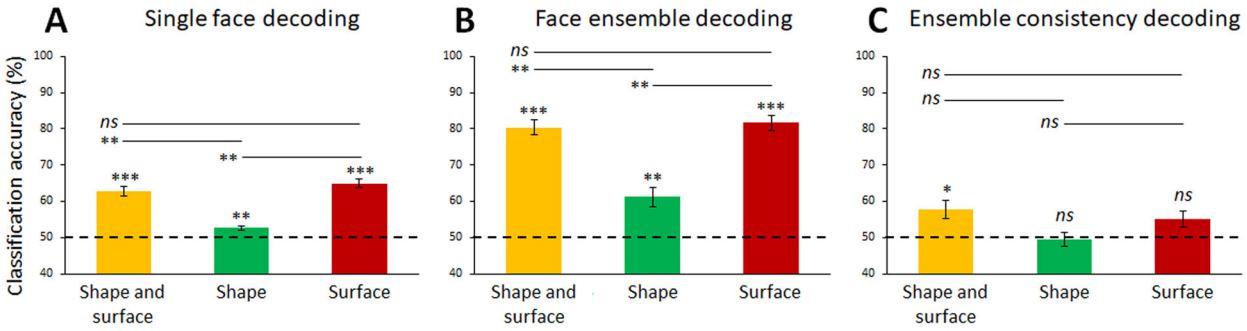


Figure 6. Temporally cumulative pairwise decoding of (***A***) single faces, (***B***) face ensembles, and (***C***) consistent versus inconsistent face ensembles. Significant decoding was observed in all cases except for decoding of only shape or only surface for consistency. Pairwise differences were significant for comparisons of shape versus surface (***A,B***), and there were no differences between shape and surface versus just surface. ns, not significant, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; Bonferroni-corrected; error bars show ± 1 *SE*.

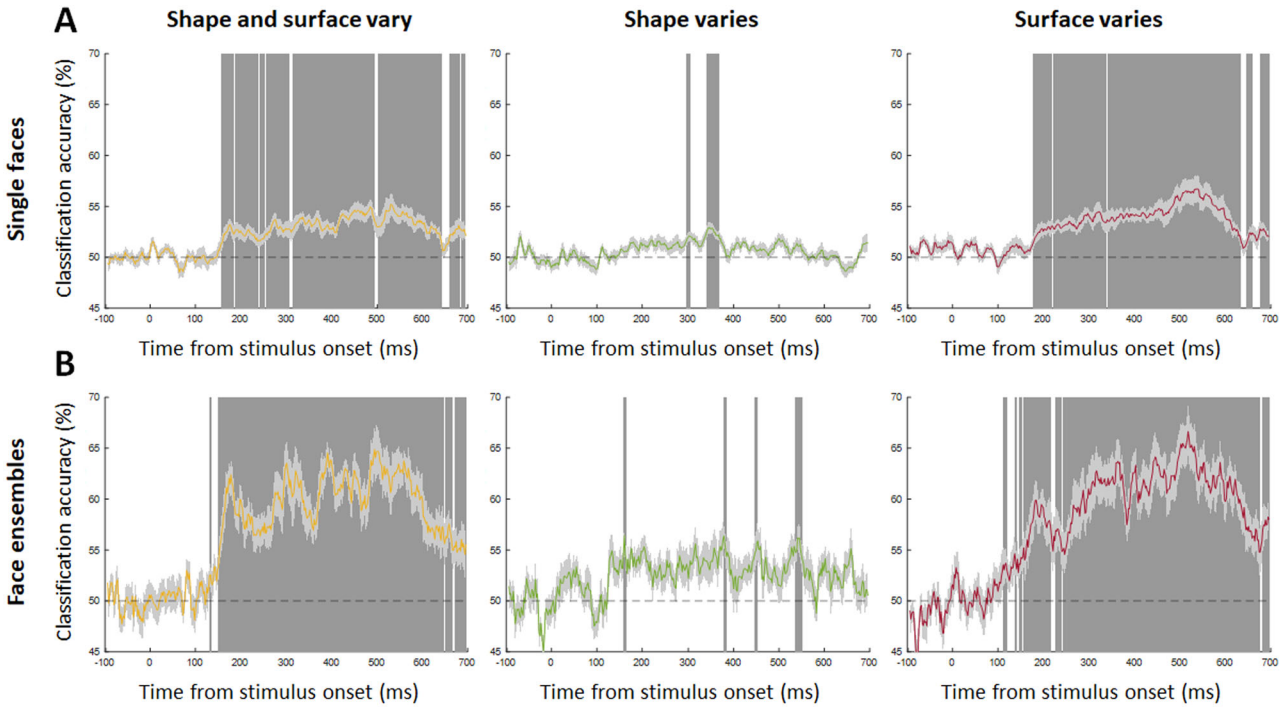


Figure 7. The time course of pairwise decoding across attribute groups for (***A***) single faces and (***B***) face ensembles. Consistent with the temporally cumulative results (Fig. 6), surface information, by itself or combined with shape, contributed more extensively to decoding compared with shape by itself. Each position on the time axis represents accuracy at that given time point centered on a 10 ms window. Accuracy across time bins is compared against 50% (hashed line) via sign test. Resulting p values are FDR-corrected ($q < 0.05$), and significant time bins are denoted by dark gray shading; light-shaded gray indicates ± 1 *SE*

shape and surface began 30 ms earlier than for surface alone, highlighting the joint contribution of shape and surface properties in face identity processing (Dzhelyova and Rossion, 2014).

In sum, we show reliable decoding of single-face identity from both shape and surface properties, with the latter serving as the dominant property. Importantly, these results are in broad agreement with previous work on single-face identity decoding (Nemrodov et al., 2016, 2019a,b; Dobs et al., 2019; Smith and Smith, 2019; Bae, 2020) and serve to ground our investigation of face ensembles.

Face ensembles

Overall, we found significant differences in temporally cumulative decoding accuracy for face ensembles ($\chi^2_{(2)} = 14.46$, $p < 0.001$, $W = 0.452$, $BF_{10} = 680{,}885.92$). Decoding accuracy was significant for all three attribute groups (all accuracy estimates >61.2 ± 2.1%, all zs > 3.05, all ps < 0.003, all RCs > 0.868, all BF_{10} > 88.02; Fig. 6*B*).

As observed with single faces, the surface, both by itself or when combined with shape, contributed more to ensemble decoding accuracy compared with shape (both mean difference >19.2 ± 1.3%, both zs > 3.05, both ps < 0.002, both RC > 0.868, both $BF_{10} > 157.09$), and there was no significant difference between combined shape and surface versus just surface (mean difference = 1.2 ± 1.6%, $z = 0.73$, $p > 0.999$, $BF_{10} = 0.33$).

The time course of ensemble decoding (Fig. 7*B*) revealed a similar pattern to that of single faces. Notably, the onset of ensemble decoding began around ~150 ms and tapered off after 550 ms. Moreover, surface, by itself or when combined with shape, contributed more extensively to ensemble decoding compared with shape by itself.

Next, we found significant decoding of consistent versus inconsistent ensembles (accuracy = 54.1 ± 1.6%, $z = 2.33$, $p = 0.010$, RC = 0.662, $BF_{10} = 6.36$) and an overall difference in accuracy across the three attribute groups ($\chi^2_{(2)} = 8.79$, $p = 0.012$, $W = 0.275$, $BF_{10} = 3.86$), which, interestingly, was driven by combined shape and surface properties (shape and surface accuracy = 57.8 ± 2.5%, $z = 2.53$, $p = 0.017$, RC = 0.721, $BF_{10} = 14.27$; shape accuracy = 49.5 ± 2.0%, $z = -0.41$, $p > 0.999$, $BF_{10} = 0.22$; surface accuracy = 55.0 ± 2.2%, $z = 1.86$, $p = 0.094$, $BF_{10} = 3.26$; Fig. 6*C*). The difference between consistency decoding for shape and surface versus just shape approached significance (mean difference = 8.3 ± 3.0%, $z = 2.38$, $p = 0.052$, $BF_{10} = 3.80$; both remaining comparisons, mean difference <5.5 ± 2.8%, both zs < |1.73|, both ps > 0.249, both $BF_{10} < 1.15$).

To shed more light on the above results, we assessed the neural dynamics of surround and central face processing separately. To this aim, during classification, stimuli were grouped by controlling for center and surround information, respectively (see Table 1 for examples). Both surrounds (accuracy = 57.0 ± 1.3%, $z = 3.10$, $p = 0.002$, RC = 0.882, $BF_{10} = 673.14$) and central faces (accuracy = 80.1 ± 1.3%, $z = 3.52$, $p < 0.001$, RC = 1.000, $BF_{10} = 3.60 \times 10^{10}$) yielded significant decoding accuracy across attribute groups (Fig. 8*A*), with central faces yielding higher accuracy (mean difference = 23.1 ± 1.9%, $z = 3.52$, $p < 0.001$, RC = 1.000, $BF_{10} = 3.40 \times 10^{6}$).

Further, we found an overall difference in decoding across attribute groups both for surrounds ($\chi^2_{(2)} = 8.81$, $p = 0.012$, $W = 0.275$, $BF_{10} = 2.02$) and for central faces ($\chi^2_{(2)} = 14.51$, $p < 0.001$, $W = 0.453$, $BF_{10} = 1.30 \times 10^{6}$). Surround decoding was driven by combined shape and surface properties (shape and surface accuracy = 60.9 ± 2.2%, $z = 3.00$, $p = 0.004$, RC = 0.853, $BF_{10} = 341.70$; shape accuracy = 52.7 ± 2.4%, $z = 0.78$, $p = 0.657$, $BF_{10} = 0.79$; surface accuracy = 57.4 ± 2.4%, $z = 2.53$, $p = 0.017$, RC = 0.721, $BF_{10} = 13.09$; Fig. 8*B*), as the highest level of decoding accuracy was found for combined shape and surface information (shape vs

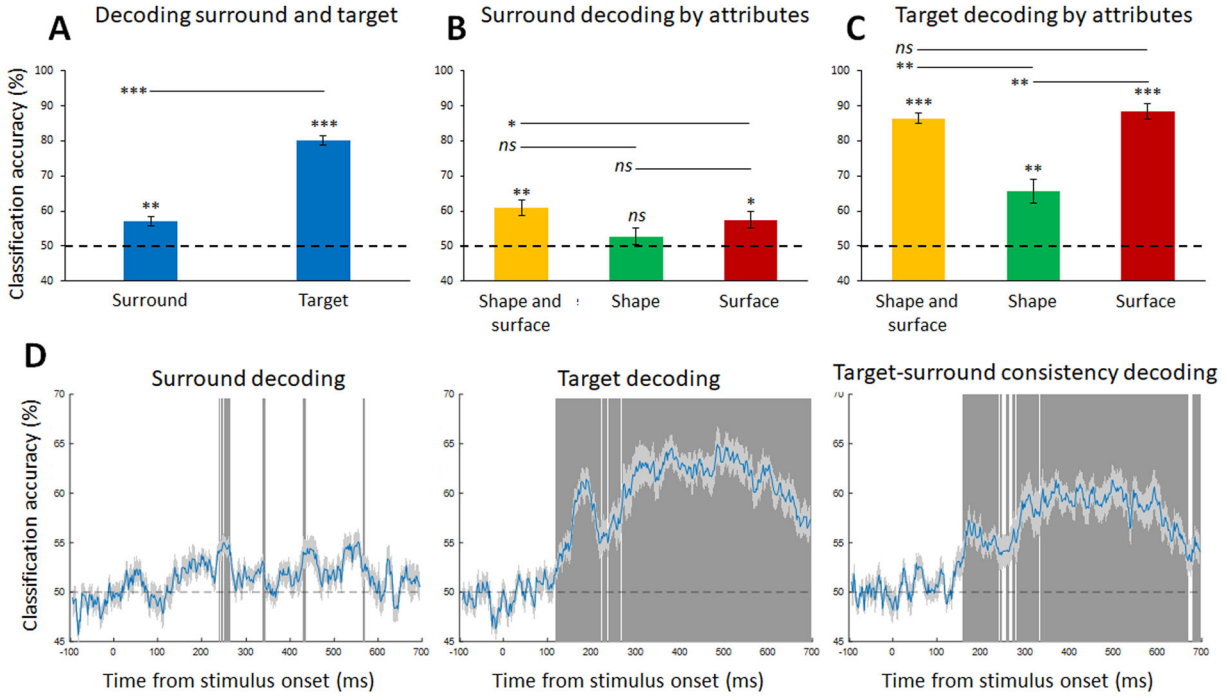


Figure 8. Separate decoding of surround and center faces. ***A***, Surround and center faces are decoded significantly above chance (collapsed across all three attribute conditions). ***B***, Surround decoding is maximized for surface and shape properties combined. ***C***, Central face decoding is driven by both shape and surface properties, with a higher contribution from surface information. ***D***, The time course of decoding surround faces, center faces, and center-surround consistency is shown collapsed across attribute conditions. Both the ensemble surround and center, as well as their consistency, show intervals of above-chance classification. ns, not significant, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; Bonferroni-corrected; error bars show ± 1 *SE*.

surface, mean difference = 4.7 ± 3.9%, $z = 1.42$, $p = 0.467$, $BF_{10} = 0.47$; shape and surface vs shape by itself, mean difference = 8.2 ± 2.5%, $z = 2.76$, $p = 0.018$, RC = 0.808, $BF_{10} = 8.65$; shape and surface vs surface by itself, mean difference = 3.5 ± 3.3%, $z = 0.70$, $p > 0.999$, $BF_{10} = 0.41$).

Next, significant central face decoding occurred for both shape and surface properties (shape and surface combined accuracy = 86.4 ± 1.5%, $z = 3.52$, $p < 0.001$, RC = 1.000, $BF_{10} = 3.38 \times 10^{10}$; shape accuracy = 65.6 ± 3.4%, $z = 3.10$, $p = 0.003$, RC = 0.882, $BF_{10} = 181.11$; surface accuracy = 88.4 ± 2.1%, $z = 3.52$, $p < 0.001$, RC = 1.000, $BF_{10} = 1.21 \times 10^{9}$; Fig. 8*C*). Again, surface properties made the greatest contribution to decoding accuracy (shape vs surface, mean difference = 22.8 ± 4.5%, $z = 3.21$, $p = 0.004$, RC = 0.912, $BF_{10} = 215.27$; shape and surface vs shape by itself, mean difference = 20.8 ± 4.3%, $z = 3.21$, $p = 0.004$, RC = 0.912, $BF_{10} = 158.18$; shape and surface vs surface by itself, mean difference = 2.0 ± 1.6%, $z = 1.31$, $p = 0.575$, $BF_{10} = 2.73$).

Further, we observed distinct temporal profiles for decoding surround versus central faces. The onset of significant classification for surrounds began at around 250 ms and ended at around 600 ms (Fig. 8*D*, left panel). Significant decoding of central faces occurred earlier, starting around 120 ms, and continued for the duration of the epoch (Fig. 8*D*, center panel). Center-surround consistency exhibited significant classification from 150 ms to around 680 ms (Fig. 8*D*, right panel).

Of note, the temporal profile of the surround faces aligns with previous work on ensemble face decoding using only surround faces (Roberts et al., 2019), with an initial peak at 250 ms, followed by other peaks at 300 ms and 400 ms. At the same time, the profile of central faces is comparable to that observed for the classification of single faces (Fig. 7*A*) as well as to that from previous research on single-face processing (Nemrodov et al., 2016, 2018; Roberts et al., 2019). Hence, it appears that the surround and the center of an ensemble are encoded differently, without impacting the time course of each other's processing.

The time course of information processing for single faces versus face ensembles

For a detailed comparison of the neural dynamics of information processing between single and ensemble faces, we grouped selected time course profiles by type of decoding and stimulus attributes (Fig. 9). Visual comparisons are complemented via Pearson correlation between time course data. Here, we applied Fisher z transformation to r values computed separately for each participant and performed a one-sample t test comparing mean z values ($\bar{z}$) to zero. The temporal window was restricted between 100 ms and 600 ms post-stimulus onset, where most classification results occurred—this also prevents inflating the correlation coefficient when including near-chance accuracy values outside this interval. The resulting p values are Bonferroni-corrected. This procedure highlights four notable findings and offers a clearer visualization of our time course results.

First, decoding profiles for single and ensemble central faces are quite similar (Fig. 9, yellow band for single faces vs purple band for the latter) and are significantly correlated (mean $r = 0.28$, $\bar{z} = 0.31 \pm 0.06$, $t_{(15)} = 4.91$, $p < 0.001$, $d = 1.23$, $BF_{10} = 164.40$), which points to comparable processing dynamics for ensemble central faces and single faces.

Second, the time course of surface-based decoding is similar for single faces and face ensembles (Fig. 9, red bands; mean $r = 0.34$, $\bar{z} = 0.37 \pm 0.06$, $t_{(15)} = 5.94$, $p < 0.001$, $d = 1.48$, $BF_{10} = 899.44$). This was not found with shape (green bands, mean $r = 0.02$, $\bar{z} = 0.02 \pm 0.04$, $t_{(15)} = 0.46$, $p > 0.999$, $BF_{10} = 0.28$), which, interestingly, contributed more to the processing of ensembles compared with single faces.

Third, the time course of shape-based decoding differs substantially for ensembles compared with single faces, as reflected by an earlier onset and longer duration of significant decoding for face ensembles. One caveat in this respect though concerns the smaller number of classification patterns for single face versus ensemble decoding (i.e., 12 vs 16). This, along with the less prominent contribution of shape-to-face processing, may have obscured the precise involvement of shape in single-face processing and the comparison with its counterpart in ensemble processing.

Fourth, the time courses of surround and center decoding were not correlated (mean $r = 0.01$, $\bar{z} = 0.01 \pm 0.05$, $t_{(15)} = 0.18$, $p > 0.999$, $BF_{10} = 0.26$), and the time course of the surround was also not correlated with that of single faces (mean $r = 0.06$, $\bar{z} = 0.07 \pm 0.05$, $t_{(15)} = 1.31$, $p > 0.999$, $BF_{10} = 0.53$).

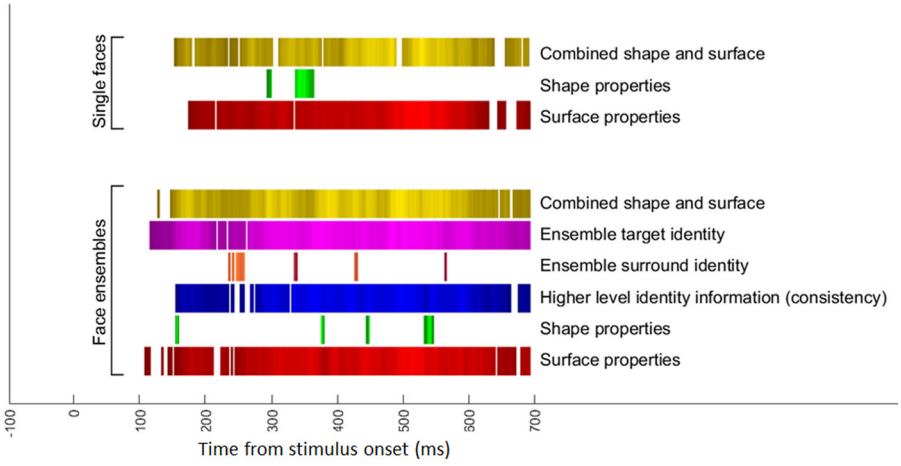


Figure 9. A comparison of the temporal profiles for ensemble and single-face processing (relying on information from Figs. 7, 8*D*). Colored bands indicate significant temporal intervals of decoding. Brightness relates to decoding accuracy (brighter colors indicate higher accuracy) at a given time point, standardized to aid visual comparison. The temporal profile of surface information is similar for both single and ensemble faces, whereas that of shape vastly differs, with an earlier decoding onset and longer duration for ensembles compared with single faces. The profiles for single and ensemble center faces are also highly similar. Relating center and surround information (i.e., consistency) occurs across the entire epoch.

Discussion

Our study seeks to elucidate the dynamics of face ensemble encoding, addressing several important aspects of their neural processing. To this end, we rely on stimuli varying in facial identity separately with respect to shape and surface properties, matched across low-level image differences and presented in a novel center-surround ensemble paradigm. In short, with the aid of our stimulus design and multivariate pattern analysis, we show distinct neural profiles for the processing of single faces versus face ensembles, different contributions of shape and surface properties, and parallel extraction of center and surround face information from ensembles.

Our univariate results point to an earlier N170 onset for face ensembles, consistent with previous research (Puce et al., 2013; Roberts et al., 2019). However, the general lack of univariate differences, such as that reflecting ensemble center-surround consistency, underscored the importance of using more sensitive techniques.

Accordingly, multivariate analyses revealed a more extensive decoding profile for the surface than shape in the case of both single and ensemble faces, in agreement with prior work arguing for a more dominant role of surface information (Bruce and Langton, 1994; Hole et al., 2002; Russell et al., 2006; Kaufmann and Schweinberger, 2008; Parr and Taubert, 2010; Nakajima et al., 2012; Dzhelyova and Rossion, 2014; Nemrodov et al., 2019a; Rogers et al., 2022). The finding that the combination of shape and surface properties provided little benefit compared to surface cues alone, but decoding based on shape alone was still significant, may indicate that the visual system can recruit both properties for single face and ensemble processing but relies predominantly on surface properties. This supports prior research that describes ensemble encoding as a form of texture perception (Dakin and Watt, 1997; Parkes et al., 2001; Morgan et al., 2008, 2014; Im and Halberda, 2012; Whitney et al., 2014).

Interestingly, surface and shape cues may contribute differentially to the encoding of face ensembles. For example, both properties were utilized when decoding ensemble central faces, whereas only surface cues were utilized when decoding surrounds. Further, consistency decoding only reached significance when shape and surface information were combined.

Given the robust decoding of surface information, these results suggest that, similar to single faces, ensembles are processed primarily in terms of surface information. Shape, however, appears to play a more complex role. Shape information is extracted for both single and ensemble faces, though the profile is markedly different across the two types of stimuli. Shape may play a more general role in face ensemble encoding, involving the entire stimulus, and not just one aspect of it (i.e., not just the central face). However, our analyses did not find clear evidence for the implication of shape in surround or consistency decoding (beyond a small advantage to combining shape and surface relative to the only surface in surround decoding).

For surface properties, it appears that both ensemble and single faces rely on similar processing mechanisms. It is possible that facial surface properties are processed similarly for ensemble and single items but in separate cortical regions. For single faces, surface cues are likely encoded in the cortical face network (Grill-Spector et al., 2017). For ensembles, surface cues may be processed in regions implicated in general ensemble encoding. One candidate region includes the parahippocampal gyrus, which responds to variations in the surface properties of object ensembles (Cant and Xu, 2012, 2017). Recent behavioral evidence suggests a domain-general ability involved in complex object ensemble encoding (Chang and Gauthier, 2022). Whether this generalizes to face ensembles and whether this is subserved by the parahippocampal gyrus are questions that await further investigation.

Next, we assessed the dynamics of information processing for face ensembles. Central face identity is decoded early, starting around 120 ms, while surround information is processed next, from 200 ms. Center-surround consistency decoding occurs during most of the temporal epoch. Thus, a relatively straightforward progression of ensemble processing involves the initial encoding of a target (in this case the central face), followed by the surround, with the consistency of the two being simultaneously compared. The underlying mechanism driving consistency decoding is unclear and remains subject to future research. One possibility is that it reflects a neural signature corresponding to outlier detection. However, if that were the case, one may expect differences between consistent and inconsistent ensembles in our univariate results, given the strong neural response to outliers in ensemble processing (Cant and Xu, 2020).

Of note, we sought to study single-face processing within the context of face ensembles, through the development of a novel center-surround paradigm. This differs from typical ensemble studies, which often involve set membership judgments of a single probe following an ensemble display. These paradigms have effectively shown that the perception of single faces from the ensemble is biased toward the ensemble average (de Fockert and Wolfenstein, 2009; Haberman and Whitney, 2009), which undermines exemplar representation in ensemble encoding. However, participants in these studies often do not fixate on any particular face in the ensemble. Here, we asked how single faces at the center of fixation are represented within the broader context of an ensemble (e.g., as one would view another person during a face-to-face conversation on a crowded subway). Our paradigm involves foveating a central face, mimicking single-face judgment tasks, while, simultaneously, presenting this face as a member of a larger ensemble. Furthermore, the central face was rendered relevant to the behavioral task, which served to maintain participants' attention. Arguably, this provides an opportunity to bridge ensemble and single-face processing and to explore the role of attention in this process. For example, distributed versus focused attention toward targeted faces may help elucidate the role of attention in ensemble processing, a topic of ongoing debate (Jackson-Nielsen et al., 2017; Chen et al., 2020).

Our results indicate that central ensemble faces are processed similarly to single faces, and, importantly, that the ensemble surround can be captured in the absence (or minimal deployment) of direct attention, given our experimental task. Moreover, the temporal processing profiles for center and surround faces were uncorrelated, suggesting relatively independent processing of center ensemble (and single faces) versus surround ensemble faces. These findings parallel what has been found for single objects versus object ensembles (Cant et al., 2015), as well as research that argues for two separate modes of visual processing for ensemble perception versus focused attention (Baek and Chong, 2020). Indeed, our surround decoding results are similar to previous ensemble decoding research which controlled for attentional deployment (Roberts et al., 2019). Arguably, these results demonstrate a novel and more ecologically valid way to explore the link between single and ensemble face perception and demonstrate the utility of center-surround paradigms as a

way to compare the dynamics between single and ensemble processing for both faces and non-face objects.

At the same time, we note that the "target" face in our study was always centrally located, hence, comparing the target face to the surround also entails a comparison of central versus peripheral vision. To disentangle these aspects of our design, future work can vary the spatial location of the target face within the ensemble (e.g., by randomly choosing one of the six surround faces from the periphery).

Further, we note that our current results are based on ERP patterns across bilateral OT electrodes. Our approach is motivated by the overall decoding advantage of this electrode subset over that based on the entire set and, also, by the lack of significant differences for laterality. However, future work is needed to clarify the extent to which different groups of electrodes carry relevant information (e.g., about shape vs surface information or center vs surround faces). EEG-based spatiotemporal searchlight analysis (Dalski et al., 2023) and/or feature selection (Nemrodov et al., 2019b) may provide valuable insights in this sense.

In conclusion, here, we establish the contribution of shape and surface properties in face ensemble processing, and we reveal both similarities and differences relative to single-face processing. Specifically, we show a different role of shape information and a global involvement of surface information in ensemble processing relative to single-face processing. Moreover, we demonstrate the utility of a novel center-surround paradigm developed to complement traditional exemplar-cueing paradigms in the study of ensemble perception. Together, these results serve to elucidate the neural dynamics of ensemble face perception, the visual information underlying this process, and its relationship with single-face processing.

References

Alvarez GA (2011) Representing multiple objects as an ensemble enhances visual cognition. Trends Cogn Sci 15:122–131.

Andrews TJ, Baseler H, Jenkins R, Burton MA, Young AW (2016) Contributions of feature shapes and surface cues to the recognition and neural representation of facial identity. Cortex 83:280–291.

Bae G-Y (2020) The time course of face representations during perception and working memory maintenance. Cereb Cortex Commun 2:tgaa093.

Baek J, Chong SC (2020) Ensemble perception and focused attention: two different modes of visual processing to cope with limited capacity. Psychon Bull Rev 27:602–606.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436.

Bruce V, Langton S (1994) The use of pigmentation and shading information in recognising the sex and identities of faces. Perception 23:803–822.

Bruce V, Young A (1998) *In the eye of the beholder: the science of face perception*. Oxford University Press.

Bruce V, Burton MA, Hanna E, Healey P, Mason O (1993) Sex discrimination: how well do we tell the difference between male and female faces? Perception 22:131–152.

Burton MA, Scheinberger SR, Jenkins R, Kaufmann JM (2015) Arguments against a configural processing account of familiar face recognition. Perspect Psychol Sci 10:482–496.

Cant JS, Xu Y (2012) Object ensemble processing in human anterior-medial ventral visual cortex. J Neurosci 32:7685–7700.

Cant JS, Xu Y (2017) The contribution of object shape and surface properties to object ensemble representation in anterior-medial ventral visual cortex. J Cogn Neurosci 29:398–412.

Cant JS, Xu Y (2020) One bad apple spoils the whole bushel: the neural basis of outlier processing. NeuroImage 211:116629.

Cant JS, Sun SZ, Xu Y (2015) Distinct cognitive mechanisms involved in the processing of single objects and object ensembles. J Vis 15:12.

Cha O, Blake R, Gauthier I (2020) The role of category- and exemplar-specific experience in ensemble processing of objects. Atten Percept Psychophys 83:1080–1093.

Chang T-Y, Gauthier I (2022) Domain-general ability underlies complex object ensemble processing. J Exp Psychol Gen 151:966–972.

Chen Z, Ran Z, Xiaolin W, Ren Y, Abrams RA (2020) Ensemble perception without attention depends upon attentional control settings. Atten Percept Psychophys 83:1240–1250.

Corbett JE, Utochkin I, Hochstein S (2023) *The pervasiveness of ensemble perception: not just your average review*. Cambridge University Press.

Cowan N (2010) The magical mystery four: how is working memory capacity limited, and why? Curr Dir Psychol Sci 19:51–57.

Dakin SC, Watt RJ (1997) The computation of orientation statistics from visual texture. Vis Res 37:3181–3192.

Dalski A, Kovács G, Ambrus GG (2023) No semantic information is necessary to evoke general neural signatures of face familiarity: evidence from cross-experiment classification. Brain Struct Funct 228:449–462.

de Fockert JW, Wolfenstein C (2009) Rapid extraction of mean identity from sets of faces. Q J Exp Psychol 62:1716–1722.

Delorme A, Sejnowski T, Makeig S (2007) Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. NeuroImage 34:1443–1449.

Dobs K, Isik L, Pantazis D, Kanwisher N (2019) How face perception unfolds over time. Nat Commun 10:1258.

Duchaine BC, 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 44:576–585.

Duchaine BC, Yovel G (2015) A revised neural framework for face processing. Annu Rev Vis Sci 1:393–416.

Dzhelyova M, Rossion B (2014) Supra-additive contribution of shape and surface information to individual face discrimination as revealed by fast periodic visual stimulation. J Vis 14:15.

Ghuman AS, Brunet NM, Konecky RO, Pyles JA, Walls SA, Destefino V, Richardson MR (2014) Dynamic encoding of face information in the human fusiform gyrus. Nat Commun 5:5672.

Grill-Spector K, Weiner KS, Kay K, Gomez J (2017) The functional neuroanatomy of human face perception. Annu Rev Vis Sci 15:167–196.

Haberman J, Whitney D (2009) Seeing the men: ensemble coding for sets of faces. J Exp Psychol Hum Percept Perform 35:718–734.

Haberman J, Harp T, Whitney D (2009) Averaging facial expression over time. J Vis 9.

Halit H, de Haan M, Johnson MH (2000) Modulation of event-related potentials by prototypical and atypical faces. Neuroreport 11:1871–1875.

Hole GJ, Goerge PA, Eaves K, Rasek A (2002) Effects of geometric distortions on face-recognition performance. Perception 31:1221–1240.

Im HY, Halberda J (2012) The effects of sampling and internal noise on the representation of ensemble average size. Atten Percept Psychophys 75: 278–286.

Jackson-Nielsen M, Cohen MA, Pitts MA (2017) Perception of ensemble statistics requires attention. Conscious Cogn 48:149–160.

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.

Ji L, Pourtois G, Sweeny TD (2020) Averaging multiple facial expressions through subsampling. Vis Cogn 28:41–58.

Jiang F, Dricot L, Blanz V, Goebel R, Rossion B (2009) Neural correlates of shape and surface reflectance information in individual faces. Neuroscience 163:1078–1091.

Kaufmann JM, Schweinberger SR (2008) Distortions in the brain? ERP effects of caricaturing familiar and unfamiliar faces. Brain Res 1228:177–188.

Kramer RSS, Jenkins R, Burton MA (2017) Interface: a software package for face image warping, averaging, and principal component analysis. Behav Res Methods 49:2002–2011.

Kuefner D, de Heering A, Jacques C, Palmero-Soler E, Rossion B (2010) Early visually evoked electrophysiological responses over the human brain (P1, N170) show stable patterns of face-sensitivity from 4 years to adulthood. Front Hum Neurosci 3:67.

Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, Knippenberg AV (2010) Presentation and validation of the radboud faces database. Cogn Emot 24:1377–1388.

Little AC, Jones BC, DeBruine LM (2011) The many faces of research on face perception. Philos Trans R Soc Lond, B, Biol Sci 366:1634–1637.

Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. Nature 390:279–281.

Ly A, Verhagen J, Wagenmakers E-J (2016) Harold Jeffreys's default Bayes factor hypothesis tests: explanation, extension, and application in psychology. J Math Psychol 72:19–32.
Mercure E, Dick F, Johnson MH (2008) Featural and configural face processing differentially modulate ERP components. Brain Res 1239:162–170.
Morgan M, Chubb C, Solomon JS (2008) A 'dipper' function for texture discrimination based on orientation variance. J Vis 8.
Morgan MJ, Raphael S, Tibber MS, Dakin SC (2014) A texture-processing model of the 'visual sense of number'. Proc R Soc B Biol Sci 281: 20141137.
Mouraux A, Iannetti GD (2008) Across-trial averaging of event-related EEG responses and beyond. Magn Reson Imaging 26:1041–1054.
Muukkonen I, Ölander K, MNumminen J, Salmela VR (2020) Spatio-temporal dynamics of face perception. NeuroImage 209:116531.
Nakajima K, Minami T, Nakauchi S (2012) The face-selective N170 component is modulated by facial color. Neuropsychologia 50:2499–2505.
Nemrodov D, Niemeier M, Mok JNY, Nestor A (2016) The time course of individual face recognition: a pattern analysis of ERP signals. NeuroImage 132:469–476.
Nemrodov D, Niemeier M, Patel A, Nestor A (2018) The neural dynamics of facial identity processing: insights from EEG-based pattern analysis and image reconstruction. eNeuro 5.
Nemrodov D, Behrmann M, Niemeier M, Drobotenko N, Nestor A (2019a) Multimodal evidence on shape and surface information in individual face processing. NeuroImage 184:813–825.
Nemrodov D, Ling S, Nudnou I, Roberts T, Cant JS, Lee ACH, Nestor A (2019b) A multivariate investigation of visual word, face, and ensemble processing: perspectives from EEG-based decoding and feature selection. Psychophysiology 57:e13511.
Nestor A, Tarr MJ (2008) The segmental structure of faces and its use in gender recognition. J Vis 8:12.
O'Toole AJ, Vetter T, Blanz V (1999) Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: an application of three-dimensional morphing. Vis Res 39:3145–3155.
Parkes L, Lund J, Angelucci A, Solomon JA, Morgan M (2001) Compulsory averaging of crowded orientation signals in human vision. Nat Neurosci 4:739–744.
Parr LA, Taubert J (2010) The importance of surface-based cues for face discrimination in non-human primates. Proc R Soc B Biol Sci 278:1964–1972.
Piepers DW, Robbins RA (2012) A review and clarification of the terms "holistic," "configural," and "relational" in the face perception literature. Front Psychol 3:559.
Puce A, McNeely ME, Berrebi ME, Thompson JC, Hardee J, Brefczynski-Lewis J (2013) Multiple faces elicit augmented neural activity. Front Hum Neurosci 14:282.
Raffone A, Wolters G (2001) A cortical mechanism for binding in visual working memory. J Cogn Neurosci 13:766–785.
Roberts T, Cant JS, Nestor A (2019) Elucidating the neural representation and the processing dynamics of face ensembles. J Neurosci 39:7737–7747.
Rogers D, Baseler H, Young AW, Jenkins R, Andrews TJ (2022) The roles of shape and texture in the recognition of familiar faces. Vis Res 194:108013.
Rouder JN, Morey RD, Speckman PL, Province JM (2012) Default Bayes factors for ANOVA designs. J Math Psychol 56:356–374.
Russell R, Sinha P, Biederman I, Nederhouser M (2006) Is pigmentation important for face recognition? Evidence from contrast negation. Perception 35:749–759.
Russell R, Biederman I, Nederhouser M, Sinha P (2007) The utility of surface reflectance for the recognition of upright and inverted faces. Vis Res 47: 157–165.
Russell R, Chatterjee G, Nakayama K (2012) Developmental prosopagnosia and super-recognition: no special role for surface reflectance processing. Neuropsychologia 50:334–340.
Smith FW, Smith ML (2019) Decoding the dynamic representation of facial expressions of emotion in explicit and incidental tasks. NeuroImage 195:261–271.
Tanaka H (2018) Face-sensitive P1 and N170 components are related to the perception of two-dimensional and three-dimensional objects. Neuroreport 29:583–587.
Tsao DY, Livingstone MS (2008) Mechanisms of face perception. Annu Rev Neurosci 31:411–437.
Vida MD, Nestor A, Plaut DC, Behrmann M (2016) Spatiotemporal dynamics of similarity-based neural representations of facial identity. Proc Natl Acad Sci U S A 114:388–393.
Wagenmakers E-J, Wetzels R, Borsboom D, van der Maas HLJ (2011) Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). J Pers Soc Psychol 100:426–432.
Wang H, Sun P, Ip C, Zhao X, Fu S (2015) Configural and featural face processing are differently modulated by attentional resources at early stages: an event-related potential study with rapid serial visual presentation. Brain Res 1602:75–84.
Westfall PH, Johnson WO, Utts JM (1997) A Bayesian perspective of the Bonferroni adjustment. Biometrika 84:419–427.
Whitney D, Yamanashi Leib A (2018) Ensemble perception. Annu Rev Psychol 69:105–129.
Whitney D, Haberman J, Sweeny TD (2014) From textures to crowds: multiple levels of summary statistical perception. In: *The new visual neurosciences* (Werner JS, Chalupa LM, eds), pp. 695–710. MIT Press.