Feature-based face representations and image reconstruction from behavioral and neural data

Adrian Nestor[a,1], David C. Plaut[b,c], and Marlene Behrmann[b,c,1]

[a]Department of Psychology at Scarborough, University of Toronto, Toronto, ON, M1C 1A4, Canada; [b]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213-3890; and [c]Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213-3890

Contributed by Marlene Behrmann, November 12, 2015 (sent for review October 4, 2015; reviewed by Garrison W. Cottrell and Pawan Sinha)

The reconstruction of images from neural data can provide a unique window into the content of human perceptual representations. Although recent efforts have established the viability of this enterprise using functional magnetic resonance imaging (MRI) patterns, these efforts have relied on a variety of prespecified image features. Here, we take on the twofold task of deriving features directly from empirical data and of using these features for facial image reconstruction. First, we use a method akin to reverse correlation to derive visual features from functional MRI patterns elicited by a large set of homogeneous face exemplars. Then, we combine these features to reconstruct novel face images from the corresponding neural patterns. This approach allows us to estimate collections of features associated with different cortical areas as well as to successfully match image reconstructions to corresponding face exemplars. Furthermore, we establish the robustness and the utility of this approach by reconstructing images from patterns of behavioral data. From a theoretical perspective, the current results provide key insights into the nature of high-level visual representations, and from a practical perspective, these findings make possible a broad range of image-reconstruction applications via a straightforward methodological approach.

image reconstruction | face space | reverse correlation

Face recognition relies on visual representations sufficiently complex to distinguish even among highly similar individuals despite considerable variation due to expression, lighting, viewpoint, and so forth. A longstanding conceptual framework, termed "face space" (1–6), suggests that individual faces are represented in terms of their multidimensional deviation from an "average" face, but the precise nature of the dimensions or features that capture these deviations, and the degree to which they preserve visual detail, remain unclear. Thus, the featural basis of face space along with the neural system that instantiate it remain to be fully elucidated. The present investigation aims not only to uncover fundamental aspects of neural representations but also to establish their plausibility and utility through image reconstruction. Concretely, the current study addresses the issues above in the context of two distinct challenges, first, by determining the visual features used in face identification and, second, by validating these features through their use in facial image reconstruction.

With respect to the first challenge, recent studies have demonstrated distinct sensitivity to local features (e.g., the size of the mouth) compared with configural features (e.g., the distance between the eyes and the mouth) in human face-selective cortex (7–10). Also, neurophysiological investigations (1, 11) of monkey cortex have found sensitivity to several facial features, particularly in the eye region of the face. However, most investigations consider only a few handpicked features. Thus, a comprehensive, unbiased assessment of face space still remains to be conducted. Furthermore, most studies target shape at the expense of surface features (e.g., skin tone) despite the relevance of the latter for recognition (12, 13).

With respect to the second challenge, a number of studies have taken steps toward image reconstruction from functional magnetic resonance imaging (fMRI) signals in visual cortex, primarily exploiting low-level visual features (14–16; but see ref. 17). The recent extension of this work to the reconstruction of face images (18) has demonstrated the promise of exploiting category-specific features (e.g., facial features) associated with activation in higher visual cortex. However, the substantial variability across individual faces in this latter study (due to race, age, image background, etc.) limits its conclusions with regard to facial identification and the representations underlying it. Moreover, this attempt deployed prespecified image features due to their reconstruction potential rather than as an argument for their biological plausibility.

The current work addresses the challenges above by adopting a broad, unbiased methodological approach. First, we map cortical areas that exhibit separable patterns of activation to different facial identities. We then construct confusability matrices from behavioral and neural data in these areas to determine the general organization of face space. Next, we extract the visual features accounting for this structure by means of a procedure akin to reverse correlation. And last, we deploy the very same features for the purpose of face reconstruction. Importantly, our approach relies on an extensive but carefully controlled stimulus set ensuring our focus on fine-grained face identification.

The results of our investigation show that (*i*) a range of facial properties such as eyebrow salience and skin tone govern face encoding, (*ii*) the broad organization of behavioral face space reflects that of its neural homolog, and (*iii*) high-level face representations retain sufficient detail to support reconstructing the visual appearance of different facial identities from either neural or behavioral data.

Significance

The present work establishes a novel approach to the study of visual representations. This approach allows us to estimate the structure of human face space as encoded by high-level visual cortex, to extract image-based facial features from this structure, and to use such features for the purpose of facial image reconstruction. The derivation of visual features from empirical data provides an important step in elucidating the nature and the specific content of face representations. Further, the integrative character of this work sheds new light on the existing concept of face space by rendering it instrumental in image reconstruction. Last, the robustness and generality of our reconstruction approach is established by its ability to handle both neuroimaging and psychophysical data.

Author contributions: A.N., D.C.P., and M.B. designed research; A.N. and M.B. performed research; A.N. analyzed data; and A.N., D.C.P., and M.B. wrote the paper.

Reviewers: G.W.C., University of California, San Diego; and P.S., Massachusetts Institute of Technology.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: behrmann@cmu.edu or anestor@utsc.utoronto.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1514551112/-/DCSupplemental.

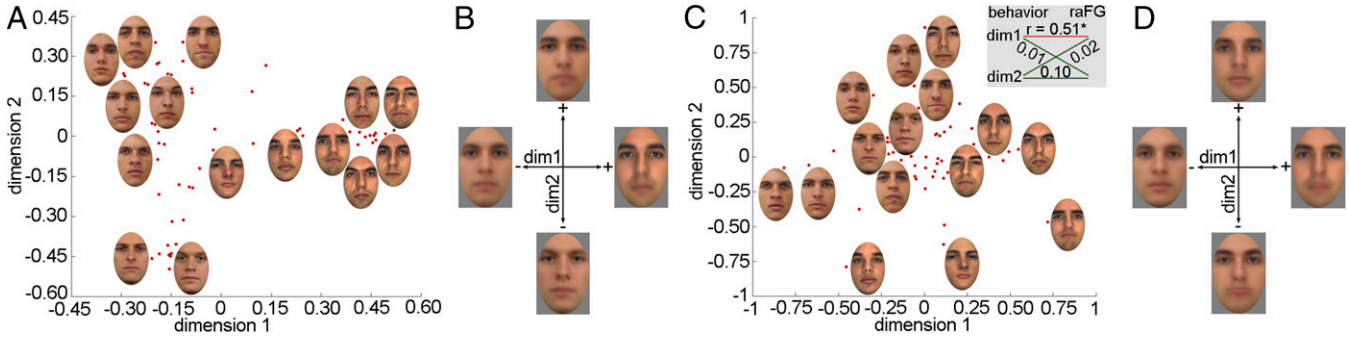

Fig. 1. Behavioral and neural face space topography estimated through MDS. Plots show the distribution of facial identities across the first two dimensions for (*A*) behavioral and (*C*) right anterior fusiform gyrus (raFG) data. Each dot represents a single identity (for simplicity only a subset of neutral images is shown in each plot). First-dimension coefficients corresponding to different facial identities correlate significantly across data types (*C Inset*, Pearson correlation, $*P < 0.05$). Pairs of opposing face templates are constructed for each dimension and data type (*B*, behavioral templates; *D*, raFG templates) for visualization and interpretation purposes. Images reproduced from refs. 46–50.

Pattern-Based Mapping of Facial Identity

Participants viewed a set of 120 face images (60 identities × 2 expressions), carefully controlled with respect to both high-level and low-level image properties (*SI Text*). Each image was presented at least 10 times per participant across five fMRI sessions using a slow event-related design (100-ms stimulus cue, followed by 900-ms stimulus presentation and 7-s fixation). Participants performed a one-back identity task across variation in expression (accuracy was high for each participant scoring above 92%).

Multivoxel pattern-based mapping (19) was carried out to localize cortical regions responding with linearly discriminable patterns to different facial identities. To this end, we separately computed at each location within a cortical mask the discriminability of every pair of identities using linear support vector machine (SVM) classification and leave-one-run-out cross-validation (*Methods* and *SI Text*). The resulting information-based map of each participant was normalized to a common space and analyzed at the group level to assess the presence of identity-related information and its approximate spatial location.

This analysis revealed multiple regions (Fig. S1) in the bilateral fusiform gyrus (FG), the inferior frontal gyrus (IFG), and the right posterior superior temporal sulcus (pSTS). Discrimination levels were compared against chance via one-sample two-tailed t tests across participants [false discovery rate (FDR)-corrected; $q < 0.05$]. Overall discriminability peaked in a region of interest (ROI) covering parts of the right anterior FG and the parahippocampal gyrus ($t_7 = 12.07$, $P < 10^{-5}$).

To ensure that other key regions were not missed, we included another region of potential interest for face processing localized in the anterior medial temporal gyrus (aMTG) (20) at a less conservative threshold ($q < 0.10$). Further, above-chance discrimination accuracy was confirmed in the bilateral fusiform face area (FFA) (21) in agreement with previous work (7, 22, 23) but not in the early visual cortex (EVC) (*SI Text*).

In summary, a total of eight ROIs localized through pattern-based mapping along with the bilateral FFA were found to be likely candidates for hosting representations of facial identity. Accordingly, these regions formed the basis for the investigation of neural representations reported below.

The Similarity Structure of Human Face Space

To evaluate the structure of the neural data relevant for identity representation, we extracted the discriminability values of all pairs of facial identities for each of 10 ROIs. Specifically, after mapping these ROIs in each participant, we separately stored, for each participant and ROI, all pairwise discrimination values corresponding to 1,770 identity pairs (i.e., all possible pairs derived from 60 identities).

Analogous behavioral measurements were collected in a separate experiment in which the confusability of the stimuli was assessed. Briefly, pairs of faces with different expressions were presented sequentially for 400 ms, and participants were asked to perform a same/different identity task. Participants were tested with all facial identity pairs across four behavioral sessions before fMRI scanning. The average accuracy in discriminating each identity pair provided the behavioral counterpart of our neural pattern-discrimination data.

Next, metric multidimensional scaling (MDS) was applied to behavioral and neural discriminability vectors averaged across participants. This analysis forms a natural bridge between recent examinations of neural-based similarity matrices in visual perception (24) and traditional investigations of behavioral-based similarity in the study of face space (2, 25). The outcome of this analysis provides us with the locations of each facial identity in a multidimensional face space. Fig. 1 *A* and *C* illustrates the distribution of facial identities across the first two dimensions for behavioral data and for right anterior FG data; we focus on this ROI both because of the robustness of its mapping and because of its central role in the processing of facial identity (23). The first two MDS dimensions are particularly relevant, because, as detailed below, they contribute important information for reconstruction purposes.

An examination of the results suggests an intuitive clustering of faces based on notable traits such as eyebrow salience. To facilitate the interpretation of these dimensions, faces were separately averaged on each side of the origin proportionally to their coefficients on each axis. This procedure yielded two opposing templates per dimension whose direct comparison informs the perceptual properties encoded by that particular dimension (Fig. 1 *B* and *D*). The comparison of these templates reveals a host of differences, such as eyebrow thickness, facial hair (i.e., stubble), skin tone, nose shape, and mouth height.

Further, the analysis of the behavioral data produced results similar to that of the fMRI data. To evaluate this correspondence, we correlated the coefficients of each facial identity across dimensions extracted for the two data types. This analysis confirmed the similarity between the organization of the first dimensions across behavioral and right aFG data (Fig. 1*C*, *Inset*); a broader evaluation of this correspondence is targeted by the assessment of image reconstructions below.

In sum, the present findings verify the presence of consistent structure in our data, assess its impact on the correspondence between behavior and neural processing, and account for this structure in terms of major visual characteristics spanning a range of shape and surface properties.

Derivation of Facial Features Underlying Face Space

The organization of face space is arguably determined by visual features relevant for identity encoding and recognition. An

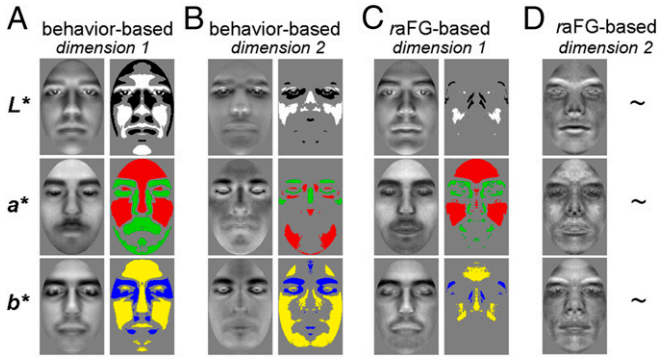

Fig. 2. CIs derived from MDS analyses of behavioral (*A* and *B*) and raFG (*C* and *D*) data. Each pair shows a raw CI (*Left*) and its analysis (*Right*) with a permutation test ($q < 0.05$, FDR correction across pixels; ~, not significant). Bright/dark, red/green, and yellow/blue regions in analyzed CIs mark areas of the face brighter (L*), redder (a*), or more yellow (b*) than chance for positive versus negative templates in Fig. 1. Results are shown separately for the first (*A* and *C*) and second (*B* and *D*) dimensions of the data.

inspection of the results above (Fig. 1) suggests that a simple linear method can capture at least some of these features.

For this purpose, we deployed the following procedure. First, for each dimension, we subtracted each corresponding template from its counterpart, thereby obtaining another template akin to a classification image (CI) (26–28)—that is, a linear estimate of the image-based template that best accounts for identity-related scores along a given dimension (*Methods*). Then, this template was assessed pixel-by-pixel with respect to a randomly generated distribution of templates (i.e., by permuting the scores associated with facial identities) to reveal pixel values lower or higher than chance (two-tailed permutation test; FDR correction across pixels, $q < 0.05$). These analyses were performed separately for each color channel after converting images to CIEL*a*b* (where L*, a*, and b* approximate the lightness, red:green, and yellow:blue color-opponent channels of human vision). Examples of raw CIs and their analyses are shown in Fig. 2.

Consistent with our observations regarding face clustering in face space, the CIs revealed extensive differences in surface properties such as eyebrow thickness and skin tone as well as in shape properties such as nose shape and relative mouth height. For instance, wide bright patches over the forehead and cheeks reflect sensitivity to color differences, whereas thinner patches along the length of the nose and the mouth provide information on shape differences. Also, these differences extend beyond lightness to chromatic channels, whether accompanied or not by similar L* differences. Further, behavioral CIs exhibited larger differences than their neural counterparts. However, most of the ROIs appeared to exhibit some sensitivity to image-based properties.

Thus, our methods were successful in using the similarity structure of neural and behavioral data to derive visual features that capture the topography of human face space.

Facial Image Reconstruction

An especially compelling way to establish the degree of visual detail captured by a putative set of face space features is to determine the extent to which they support identifiable reconstructions of face images. Accordingly, we carried out the following procedure (Fig. 3).

First, we systematically left out each facial identity and estimated the similarity space for the remaining 59 identities. This space is characterized by a set of visual features corresponding to each dimension as well as an average face located at the origin of the space. Second, the left-out identity was projected into this space based on its neural or behavioral similarity with the other faces, and its coordinates were retrieved for each dimension. Third, significant features were weighted by the corresponding coordinates and linearly combined along with the average face to generate an image reconstruction. This procedure was carried out separately for behavioral data and for each ROI to generate exemplars with both emotional expressions. Last, reconstructions were combined across all ROIs to generate a single set of neural-based reconstructions (*SI Text*).

Reconstruction accuracy was quantified both objectively, with the use of a low-level L2 similarity metric, and behaviorally, by asking naïve participants to identify the correct identity of a stimulus from two different reconstructions using a two-alternative forced choice task.

Overall, we found that reconstructions for each emotional expression were accurate above chance by either type of evaluation and emotional expression (one-sample *t* tests against chance) (see Fig. 4 for reconstruction exemplars and Fig. 5 *A* and *B* for accuracy estimates). Behavioral estimates surpassed their neural counterparts in accuracy for both evaluation metrics ($P < 0.01$, two-way analysis of variance across data types and emotional expression); no difference across expression and no interaction with data type were found. Additional analyses found significant variation in objective accuracy across the 10 ROIs ($P < 0.01$, one-way analysis of variance). Interestingly, further tests against chance-level performance showed that only three ROIs in the bilateral FG provided significant accuracy estimates (Fig. 5*C*).

Next, we constructed pixelwise accuracy maps separately for each color channel and data type to quantify reconstruction

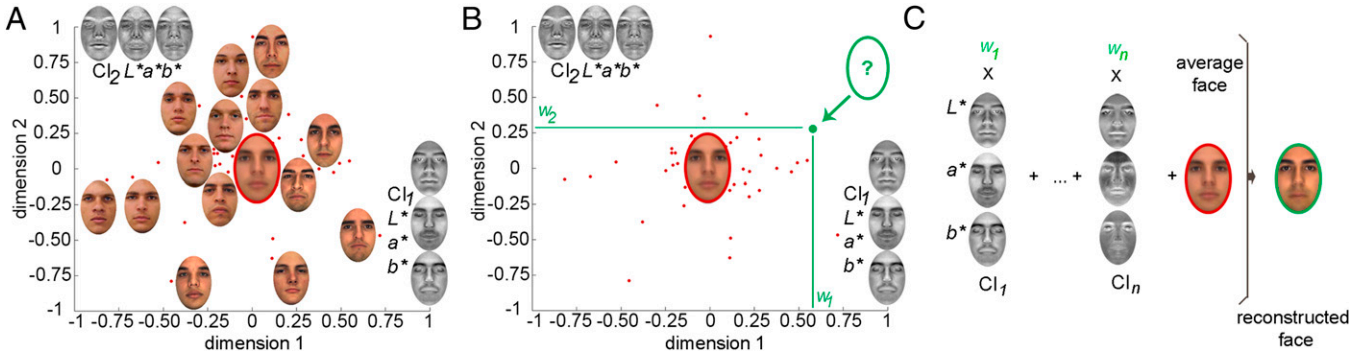

Fig. 3. Steps involved in the reconstruction procedure: (*A*) We estimate a multidimensional face space associated with a cortical region, and we derive CI features for each dimension in CIEL*a*b* color space along with an average face (for simplicity, only two dimensions are displayed above); (*B*) we project a new face in this space based on its neural similarity with other faces, and we recover its coordinates; and (*C*) we combine CI features proportionately with the coordinates of the new face to achieve reconstruction. Thus, as long as we can estimate the position of a stimulus in face space we are able to produce an approximation of its visual appearance. Images reproduced from refs. 46–50.

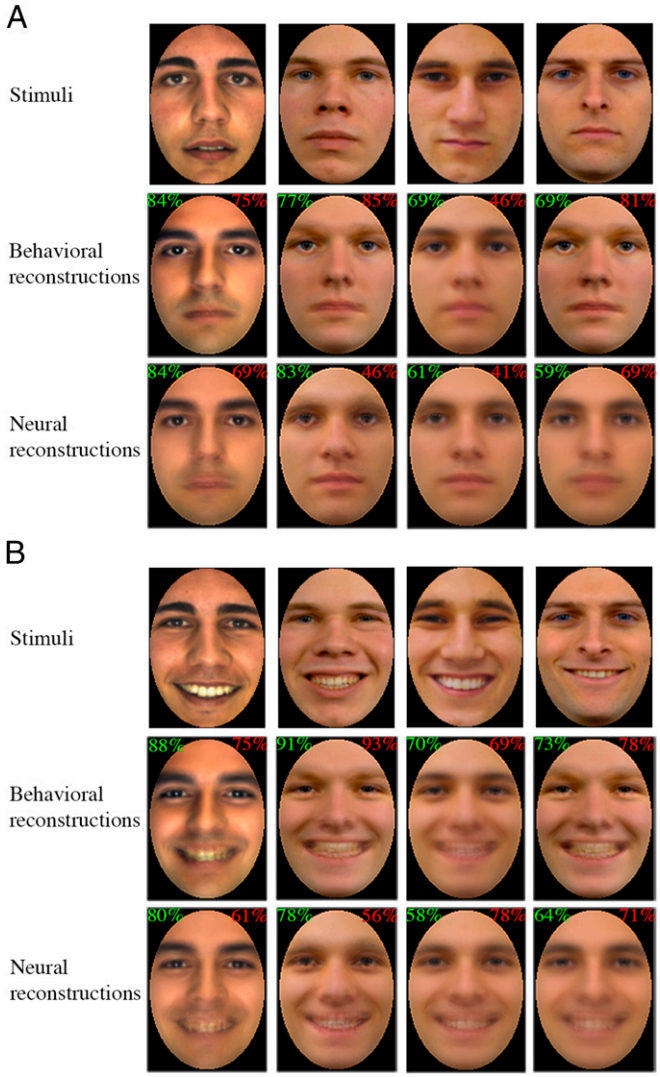

Fig. 4. Examples of face stimuli and their reconstructions from behavioral and fMRI data across (*A*) neutral and (*B*) happy expressions. Numbers in the top corners of each reconstruction show its average experimentally based accuracy (green) along with its image-based accuracy (red). Images reproduced from refs. 48–50.

quality across the structure of a face (Fig. S2). In agreement with our evaluation of visual features, the results suggest that a variety of shape and surface properties across all color channels contribute to reconstruction success.

To compare behavioral and neural reconstructions, we related accuracy estimates for the two types of data across face exemplars. Overall, we found significant correlations for both experimentally derived estimates of accuracy (Pearson correlation; $r = 0.39$, $P < 0.001$) and for image-based estimates ($r = 0.45$, $P < 0.001$), confirming a general correspondence between the two types of data. Similar results were also confirmed by comparing image-based estimates of accuracy for the ROIs capable of supporting reconstruction and their behavioral counterpart (right aFG, $r = 0.45$; right pFG, $r = 0.56$; left pFG, $r = 0.44$; $P < 0.001$).

Further, a more thorough examination of the ROIs involved in reconstruction considered the similarity of their MDS spaces to their behavioral counterpart. Specifically, we computed the goodness of fit for Procrustes alignments between neural and behavioral spaces. This analysis found a systematic correspondence between the two types of space (Fig. S3*A*), especially for the bilateral posterior FG ($P < 0.001$, two-tailed permutation test). Also, fit estimates for FG regions, but not for a control ROI in the right aMTG, were adversely impacted by increases in space dimensionality as revealed by positive correlations between alignment error and the number of dimensions (right aFG, $r = 0.67$; right pFG, $r = 0.71$; left pFG, $r = 0.81$; $P < 0.01$). Last, a closer examination of 2D space alignments within each participant (Fig. S3*B*) showed significant variation in fit estimates across ROIs ($P < 0.05$, one-way analysis of variance); pairwise comparisons also found that the left pFG provided a better fit with the behavioral data than the right aMTG ($t_7 = 3.09$, $P < 0.05$).

To conclude, visual features derived from neural or behavioral data were capable of supporting facial image reconstruction, with a good degree of agreement between the two, although neural reconstructions were driven primarily by bilateral FG activation.

Discussion

How is facial identity represented by the human visual system? To address this question, we undertook a comprehensive investigation that combines multiple, converging methods in the study of visual recognition, as detailed below.

Cortical Mapping of Facial Identity. Growing sophistication in the analysis of neuroimaging data has facilitated the mapping of the neural correlates of face identification. The examination of face-selective cortex has implicated areas of the FG, STS, and the anterior temporal lobe (ATL) in identification (29–31). Recent investigations relying on multivoxel pattern analysis have extended this work by identifying regions responding with separable patterns of activation to different facial identities, regardless of whether they are accompanied by face selectivity (7, 19, 22, 23, 32).

In contrast to previous studies, which have explored the neural code associated with a relatively small number of facial identities, the present study examines the neural and psychological representations underlying an extensive, homogeneous set of unfamiliar faces. This constitutes an exacting test of identity mapping based on fine-grained sensitivity to perceptual differences.

Consistent with previous studies, our investigation found above-chance discrimination in multiple FG, IFG, and STS regions as well as in the FFA. However, the ability of a region to support identity discrimination does not necessarily imply that it encodes visual face representations. Higher level semantic information (33) or even a variety of unrelated task/stimulus

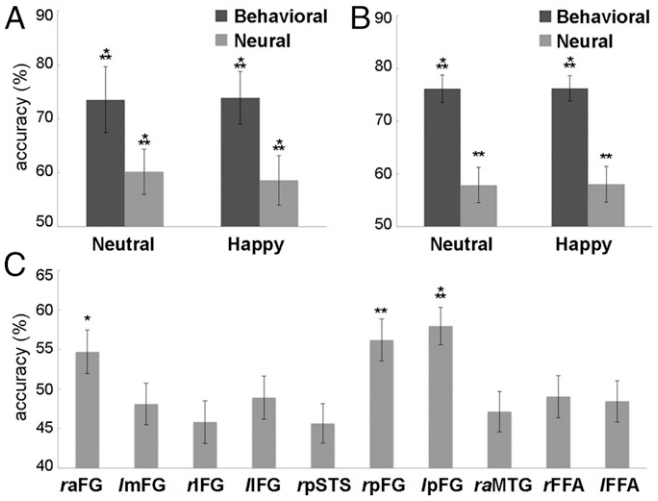

Fig. 5. Reconstruction accuracy using (*A*) experimentally based and (*B*) image-based estimates of behavioral and neurally derived reconstructions. Image-based estimates are also separately shown for each ROI collapsed across expressions (*C*). Error bars show ±1 SE across (*A*) participants and (*B* and *C*) reconstructions (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). FG, fusiform gyrus; IFG, inferior frontal gyrus; MTG, middle temporal gyrus; STS, superior temporal sulcus.

properties may account for pattern discrimination (34). The latter possibility is a source of concern, especially given certain limitations of the fMRI signal in decoding facial identity (35).

The present findings address this concern by demonstrating that at least certain regions localized via pattern-based mapping contain visual information critical for facial identification. Specifically, three regions of the posterior and anterior FG were able to support identifiable reconstructions of face images. Interestingly, the FFA did not support similar results. However, recent work has shown that the FFA is particularly sensitive to templates driving face detection (36) and can even support the visual reconstruction of such templates (27). Thus, the current results agree with the involvement of the FFA primarily in face detection and, only to a lesser extent, in identification (37, 38). Also, the inability of more anterior regions of the IFG and aMTG to support image reconstruction is broadly consistent with their involvement in processing higher level semantic information (39).

Thus, our results confirm that facial identification relies on an entire network of cortical regions, and importantly, they point to multiple FG regions as responsible for encoding image-based visual information critical for the representation of facial identity.

Human Face Space and Its Visual Features. What properties dominate the organization of human face space? To be clear, our investigation does not target the entirety of face space but rather a specific subdomain: young adult Caucasian males. Furthermore, we avoid large differences in appearance due to hair, outer contour, or aspect ratio, which are obvious, well-known cues to recognition (2, 40). Instead, we reason that understanding the structure of face representation in a carefully restricted domain challenges the system maximally and is instrumental in understanding recognition at its best. Our expectation, though, is that the principles revealed here generalize to face recognition as a whole.

A combination of MDS and CI analyses allowed us to assess and visualize the basic organization of face space in terms of both shape and surface properties. Our results reveal a host of visual properties across multiple areas of the face and across different color channels. Notably, we find evidence for the role of eyebrow salience, nose shape, mouth size, and positioning, as well as for the role of skin tone. Interestingly, these results agree with previous behavioral work; for instance, the critical role of the eyebrows in individuation has been specifically tested and confirmed (41). Also, several of the properties above appear to be reflected in the structure of both behavioral and neural data.

At the same time, we note that our analyses reveal only a handful of significant features relying on low-dimensional spaces, whereas human face space is believed to be high-dimensional (42). Critically, the number of significant dimensions recovered from the data depends on the signal-to-noise ratio (SNR) of the data as well as on the size of the stimulus set. Here, the restricted number of trials (e.g., 10 presentations per stimulus during scanning) imposes direct limitations on the SNR and allows only the estimation of the most robust features and of the dimensions associated with them. Hence, the current results do not speak directly to the dimensionality of face space, but they do open up the possibility of its future investigation with the aid of more advanced imaging techniques and designs.

Last, regarding the presence of visual information across multiple color channels, we note that, traditionally, the role of color in face identification has been downplayed (43, 44). However, recent work has pointed to the value of color in face detection (12) and even in identification (13). As previously suggested (45), color may aid identification when the availability of other cues is diminished. More generally, the difficulty of the task, as induced here by the homogeneity of the stimulus set, could lead to the recruitment of relevant color cues. From a representational standpoint, the current findings suggest that color information is included in the structure of face space and, thus, available when needed. However, additional investigations targeting this hypothesis are needed to ascertain its precise scope and its validity.

In sum, the present results establish the relevance of specific properties, their relative contribution to molding the organization of face space, and conversely, our ability to derive them from neural and behavioral data. More generally, we conclude that specific perceptual representations are encoded in high-level visual cortex and that these representations are fundamentally structured by the visual properties described here.

Facial Identity and Image Reconstruction. The fundamental idea grounding our reconstruction method is that, as long as the relative position of an identity in face space is known, its visual appearance can be reconstructed from that of other faces in that space. In a way, our method validates the classic concept of face space (6) by making it instrumental in the concrete enterprise of image reconstruction. Also, from a practical perspective, the presumed benefit of this approach is a more efficient reconstruction method relying on empirically derived representations rather than on hypothetical, prespecified features. For instance, our reconstruction procedure involves only a handful of features (*SI Text*) whose relevance for representing facial identity is ensured by the process of their derivation and selection. The successful reconstruction of face images drawn from a homogeneous stimulus set provides strong support for this method.

Overall, the application of reconstruction to both behavioral and fMRI data has extensive theoretical and methodological implications. Theoretically, it points to the general correspondence of face space structure across behavioral and neural data; at the same time, it highlights the variation of face space across different cortical regions, only some of which contain relevant visual information. Methodologically, the generality and the robustness of the current approach allow its extension to other neuroimaging modalities as well as to data gleaned from patient populations (e.g., to examine distortions of visual representations in prosopagnosia or autism). Thus, our reconstruction results not only provide specific information about the nature of face space but also allow a wide range of future investigations into visual representations and their application to image reconstruction.

In conclusion, our work sheds light on the representational basis of face recognition regarding its cortical locus, its underlying features, and their visual content. Our findings reveal a range of shape and surface properties dominating the organization of face space, they show how to synthesize these properties into image-based features of facial identity, they establish a general method for using these features in image reconstruction, and last, they validate their behavioral and neural plausibility. More generally, this work demonstrates the strengths of a multipronged multivariate paradigm that brings together functional mapping, investigations of behavioral and neural similarity space, as well as feature derivation and image reconstruction.

Methods

Stimuli and Design. A total of 120 images of adult Caucasian males (60 identities × 2 expressions) were selected from multiple face databases and further processed to ensure their homogeneity.

Eight right-handed Caucasian adults participated in nine 1-h experimental sessions (four behavioral and five fMRI). During behavioral sessions, participants viewed pairs of faces, presented in succession, and judged whether they represented the same/different individuals. During fMRI scans, participants performed a continuous one-back version of the same task using a slow event-related design (8-s trials). We imaged 27 oblique slices covering the ventral cortex at 3T [2.5^3 mm voxels, 2 s time-to-repeat (TR)]. Informed consent was obtained from all participants, and all procedures were approved by the Institutional Review Board of Carnegie Mellon University.

Pattern-Based Brain Mapping. Multivoxel pattern-based mapping was performed by walking a spherical searchlight voxel-by-voxel across a cortical mask (see Fig. S1*A* for a group mask). At each location, a single average voxel pattern was extracted per run for every facial identity. To estimate neural discriminability, linear SVM classification was applied across these patterns for each identity pair using leave-one-run-out cross-validation. Then, participant-specific maps were constructed by voxel-wise averaging of discrimination estimates across identity pairs. For the purpose of group

analysis, all maps were brought into Talairach space, and statistical effects were computed across participants (two-tailed *t* test against chance).

Similarity Structure Analyses. Behavioral estimates of pairwise face (dis)similarity were computed based on the average discrimination accuracy of each participant during behavioral sessions. Homologous neural-based estimates were computed with the aid of pattern classification in different ROIs based on average discrimination sensitivity. For both data types, this procedure yielded a vector of 1,770 pairwise discrimination values.

Further, discrimination vectors were encoded as facial dissimilarity matrices. These matrices were then averaged across participants and analyzed by metric MDS. To interpret the perceptual variation encoded by MDS dimensions (Fig. 1 *A* and *C*), individual faces were averaged on each side of the origin proportionally to their dimension-specific coefficients. The resulting templates were assessed using a reverse correlation approach (26). Concretely, each pair of templates thus obtained were subtracted from each other to derive a CI summarizing the perceptual differences specific to that dimension. Each CI was next analyzed, pixel-by-pixel, by comparison with a group of randomly generated CIs (*t* test; $q < 0.05$ correction across pixels). This analysis was separately conducted for the L*, a*, and b* components of face images.

Image Reconstruction Method. For every facial identity, an independent estimate of face space was constructed through MDS using all other identities. Then, the left-out identity was projected in this space via Procrustes alignment. Concretely, the MDS solution derived for all 60 identities was mapped onto the first solution, providing us with the coordinates of the target face in the original face space. The resulting coordinates are next used to weight the contribution of significant CIs in the reconstruction process; relevant CIs are selected based on the presence of significant pixels in any of three color channels via a permutation test (FDR correction across pixels; $q < 0.10$). The linear combination of significant CIs along with that of an average face is used to approximate the visual appearance of the target face. This method was conducted separately for 10 ROIs and for behavioral data. Last, a single set of neural reconstructions was derived through the linear combination of ROI-specific reconstructions via an L2 regularized regression model and a leave-one-identity-out procedure.

Neural and Behavioral Face Space Correspondence. The global correspondence between the two types of face space was assessed by bringing ROI-specific spaces into alignment with behavioral space. Goodness of fit was then estimated via sum of squared errors (SSE) between Procrustes-aligned versions of neural space and behavioral space. Fit estimates were compared with chance across systematic differences in MDS-derived space dimensionality (from 2 to 20 dimensions) via permutation tests. Last, fit estimates across ROIs were compared with each other through parametric tests across participants.

ACKNOWLEDGMENTS. This research was supported by the Natural Sciences and Engineering Research Council of Canada (A.N.), by a Connaught New Researcher Award (to A.N.), by National Science Foundation Grant BCS0923763 (to M.B. and D.C.P.), and by Temporal Dynamics of Learning Center Grant SMA-1041755 (to M.B.).

1. Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. *Nat Neurosci* 12(9):1187–1196.
2. Johnston RA, Milne AB, Williams C, Hosie J (1997) Do distinctive faces come from outer space? An investigation of the status of a multidimensional face-space. *Vis Cogn* 4:59–67.
3. Leopold DA, O'Toole AJ, Vetter T, Blanz V (2001) Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat Neurosci* 4(1):89–94.
4. Loffler G, Yourganov G, Wilkinson F, Wilson HR (2005) fMRI evidence for the neural representation of faces. *Nat Neurosci* 8(10):1386–1390.
5. O'Toole AJ (2011) *The Oxford Handbook of Face Perception*, eds Calder AJ, Rhodes G, Johnson M, Haxby JV (Oxford Univ Press, Oxford), pp 15–30.
6. Valentine T (1991) A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q J Exp Psychol A* 43(2):161–204.
7. Goesaert E, Op de Beeck HP (2013) Representations of facial identity information in the ventral visual stream investigated with multivoxel pattern analyses. *J Neurosci* 33(19):8549–8558.
8. Harris A, Aguirre GK (2010) Neural tuning for face wholes and parts in human fusiform gyrus revealed by FMRI adaptation. *J Neurophysiol* 104(1):336–345.
9. Liu J, Harris A, Kanwisher N (2010) Perception of face parts and face configurations: An FMRI study. *J Cogn Neurosci* 22(1):203–211.
10. Maurer D, et al. (2007) Neural correlates of processing facial identity based on features versus their spacing. *Neuropsychologia* 45(7):1438–1451.
11. Issa EB, DiCarlo JJ (2012) Precedence of the eye region in neural processing of faces. *J Neurosci* 32(47):16666–16682.
12. Bindemann M, Burton AM (2009) The role of color in human face detection. *Cogn Sci* 33(6):1144–1156.
13. Nestor A, Plaut DC, Behrmann M (2013) Face-space architectures: Evidence for the use of independent color-based features. *Psychol Sci* 24(7):1294–1300.
14. Miyawaki Y, et al. (2008) Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron* 60(5):915–929.
15. Nishimoto S, et al. (2011) Reconstructing visual experiences from brain activity evoked by natural movies. *Curr Biol* 21(19):1641–1646.
16. Thirion B, et al. (2006) Inverse retinotopy: Inferring the visual content of images from brain activation patterns. *Neuroimage* 33(4):1104–1116.
17. Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. *Neuron* 63(6):902–915.
18. Cowen AS, Chun MM, Kuhl BA (2014) Neural portraits of perception: Reconstructing face images from evoked brain activity. *Neuroimage* 94:12–22.
19. Kriegeskorte N, Formisano E, Sorger B, Goebel R (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc Natl Acad Sci USA* 104(51):20600–20605.
20. Von Der Heide RJ, Skipper LM, Olson IR (2013) Anterior temporal face patches: A meta-analysis and empirical study. *Front Hum Neurosci* 7:17.
21. Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci* 17(11):4302–4311.
22. Anzellotti S, Fairhall SL, Caramazza A (2014) Decoding representations of face identity that are tolerant to rotation. *Cereb Cortex* 24(8):1988–1995.
23. Nestor A, Plaut DC, Behrmann M (2011) Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc Natl Acad Sci USA* 108(24):9998–10003.
24. Kriegeskorte N, et al. (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60(6):1126–1141.
25. Dailey MN, Cottrell GW, Padgett C, Adolphs R (2002) EMPATH: A neural network that categorizes facial expressions. *J Cogn Neurosci* 14(8):1158–1173.
26. Murray RF (2011) Classification images: A review. *J Vis* 11(5):11.
27. Nestor A, Vettel JM, Tarr MJ (2013) Internal representations for face detection: An application of noise-based image classification to BOLD responses. *Hum Brain Mapp* 34(11):3101–3115.
28. Smith ML, Gosselin F, Schyns PG (2012) Measuring internal representations from behavioral and brain data. *Curr Biol* 22(3):191–196.
29. Fox CJ, Moon SY, Iaria G, Barton JJ (2009) The correlates of subjective perception of identity and expression in the face network: An fMRI adaptation study. *Neuroimage* 44(2):569–580.
30. Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. *Trends Cogn Sci* 4(6):223–233.
31. Rotshtein P, Henson RN, Treves A, Driver J, Dolan RJ (2005) Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat Neurosci* 8(1):107–113.
32. Natu VS, et al. (2010) Dissociable neural patterns of facial identity across changes in viewpoint. *J Cogn Neurosci* 22(7):1570–1582.
33. Çukur T, Huth AG, Nishimoto S, Gallant JL (2013) Functional subdomains within human FFA. *J Neurosci* 33(42):16748–16766.
34. Todd MT, Nystrom LE, Cohen JD (2013) Confounds in multivariate pattern analysis: Theory and rule representation case study. *Neuroimage* 77:157–165.
35. Dubois J, de Berker AO, Tsao DY (2015) Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J Neurosci* 35(6):2791–2802.
36. Gilad S, Meng M, Sinha P (2009) Role of ordinal contrast relationships in face encoding. *Proc Natl Acad Sci USA* 106(13):5353–5358.
37. Nestor A, Vettel JM, Tarr MJ (2008) Task-specific codes for face recognition: How they shape the neural representation of features for detection and individuation. *PLoS One* 3(12):e3978.
38. Tong F, Nakayama K, Moscovitch M, Weinrib O, Kanwisher N (2000) Response properties of the human fusiform face area. *Cogn Neuropsychol* 17(1):257–280.
39. Simmons WK, Reddish M, Bellgowan PS, Martin A (2010) The selectivity and functional connectivity of the anterior temporal lobes. *Cereb Cortex* 20(4):813–825.
40. Mondloch CJ, Le Grand R, Maurer D (2002) Configural face processing develops more slowly than featural face processing. *Perception* 31(5):553–566.
41. Sadr J, Jarudi I, Sinha P (2003) The role of eyebrows in face recognition. *Perception* 32(3):285–293.
42. Sirovich L, Meytlis M (2009) Symmetry, probability, and recognition in face space. *Proc Natl Acad Sci USA* 106(17):6895–6899.
43. Bruce V, Young A (1998) *In the Eye of the Beholder: The Science of Face Perception* (Oxford Univ Press, New York).
44. Kemp R, Pike G, White P, Musselman A (1996) Perception and recognition of normal and negative faces: The role of shape from shading and pigmentation cues. *Perception* 25(1):37–52.
45. Yip AW, Sinha P (2002) Contribution of color to face recognition. *Perception* 31(8):995–1003.
46. Phillips JP, Moon H, Rizvi SA, Rauss PJ (2000) The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans Pattern Anal Mach Intell* 22(10):1090–1104.
47. Phillips JP, Wechsler H, Huang J, Rauss PJ (1998) The FERET database and evaluation procedure for face-recognition algorithms. *Image Vis Comput* 16(5):295–306.
48. Thomaz CE, Giraldi GA (2010) A new ranking method for principal component analysis and its application to face image analysis. *Image Vis Comput* 28(6):902–913.
49. Martinez AR, Benavente R (1998) *The AR Face Database*, CVC Technical Report #24. Available at www.cat.uab.cat/Public/Publications/1998/MaB1998/CVCReport24.pdf. Accessed December 18, 2015.
50. Langner O, et al. (2010) Presentation and validation of the Radboud Faces Database. *Cogn Emotion* 24(8):1377–1388.
51. Cox RW (1996) AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29(3):162–173.
52. Sinha P, Russell R (2011) A perceptually based comparison of image similarity metrics. *Perception* 40(11):1269–1281.