Spatiotemporal dynamics of similarity-based neural representations of facial identity

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

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

Contributed by Marlene Behrmann, November 22, 2016 (sent for review September 6, 2016; reviewed by Ming Meng and Jim W. Tanaka)

Humans' remarkable ability to quickly and accurately discriminate among thousands of highly similar complex objects demands rapid and precise neural computations. To elucidate the process by which this is achieved, we used magnetoencephalography to measure spatiotemporal patterns of neural activity with high temporal resolution during visual discrimination among a large and carefully controlled set of faces. We also compared these neural data to lower level "image-based" and higher level "identity-based" model-based representations of our stimuli and to behavioral similarity judgments of our stimuli. Between ∼50 and 400 ms after stimulus onset, face-selective sources in right lateral occipital cortex and right fusiform gyrus and sources in a control region (left V1) yielded successful classification of facial identity. In all regions, early responses were more similar to the image-based representation than to the identity-based representation. In the face-selective regions only, responses were more similar to the identity-based representation at several time points after 200 ms. Behavioral responses were more similar to the identity-based representation than to the image-based representation, and their structure was predicted by responses in the face-selective regions. These results provide a temporally precise description of the transformation from low- to high-level representations of facial identity in human face-selective cortex and demonstrate that face-selective cortical regions represent multiple distinct types of information about face identity at different times over the first 500 ms after stimulus onset. These results have important implications for understanding the rapid emergence of fine-grained, high-level representations of object identity, a computation essential to human visual expertise.

face processing | magnetoencephalography | decoding | representational similarity analysis | face identity

Humans can discriminate among thousands of highly similar and complex visual patterns, such as face identity, in less than half a second (1, 2). Efficient within-category discrimination of facial identity is important for real-world decisions (e.g., classifying a person as a friend or stranger) and social interactions. Progress has been made in elucidating the neural mechanisms underlying the discrimination of individual face identities in humans. Using fMRI, these studies demonstrate that individual face identities are represented by spatially distributed patterns of neural activity within occipitotemporal cortex (3–13). Because of the poor temporal resolution in fMRI studies (typically around 2 s), however, our understanding of the neural basis of discrimination among complex visual patterns in humans remains limited. For example, within a given region, different information relevant to discrimination may be represented at different times over the first few 100 ms after stimulus onset. However, current models of the neural basis of face recognition in humans do not typically allow for this possibility, because they usually assign a single functional role to each face-selective region (14).

A few previous studies (15–18) have explored the temporal properties of the neural representation of individual face identities in humans. However, the measurements of representations in these studies were based on variations in the amplitude of neural activity at just one or two time points, typically sampling from different sensors for different time points (16–18) or on the temporal dynamics of signal from a small number of intracranial electrodes in fusiform gyrus (15). Furthermore, those studies that have investigated the nature of the facial identity information encoded in the neural data have used analyses that are limited to relatively low-level visual information (e.g., manual pixel-based measurements of eye or cheek color) (15, 17, 18) and/or analyses that sample from different brain regions for different time points (18). Hence, these studies provide limited information about the temporal dynamics of neural representations of facial identity. To allow fast and accurate discrimination of face identity in the real world, the human visual system must rapidly (within the first few 100 ms) transform image-based inputs into a more abstract, less image-based representation with greater tolerance to identity-preserving image transformations (19, 20). The computations underlying the temporal emergence of these high-level representations of facial identity are largely untapped. Hence, the neural basis of human face recognition cannot be fully understood without further examination of the temporal dimension of the neural representation of face identity.

In the current study, we investigated three important and unanswered questions about the neural basis of within-category discrimination among a large and carefully-controlled set of facial

Significance

Humans can rapidly discriminate among many highly similar facial identities across identity-preserving image transformations (e.g., changes in facial expression), an ability that requires the system to rapidly transform image-based inputs into a more abstract, identity-based representation. We used magnetoencephalography to provide a temporally precise description of this transformation within human face-selective cortical regions. We observed a transition from an image-based representation toward an identity-based representation after ∼200 ms, a result suggesting that, rather than computing a single representation, a given face-selective region may represent multiple distinct types of information about face identity at different times. Our results advance our understanding of the microgenesis of fine-grained, high-level neural representations of object identity, a process critical to human visual expertise.

Author contributions: M.D.V., A.N., and M.B. designed research; M.D.V. performed research; M.D.V., A.N., and D.C.P. contributed new reagents/analytic tools; M.B. supervised the project; M.D.V., A.N., D.C.P., and M.B. analyzed data; and M.D.V., A.N., D.C.P., and M.B. wrote the paper.

Reviewers: M.M., Dartmouth College; and J.W.T., University of Victoria.

The authors declare no conflict of interest.

Data deposition: Data reported in this paper are available on figshare at https://figshare.com/articles/FST_raw_data/4233107 (doi: 10.6084/m9.figshare.4233107.v1).

[1]To whom correspondence should be addressed. Email: behrmann@cmu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1614763114/-/DCSupplemental.

identities in humans. (*i*) When do spatiotemporal patterns of activity within face-selective cortex carry information sufficient for discrimination of facial identity across changes in facial expression? (*ii*) What types of information about facial identity (e.g., low-level image-based information or higher level identity-based information, information encoded in human behavioral similarity judgments) are represented by these spatiotemporal patterns of activity, and when? (*iii*) Where in the brain are these different types of information represented (e.g., in face-selective or control regions)? To investigate these questions, we developed a paradigm that permits the characterization of the representation of a large set of individual face identities from spatiotemporal patterns of neural activity, with extremely high temporal resolution. We used a small-sample design inspired by single-cell recording studies in nonhuman primates (21–23) and psychophysics (24, 25). We recorded comprehensive brain activity with magnetoencephalography (MEG) in four adult human participants while they viewed face images from a large, carefully controlled set (91 face identities, with two facial expressions per identity; Fig. 1), with a sufficiently large number of trials for each face identity (104–112 trials per face identity, 9,464–10,192 trials per participant) to be able to evaluate the representation of individual face identities in each participant (26). We used MEG because it has excellent temporal resolution and sufficient spatial resolution for decoding of fine visual information from spatial patterns of neural activity (26, 27). In each participant, we used an independent functional localizer task in MEG to identify face-selective regions in right lateral occipital cortex and right fusiform gyrus. We also used an anatomical atlas to localize left V1. We selected V1 to serve as a control area because it is known to encode relatively low-level visual information, and we used left V1 instead of right V1 because we expected that left V1 would be less likely to be influenced by interactions with the aforementioned right-hemisphere face-selective regions. Note, however, that the structure of representations in right and left V1 seems to be qualitatively similar (see *Results*, *Left Hemisphere*).

To evaluate the extent to which spatiotemporal patterns of activity in each of the aforementioned regions discriminated among the 91 face identities, we used a pairwise k-nearest-neighbor classifier to classify all possible pairs of face identities across changes in facial expression. We then evaluated what information was encoded in the neural data by comparing the pairwise dissimilarity structure of the neural data within each region of interest to each of two representations: (*i*) an "image-based" representation computed from a neural model simulating the response of simple cells in V1 (29) and (*ii*) an "identity-based" representation, in which all face pairs have

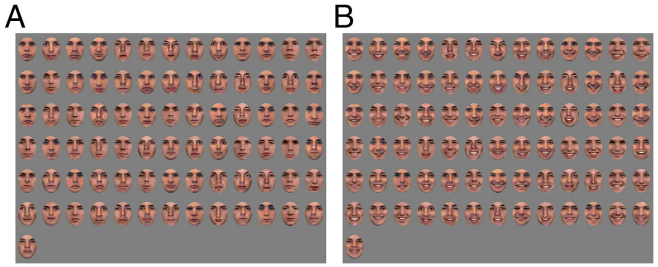

Fig. 1. Examples of stimuli presented in the current study. (*A*) All faces are presented with a neutral facial expression. Eight additional face identities (models 21, 24–27, and 32–34) from the NimStim Face Stimulus Set (28) and four additional identities from the Psychological Image Collection at Sterling Pain Expressions Set (male models 5, 6, 8, and 10) were included in our stimulus set but cannot be reproduced here under the release terms of those stimulus sets. (*B*) All details as described for *A*, with the exception that faces are presented with happy expressions.

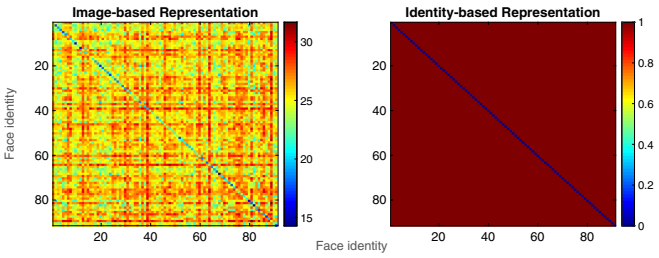

Fig. 2. Heat maps showing pairwise distance values for "image-based" and "identity-based" representations of the facial identities in our stimulus set. Each cell of each matrix shows the distance value associated with the comparison of two identities, across a change in facial expression, with hotter colors indicating a greater distance (i.e., larger difference between the representations of each identity in the pair).

dissimilarity of 0 if they are the same identity, and 1 otherwise (Fig. 2). For each postbaseline time point (where each time point is the starting point of a 60-ms sliding window used for all analyses) and each region of interest, we then examined which of these two representations was more similar to the neural data. To examine the extent to which the spatiotemporal patterns measured in the neural data could account for behavior, we also compared the pairwise structure of the neural data to pairwise behavioral judgments of a subset of the stimuli presented during the MEG experiment.

Results

Behavior During MEG Face Identity Task. In each of the 26–28 blocks of the task, participants viewed each of 91 face identities four times (twice per expression) while brain activity was recorded with MEG. Participants were instructed to maintain fixation and to press a button whenever they saw the same face identity repeated, regardless of facial expression. Across all participants and blocks, mean d-prime was 2.21 (SD = 0.52).

MEG.

Functional and anatomical regions of interest. To localize source points in the MEG source space that responded selectively to faces, we used a one-back localizer task with a block design and with stimuli from five different categories: faces, houses, objects, scrambled objects, and words. Activations from this task (faces > objects) were used to identify face-selective source points within right lateral occipital cortex (rLO-faces) and right fusiform gyrus (rFG-faces), two regions commonly implicated in neuroimaging studies of face perception (see *Materials and Methods* for details). These two face-selective regions are shown in Fig. 3. In each participant, we also used an anatomical atlas in Freesurfer (30) to identify source points within left V1. We restricted all further analyses to these regions, and to corresponding regions in the opposite hemisphere.

Classification of facial identity. To examine the extent to which each region encoded information about facial identity, we used a binary k-nearest-neighbor classifier (k = 1) to classify each possible pair of facial identities based on the spatiotemporal pattern of activity within each region, and within a 60-ms sliding temporal window. All classification was performed across a change in facial expression (*Materials and Methods*).

Fig. 4 shows classification accuracy for each region of interest. In all three regions, classification accuracy exceeded chance after ~50 ms, reaching a peak between 100 and 200 ms, with a clear secondary peak observed in rLO-faces and rFG-faces at ~250 ms and decreasing back to chance by ~400 ms. Note that here and elsewhere in this paper time is expressed as the beginning of the 60-ms sliding temporal window used for classification and other analyses of the neural data, as in a previous study using similar methods (15). Hence, the results presented for a given

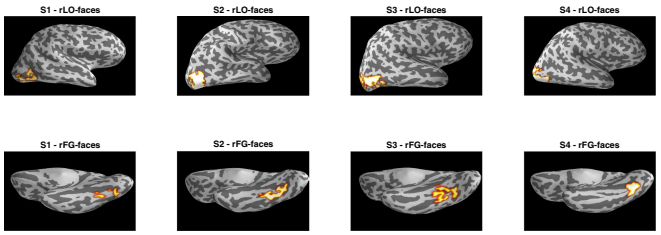

Fig. 3. Face-selective regions of interest (rLO-faces and rFG-faces) for each participant. All regions are plotted on inflated cortical surface reconstructions generated separately for each participant. rLO-faces is shown with a lateral view of the right hemisphere, and rFG-faces is shown with a ventral view.

time point can reflect measurements from that time point and up to 60 ms afterward. Between 100 and 200 ms, accuracy was higher in lV1 than in the other two regions. However, there were several periods after 200 ms after which accuracy was higher in the two face-selective regions than in lV1. Together, these results indicate that in all three regions there was sufficient information for cross-expression classification of facial identity between ∼50 ms and 400 ms after stimulus onset.

Comparison with model-based representations. To investigate what information was encoded in each region, and when, we compared the representational structure of the neural data within each region to two representations with known properties: an image-based representation based on relatively low-level visual properties of the stimuli and a higher level identity-based representation, which solely encodes whether or not two face images differ in identity and is not sensitive to any other property of the images (Fig. 2).

Correlations between the neural data and the image- and identity-based representations are shown in Fig. 5. In lV1, the neural data were significantly more similar to the image-based representation than to the identity-based representation at nearly all time points after stimulus onset. In rLO-faces, the neural data were more similar to the image-based representation between 100 and 200 ms after stimulus onset but were more similar to the identity-based representation at several time points between 200 and 300 ms after stimulus onset. The transition observed after 200 ms seems to reflect a drop in the correlation with the image-based representation, with no corresponding drop in the correlation with the identity-based representation. A similar pattern was observed in rFG-faces, with the exception that the transition to the identity-based representation was less pronounced and did not occur until after 300 ms. Together, these results suggest that spatiotemporal patterns of activity in both early visual cortical regions such as lV1 and face-selective occipital and temporal regions primarily represent image-based properties of face identity between 100- and 200-ms stimulus onset, and that only the latter face-selective regions transition toward a higher level, more identity-specific representation after 200 ms (after 300 ms in rFG-faces). We observed qualitatively similar patterns of correlations (particularly in rLO-faces and lV1) in a comparison of the neural data to layers of a deep neural network trained on our stimuli (*Supporting Information* and Fig. S1).

Behavioral Similarity Ratings. Behavioral dissimilarity ratings (Fig. 6) were strongly and positively correlated with both the identity- ($r = 0.89$) and image-based ($r = 0.79$) representations but were significantly more strongly correlated with the former than with the latter, $P < 0.0001$ (31). Correlations between the behavioral and neural data were statistically significant at most postbaseline time points (Fig. 6). Correlations were not significantly stronger for the face-selective regions than for the control region at any postbaseline time points. However, a multiple regression analysis indicated that responses in both face-selective regions predicted behavioral responses after controlling for responses in the control region (lV1) between ∼100 and 250 ms after stimulus onset, and also between 350 and 400 ms in rLO-faces (see *Materials and Methods* for details). Overall, this pattern indicates that behavioral responses primarily reflect an identity-based representation but may also reflect image-based properties to a lesser extent, and that these behavioral responses can be predicted by responses in face-selective regions during earlier (i.e., between 100 and 200 ms) and later (after 200 ms) time periods in which these regions seem to represent lower and higher level information about facial identity, respectively.

Left Hemisphere. We extended the analyses described above to the left hemisphere (see *Supporting Information* and Fig. S2 for details).

Discussion

In the current study, we investigated when spatiotemporal patterns of activity within face-selective cortex carry information sufficient for discrimination of facial identity across changes in facial expression, what type of information about facial identity (e.g., low-level image-based information, higher level identity-based information, and information encoded in pairwise behavioral judgments of the stimuli) is represented in these patterns, and where in the brain (e.g., in face-selective or control regions) these different types of information are represented at different points in time. In two face-selective regions (rLO-faces and rFG-faces) and one control region (lV1) we first measured pairwise classification among all possible pairs of 91 face identities, across changes in facial expression so as to tap into a more abstract representation, invariant over the geometry of the input. We then compared the pairwise similarity structure of the data in each of these regions to an image-based representation based on relatively low-level visual information and to a higher level identity-based representation. We also compared the neural data to behavioral similarity judgments of the stimuli. Between ∼50 and 400 ms, we were able to decode face identity successfully in each of the three regions, with accuracy first peaking at values above 70% between 100 and 200 ms, and with a secondary

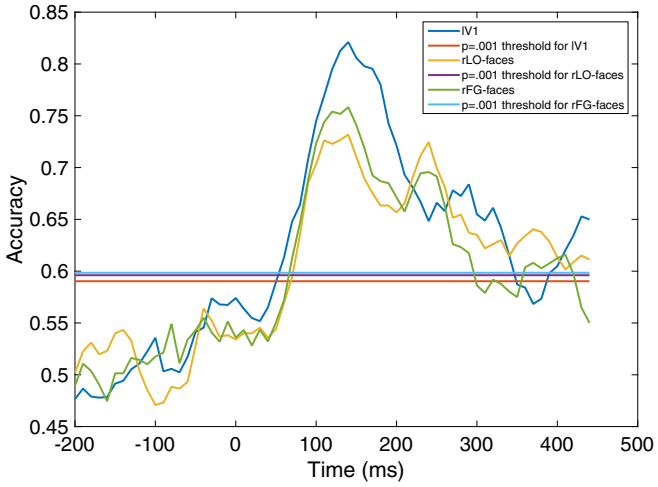

Fig. 4. Accuracy in classifying facial identity across a change in facial expression, as a function of time (milliseconds) after stimulus onset. Note that here and elsewhere in this paper time is expressed as the starting point of the 60-ms sliding window used for all analyses. Separate lines are plotted for accuracy in each region of interest (lV1, rLO-faces, and rFG-faces) and for the statistical threshold ($P = 0.001$) for each region. Accuracy is significantly greater than chance where observed values exceed the statistical threshold.

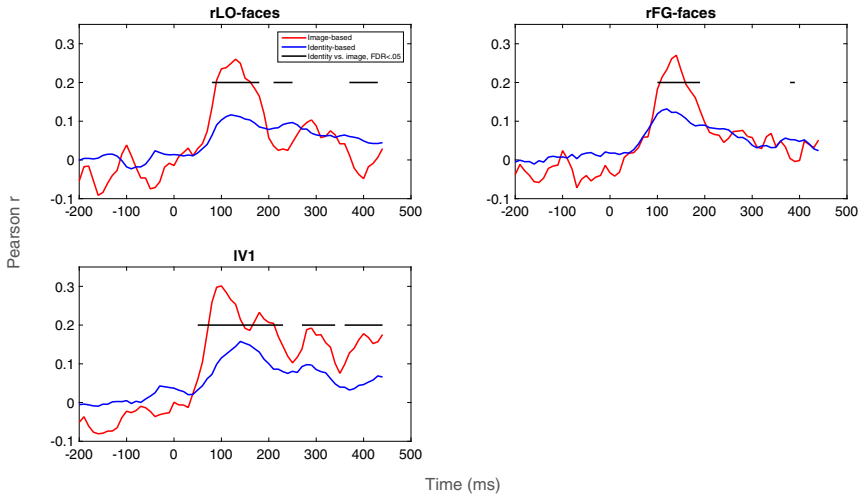

Fig. 5. Correlations between neural data and image-based and identity-based representations of our stimuli, as a function of time (milliseconds). Separate plots are given for each region of interest (rLO-faces, rFG-faces, lV1). The horizontal black line indicates postbaseline time points at which correlations differed significantly [false discovery rate (FDR) < 0.05] between the two types of representations.

peak in the face-selective regions between 200 and 300 ms. In all regions, neural responses were more similar to the image-based representation than to the identity-based representation until ∼200 ms. In the face-selective regions only, the correlation with the image-based representation dropped at several time points after 200 ms, so that responses were more similar to the identity-based representation. Behavioral responses were more similar to the identity-based representation than to the image-based representation, and their structure was predicted by responses in the face-selective regions between 100 and 250 ms, after controlling for responses in the control region.

Our finding of successful face identity classification across changes in facial expression between ∼50 and 400 ms in all regions suggests that all regions carried information relevant to discrimination over a relatively large span of time after stimulus onset. This result provides evidence that even very early visual regions can encode information sufficient for discriminating among highly similar complex visual patterns, with some degree of tolerance to identity-preserving transformations. Note, however, that the face images in the current study were carefully aligned to remove obvious cues to identity caused by misalignment, and that this alignment may amplify the extent to which relatively low-level image-based differences can be used for identity discrimination, even across changes in facial expression. Hence, it is not surprising that we were able to decode facial identity across changes in facial expression, even in left V1. That decoding accuracy reached an initial peak between 100 and 200 ms is consistent with previous findings that the N170 component in EEG and the corresponding M170 component in MEG are correlated with characteristics of individual face identities (16–18). In addition, the finding that decoding performance remained above chance at several time points after 200 ms is consistent with findings from intracranial recordings from fusiform gyrus that facial identity could be decoded between 200 and 500 ms after stimulus onset (15). Also, the finding of secondary peaks in classification accuracy in the face-selective regions at around 250 ms seems to be compatible with data showing that subordinate-level processing of faces selectively enhances the N250 ERP component in EEG, which peaks ∼250 ms after stimulus onset (32, 33). In sum, our decoding analysis based on spatiotemporal patterns of activity captured many existing findings in the EEG and MEG literatures, while allowing further analyses of the temporal dynamics of the similarity structure of the neural representation of facial identity.

A particularly compelling aspect of our results is the transition between image-based and identity-based representations observed after 200–300 ms in the face-selective regions, but not in the control region (lV1), and the relation between the

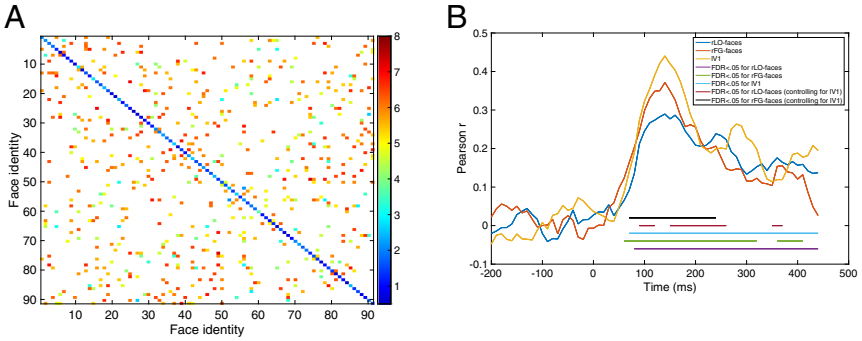

Fig. 6. (*A*) Heat map showing mean (across seven participants) pairwise cross-expression behavioral dissimilarity ratings for a subset of face pairs presented during the MEG experiment. Hotter colors indicate greater dissimilarity. Face pairs presented during the MEG experiment but not during the behavioral experiment are shown in white. (*B*) Correlations between behavioral and neural data, as a function of time (milliseconds). Horizontal lines indicate postbaseline time points at which correlations were statistically significant (FDR < 0.05) and those at which face-selective regions predicted behavioral responses after controlling for responses in the control region (FDR < 0.05).

timing of this transition and that of classification accuracy. In rLO-faces, there were two obvious peaks in which decoding accuracy exceeded 70%, one between 100 and 200 ms and the second between 200 and 300 ms. At the first peak, the similarity structure of the neural data was more similar to an image-based representation, whereas at the second peak the correlation with the image-based representation dropped, so that the neural data were more similar to the identity-based representation. This pattern was not observed in the control region (lV1), because the data were more similar to the image-based representation than to the identity-based representation at all time points. Previous studies have demonstrated that signals in occipitotemporal regions between 100 and 200 ms are related to image-based properties of faces (16–18). However, the transition toward an identity-based representation within face-selective cortical regions after 200 ms has not been observed in previous studies of discrimination of facial identity, because previous studies either used fMRI (3–13), which lacks the temporal resolution required to resolve the temporal patterns observed in the current study, and/or because their analyses of the neural representations were based on different sensors for different time points (18), and/or were limited to image-based properties of the stimuli (15, 17, 18). The observed pattern provides evidence that spatiotemporal patterns of activity in at least some face-selective regions in human cortex encode qualitatively different information about face identity at different times over the first few 100 ms after stimulus onset, with a transition from a lower level representation to a higher level representation occurring around 200–300 ms. This pattern suggests that models of the neural basis of face recognition that assign a single function to each face-selective cortical region (14) are likely to be incomplete, because they do not account for the possibility that a given face-selective region may play different functional roles at different times. Given that the identity-based representation used in the current study represents any exemplar of the same identity identically, whereas the image-based representation does not, this late transition could reflect the temporal emergence of a neural representation with high tolerance to identity-preserving transformations (19). Such a representation is highly relevant for real-world behavior, because many situations require the system to track a single identity and/or discriminate between identities across identity-preserving image transformations. Further support for the behavioral relevance of this representation comes from our finding that pairwise behavioral judgments of the stimuli were more strongly correlated with the identity-based representation than with the image-based representation, and that responses in the face-selective regions predicted behavior during temporal periods in which responses in the face-selective regions transitioned toward an identity-based representation. Given that the observed transition occurs relatively late, and that it corresponds with a secondary later peak in the classification accuracy function, it seems somewhat unlikely that it arises as a consequence of a single initial feedforward sweep (34) but instead reflects recurrent/feedback processing (19, 35).

One remaining question is how the face-selective regions identified in the current study (rLO-faces and rFG-faces) are related to the corresponding face-selective regions typically identified in fMRI studies of face processing: occipital face area (OFA) and fusiform face area (FFA). Given that all source points within rLO-faces and rFG-faces in the current study are face-selective and are located within the same anatomical subregions as OFA and FFA, respectively, it seems possible that their representations would overlap with those of OFA and FFA. However, at least two differences are likely to limit the degree of overlap. First, MEG and fMRI measure different aspects of neural activity and have different spatial signal distributions, and are therefore likely to be sensitive to different spatial patterns of activity. For example, MEG is less sensitive to signals from deeper and more gyral sources than it is to more superficial and sulcal sources (36). Given that the rFG-faces region used in the current study is deeper and more gyral than rLO-faces, it seems possible that rLO-faces would capture signals from the corresponding fMRI-defined region to a greater extent than rFG-faces. This could account for the lower sensitivity to face identity observed in rFG-faces than in rLO-faces. Second, our MEG data have rich temporal structure, but fMRI data do not, and so the MEG data are unlikely to correspond to the fMRI data at all time points. Hence, although it is possible that the representations measured from rLO-faces and rFG-faces in the current study reflect activity from OFA and FFA, they are likely to reflect different aspects of the representation of facial identity than those standardly measured in OFA and FFA with fMRI.

Taken together, our results provide important information about when spatiotemporal patterns in face-selective cortical regions discriminate among a large and carefully controlled set of face identities across changes in facial expression, and about what type of information is represented in each region, and when. Specifically, our results indicate spatiotemporal patterns of activity in both face-selective and control regions encode information about facial identity between ∼50 and 400 ms after stimulus onset. However, the face-selective regions, but not the control region, seem to encode qualitatively different information about facial identity at different times, with a transition from an image-based representation toward an identity-based representation after 200–300 ms. As described above, these results have implications for understanding the microgenesis of fine-grained, high-level neural representations of object identity, a process critical to human visual expertise (19), and perhaps for distinguishing between feedforward versus recurrent/feedback accounts of visual processing. Overall, the current investigation represents a critical advancement toward understanding the temporal dynamics of visual pattern recognition in the human brain.

Materials and Methods

Participants. All participants were Caucasian (white European), right-handed, and had normal or corrected-to-normal visual acuity and no history of eye problems. Participants in the MEG experiment were four adults (one female), aged 23–27 y. Participants in the behavioral experiment were seven adult humans (five female), aged 18–28 y, none of whom participated in the MEG experiment. No participants were tested but excluded. Protocols were approved by institutional review boards at Carnegie Mellon University and the University of Pittsburgh. All participants provided written informed consent before each session and received monetary compensation for their participation.

MEG.

Localizer task. Each participant completed four or five 3.5-h MEG sessions. During the final MEG session, each participant completed a block design category localizer adapted from an existing fMRI localizer used in previous work (12) (*Supporting Information*). The localizer data were used to identify source points that responded significantly more strongly to faces than to objects (FDR < 0.05) within two anatomical regions defined with an atlas in Freesurfer (30): right lateral occipital cortex and right fusiform gyrus (see Fig. 3 and see *Supporting Information* for details).

Face identity task. In each block, participants viewed each of the 91 face identities four times (twice per expression). Participants were instructed to maintain fixation and to respond whenever they saw the same face identity repeated, regardless of facial expression (*Supporting Information*). During each of the MEG sessions except for the last part of the final MEG session participants performed this task while MEG signals were recorded. Each participant completed between 26 and 28 blocks of the task.

MEG data acquisition and processing. MEG data were acquired at the University of Pittsburgh Medical Center Brain Mapping Center, with a 306-channel Neuromag (Elektra AB) system. Data were preprocessed using both spatial and temporal filtering approaches. Each participant's MEG data were then projected onto a cortical surface reconstructed from their anatomical MRI scan (*Supporting Information*).

Classification of facial identity. For each region and time point we used a binary k-nearest-neighbor classifier to classify all possible pairs of facial identities across a change in facial expression. We computed a statistical threshold for accuracy by shuffling the labels in the neural data. Accuracy is significantly greater than chance where observed values exceed this threshold (*Supporting Information*).

Comparison with model-based representations. For each brain region and time point, we measured the correlation between the neural data and each of two model-based representations: an image-based representation based on low-level visual information and an identity-based representation that was sensitive to face identity but was not sensitive to image-based information (Fig. 2 and *Supporting Information*). We then compared the correlation between the two types of representations to examine which type was more similar to the neural data (31).

Behavioral Similarity Ratings. In each of two 1-h sessions, participants viewed a subset of pairs of faces from the stimulus set used in the MEG experiment (Fig. 6) and rated the similarity of the face identities on an 8-point scale, with a value of 1 indicating very different face identities and a value of 8 indicating the same face identity. Within a pair, face images always differed in facial expression (*Supporting Information*). We used the methods described above to compare the behavioral data to the neural data and model-based representations. To examine whether neural data from the face-selective regions could predict the behavioral data after controlling for responses in lV1, we fit a multiple linear regression model in which the distance values for lV1 and a face-selective region were used to predict the behavioral data (37).

ACKNOWLEDGMENTS. This work was supported by Natural Sciences and Engineering Research Council PDF Award 471687-2015 (to M.D.V.), a Small Grant from the Temporal Dynamics of Learning Center (to M.D.V. and M.B.), Pennsylvania Department of Health's Commonwealth Universal Research Enhancement Program Grant SAP-14282-012 (to D.C.P.), National Science Foundation Grant BCS0923763 (to M.B. and D.C.P.), and Temporal Dynamics of Learning Center Grant SMA-1041755 (principal investigator: G. Cottrell) (to M.B.).

1. Barragan-Jason G, Besson G, Ceccaldi M, Barbeau EJ (2013) Fast and Famous: Looking for the fastest speed at which a face can be recognized. *Front Psychol* 4(4):100.
2. Ramon M, Caharel S, Rossion B (2011) The speed of recognition of personally familiar faces. *Perception* 40(4):437–449.
3. Anzellotti S, Fairhall SL, Caramazza A (2014) Decoding representations of face identity that are tolerant to rotation. *Cereb Cortex* 24(8):1988–1995.
4. Axelrod V, Yovel G (2015) Successful decoding of famous faces in the fusiform face area. *PLoS One* 10(2):e0117126.
5. Cowen AS, Chun MM, Kuhl BA (2014) Neural portraits of perception: Reconstructing face images from evoked brain activity. *Neuroimage* 94:12–22.
6. Gao X, Wilson HR (2013) The neural representation of face space dimensions. *Neuropsychologia* 51:1787–1793.
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* 39(19):8549–8558.
8. 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.
9. Natu VS, et al. (2010) Dissociable neural patterns of facial identity across changes in viewpoint. *J Cogn Neurosci* 22(7):1570–1582.
10. 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.
11. Nestor A, Behrmann M, Plaut DC (2013) The neural basis of visual word form processing: A multivariate investigation. *Cereb Cortex* 23(7):1673–1684.
12. Nestor A, Plaut DC, Behrmann M (2016) Feature-based face representations and image reconstruction from behavioral and neural data. *Proc Natl Acad Sci USA* 113(2):416–421.
13. Verosky SC, Todorov A, Turk-Browne NB (2013) Representations of individuals in ventral temporal cortex defined by faces and biographies. *Neuropsychologia* 51(11):2100–2108.
14. Freiwald W, Duchaine B, Yovel G (2016) Face processing systems: From neurons to real-world social perception. *Annu Rev Neurosci* 39:325–346.
15. Ghuman AS, et al. (2014) Dynamic encoding of face information in the fusiform gyrus. *Nat Commun* 5:5672.
16. Liu J, Harris A, Kanwisher N (2002) Stages of processing in face perception: An MEG study. *Nat Neurosci* 5(9):910–916.
17. Rousselet G, Hannah H, Ince R, Schyns P (2015) The N170 is mostly sensitive to pixels in the contralateral eye area. *J Vis* 15:687.
18. Zheng X, Mondloch CJ, Nishimura M, Vida MD, Segalowitz SJ (2011) Telling one face from another: Electrocortical correlates of facial characteristics among individual female faces. *Neuropsychologia* 49(12):3254–3264.
19. DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. *Trends Cogn Sci* 11(8):333–341.
20. Sugase-Miyamoto Y, Matsumoto N, Kawano K (2011) Role of temporal processing stages by inferior temporal neurons in face recognition. *Front Psychol* 2:141.
21. Freiwald WA, Tsao DY (2010) Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330(6005): 845–851.
22. Sugase-Miyamoto Y, Yamane S, Ueno S, Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400(6747):869–873.
23. Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science* 311(5761):670–674.
24. Braun C, Schweizer R, Elbert T, Birbaumer N, Taub E (2000) Differential activation in somatosensory cortex for different discrimination tasks. *J Neurosci* 20(1):446–450.
25. Freeman TC, Fowler TA (2000) Unequal retinal and extra-retinal motion signals process different perceived slants of moving surfaces. *Vision Res* 40(14):1857–1868.
26. Isik L, Meyers EM, Leibo JZ, Poggio T (2014) The dynamics of invariant object recognition in the human visual system. *J Neurophysiol* 111(1):91–102.
27. Cichy RM, Ramirez FM, Patazis D (2015) Can visual information encoded in cortical columns be decoded from magnetoencephalography data in humans. *Neuroimage* 121:193–204.
28. Tottenham N, et al. (2009) The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Res* 168(3):242–249.
29. Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2(11):1019–1025.
30. Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53: 1–15.
31. Steiger JH (1980) Test for comparing elements of a correlation matrix. *Psychol Bull* 87(2):245–251.
32. Scott LS, Tanaka JW, Sheinberg DL, Curran T (2006) A reevaluation of the electrophysiological correlates of expert object processing. *J Cogn Neurosci* 18(9):1453–1465.
33. Tanaka JW, Curran T, Porterfield AL, Collins D (2006) Activation of preexisting and acquired face representations: The N250 event-related potential as an index of face familiarity. *J Cogn Neurosci* 18(9):1488–1497.
34. Serre T, Oliva A, Poggio T (2007) A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci USA* 104(15):6424–6429.
35. Wyatte D, Jilk DJ, O'Reilly RC (2014) Early recurrent feedback facilitates visual object recognition under challenging conditions. *Front Psychol* 5:674.
36. Hillebrand A, Barnes GR (2002) A quantitative assessment of the sensitivity of whole-head meg to activity in the adult human cortex. *Neuroimage* 16(3 Pt 1):638–650.
37. Freckleton RP (2002) On the misuse of residuals in ecology: Regression of residuals vs. multiple regression. *J Anim Ecol* 71(3):542–545.
38. Gramfort A, et al. (2013) MEG and EEG data analysis with MNE-Python. *Front Neurosci* 7:267.
39. Gramfort A, et al. (2014) MNE software for processing MEG and EEG data. *Neuroimage* 86:446–460.
40. Dale AM, et al. (2000) Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26 (1):55–67.
41. Gross R, Matthews I, Cohn J, Kanade T, Baker S (2010) Multi-PIE. *Proc Int Conf Autom Face Gesture Recognit* 28:807–813.
42. Langner O, et al. (2010) Presentation and validation of the Radboud Faces Database. *Cognit Emot* 24(8):1377–1388.
43. Goeleven E, De Raedt R, Leyman L, Verschuere B (2008) The karolinska directed emotional faces: A validation study. *Cognit Emot* 22(6):1094–1118.
44. Martinez AM, Benavente R (1998) The AR face database. CVC Tech Rep 24.
45. Delorme A, Makeig S (2004) EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. *J Neurosci Methods* 134(1):9–21.
46. Dunn OJ (1961) Multiple comparisons among means. *J Am Stat Assoc* 56(293): 52–64.
47. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Roy Stat Soc* 51(1):289–300.
48. Fukushima K (1980) Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* 36: 193–202.
49. Hinton GE (1989) Connectionist learning procedures. *Artif Intell* 40:185–234.
50. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323:533–536.

Supporting Information

Vida et al. 10.1073/pnas.1614763114

MEG

Acquisition. To allow correction of eye movement artifacts, we recorded the electrooculogram (EOG) from four electrodes. To allow correction of heartbeat artifacts, we recorded the ECG from a pair of electrodes. Four head position indicator (HPI) coils were used to monitor the participant's head position within the MEG helmet. At the beginning of each session, a digitizing pen was used to record the shape of the participant's head and the locations of the HPI coils on the head in 3D space. For all participants, head position was recorded from the HPI coils at the beginning of each block. For participants 1, 3, and 4, continuous HPI (cHPI) signals were recorded during each block, to allow movement compensation during preprocessing. cHPI signals were not recorded for participant 2 because enabling cHPI recording seemed to produce excessive artifacts. A Panasonic PT-D7700U projector (1,024 × 768 resolution, 60-Hz refresh rate) presented the stimuli at the center of a back-projection screen placed 120 cm from the participant. Face images were 6.87° high and 5° wide at this viewing distance. To track stimulus timing, we used a photodiode that emitted a continuous signal when the stimulus was on the screen. In addition, the experimental software sent a signal to the MEG acquisition computer whenever a stimulus was presented. Participants entered responses by pressing a button with their right index fingers.

Preprocessing. We first applied signal source separation in Maxfilter (Elekta AB), with head movement compensation enabled where applicable. The temporal extension (tSSS) was enabled for subject 4, for whom this extension was required to remove an artifact caused by an orthodontic appliance. tSSS was also enabled for the localizer data in subject 3, for whom this extension significantly boosted signal. We then carried out temporal filtering and artifact rejection and correction in MNE-Python (38) and applied a band-pass filter with lower and upper cutoffs set to 1 and 100 Hz, respectively. To remove power line noise, we applied notch filters at 60, 120, and 180 Hz. Empty room data were used to create signal space projectors, which were applied to the filtered raw data to remove environmental artifacts (39). To correct eye movement and heartbeat artifacts, we used MNE-Python to fit an independent components analysis model to the ECG and EOG data and remove components correlated with the MEG data. Finally, trials with signals exceeding standard thresholds (gradiometer = $4{,}000e^{-13}$, magnetometer = $4e^{-12}$) in at least one channel were rejected (38). For the main face identity task, at least 101, 92, 101, and 97 trials per face identity were retained for participants 1–4, respectively.

Source Localization. For each participant, we acquired a T1-weighted MPRAGE anatomical MRI scan on a Siemens Verio 3T scanner (voxel size 1 mm^3, flip angle 9°, TE = 1.97 ms, TR = 2,300 ms, FOV = 256 × 256 × 176 mm). All scans were carried out at the Scientific Imaging and Brain Research Center at Carnegie Mellon University. Freesurfer reconstructions based on the anatomical scan were used in source modeling of MEG signals.

For each participant, we used dynamic statistical parametric mapping (dSPM) (40) to project single-trial MEG data onto the cortical surface reconstructed from each participant's MRI data. This approach allowed us to align and combine each participant's data across sessions and allowed us to restrict our analyses to signals estimated to originate in specific regions on the cortical surface. We used the MNE watershed tool (39) and Freesurfer output to generate boundary element models (BEMs). In mne-analyze, we used the digitizer data and the BEMs to align each participant's MEG data to their MRI anatomical image. In MNE, we generated a source space on the reconstructed cortical surface, with source points spaced 5 mm apart. We then generated a forward solution, which maps the MEG sensor space to the source space (39). Using this forward solution, we performed dSPM source localization in MNE-Python [single-trial data, noise covariance matrix calculated from baseline period (−200 to 0 ms), signal-to-noise ration 3, no depth prior] (38, 39). Note that the estimates of activity provided by our inverse model do not directly measure neural activity, but instead represent a probability density function for activity across the source space. Because the inverse model provides an estimate of activity at a given point in the source space, we use "activity" to refer to this estimate throughout the paper, for simplicity.

Face Identity Task

Stimuli. Stimuli were color frontal photographs of 91 young Caucasian (white European) male face identities selected from six databases (28, 41–44), including PICS (see Fig. 1 for examples). The following male identities from the Karolinska Directed Emotional Faces database were included: 2, 3, 5, 6, 7, 10, 11, 12, 13, 17, 18, 21, 23, 24, 25, 26, 27, 28, 31, 32, and 35. The stimulus set included two expressions (neutral or happy) per identity. To eliminate global luminance and color cues, we converted each image to CIELAB color space and set the mean of the L^* (luminance), a^* (red–green), and b^* (yellow–blue) channels to the mean values across all identities. We also set the rms contrast of the L^* channel to the mean across all identities. To minimize differences in alignment, face images were transformed without altering aspect ratio, so that the eyes were in the same positions in each image. To eliminate hair cues, we applied an oval mask of constant size to each face image. To minimize other obvious cues, we excluded faces with facial hair, uneven lighting, and/or aspect ratios (defined here as the ratio of the horizontal interocular distance to the vertical eye–mouth distance) greater than two standard deviations from the group mean.

Design. Before the MEG experiment, each participant first completed a single 1-h session of behavioral training on the task that they would complete during subsequent MEG sessions. In each block of training, participants viewed each of the 91 face identities four times (twice per expression), for a total of 364 trials. Participants were instructed to maintain fixation and to press a key on a computer keyboard whenever they saw the same face identity repeated, regardless of facial expression. During each of the MEG sessions except for the last part of the final MEG session participants did the same behavioral task as in the training session while MEG signals were recorded.

The order of trials was randomized for each block, with the exception that the same face identity was presented twice in a row on 36 trials per block. The two face images had the same expression on half of these repeat trials and had different expressions on the other half. The positions of the repeat trials within the block and the face identities to be presented on these repeat trials were randomly selected for each block. Each trial began with a white fixation dot presented for 500 ms, followed by a face image presented for 500 ms, then a blank response screen for 1,500 ms.

Localizer Task. Each run of the MEG localizer consisted of 15 blocks, with a fixation baseline (8 s) between blocks. Within each run, there were three blocks for each of five categories of images (faces, objects, scrambled objects, houses, and words) presented in a random order. Within each block, 16 images from a single category were presented in a row (900 ms per image, 100-ms interstimulus interval), in a random order. Each participant completed six runs, for a total of 288 trials per category. Participants were instructed to press a button on the response glove with their right index fingers to indicate the presence of one repeated image within each block.

To identify face-selective source points in each participant, we used the statcond function in EEGLAB (45) to carry out a nonparametric, one-way permutation test on single-trial data for each participant, source point, and postbaseline time point (10,000 permutations per test). Each test yielded a P value indicating the extent to which activity differed between faces and objects. For each participant, P values were FDR-corrected across all source points and time points. A source point was considered to be face-selective if it responded significantly more strongly to faces than to objects (FDR < 0.05) at one or more time points and did not respond significantly more strongly to objects than to faces at any time point.

Classification of Face Identity. For each brain region, time point, participant, and face identity, we extracted the spatiotemporal pattern of neural activity across all source points within the region and across all time points between the current time point and a time point 60 ms after the current time point. We then computed the pairwise Euclidean distance for each possible pair of identities. To ensure that, to the extent possible, this analysis captured information invariant to changes in facial expression (i.e., were specific to identity), all Euclidean distance values were computed across a change in facial expression. To increase power, we averaged Euclidean distance values across participants.

For a given time point, region of interest, and pair of face identities, inputs to the k-nearest-neighbor classifier were the cross-expression within-identity Euclidean distance value for the each of the two identities, and the cross-expression, cross-identity distance value. When spatiotemporal patterns of activity contain sufficient information to distinguish between face identities across a change in facial expression, the cross-identity distance will be larger than the within-identity distance. Hence, for each face in the pair, the classifier compared the within-identity distance to the cross-identity distance and decided that the smaller of the two (i.e., the nearest neighbor) was the within-identity distance. Overall classification accuracy was computed as the proportion of correct decisions across all face pairs. Because each decision is binary, accuracy fluctuates around 0.5 (50% correct) when there is no signal (i.e., in the baseline period; Fig. 4).

To evaluate whether whether classification accuracy was significantly greater than chance, we computed a statistical threshold for accuracy [Bonferroni-corrected alpha = 0.001 (46)] from a null distribution with 10,000 values. Each value in the null distribution was generated by shuffling the labels in the neural data and recomputing classification performance. To allow for the possibility that different regions might have different statistical properties, and therefore different null distributions, we computed statistical thresholds separately for each region. The threshold represents the accuracy level associated with a probability of 0.001 under the null hypothesis that the neural data do not reliably encode information about facial identity.

Comparison with Model-Based Representations. To generate an image-based representation of the stimuli, we first converted each face image to $L^*a^*b^*$ color space and extracted the luminance component of the image. We then extracted activations for each image from the first layer (S1) of HMAX, which simulates responses of a population of simple cells in V1 (29). We used the vector of activations for each image to compute a cross-expression Euclidean distance matrix of the same form used for the classification analysis described above. The identity-based representation was a 91×91 matrix in which each row and column represents a facial identity, and each cell represents a comparison between two facial identities (Fig. 2). All cells representing comparisons within identities (e.g., row 1, column 1) received a distance value of 0, which indicates identical representations. The remaining cells, which represent comparisons between identities, received a distance value of 1. This value indicates that representations of exemplars of different identities differ to the same extent.

For each brain region and postbaseline time point (from 0 ms onward) we computed the correlation between the cross-expression distance matrix for the neural data and that for each of the representations described above. We excluded the baseline period (-200 to 0 ms) from this analysis because, during this period, classification accuracy did not exceed chance, and so there was no evidence that information about face identity was encoded. We then tested whether the correlation was stronger for the identity-based representation or the image-based representation (31). The resulting P values for each region were FDR-corrected (47) across all time points.

Behavioral Similarity Ratings. In each of two 1-h sessions each participant provided similarity ratings for 377 face pairs. These face pairs included all 91 within-identity pairs and an additional 286 between-identity pairs. The latter were randomly selected from among all possible between-identity face pairs. To allow aggregation of data across participants and sessions, we used the same procedure and face pairs for each participant and session. In each session, participants rated each face pair twice. For each face pair, the two face images were always presented with different facial expressions, with all possible combinations of facial expressions presented within each session. On each trial, a face pair was presented sequentially, in a random order. Each face image was presented for 500 ms, with a 500-ms interval between images.

To allow comparisons between behavioral similarity ratings and neural and model-based representations, we converted similarity ratings to dissimilarity ratings by subtracting each similarity value from 8, so that a similarity rating of 1 was converted to a dissimilarity rating of 8, and so on.

Left Hemisphere. In the left hemisphere, we carried out the primary analyses performed for face-selective regions in the right hemisphere. In three of four participants we were able to identify face-selective regions in left lateral occipital cortex (lLO-faces) and fusiform gyrus (lFG-faces). As with the right hemisphere analyses, we used V1 in the opposite hemisphere (in this case, right V1) as a control region. With data from these regions, we performed classification of face identity and compared the neural data to model-based representations and behavioral judgments of the stimuli. Results for analyses of face-selective regions in the left hemisphere are shown in Fig. S1. All analyses of neural data include only data from the three participants in which face-selective regions could be identified in the left hemisphere.

The positions of face-selective regions in the left hemisphere seemed to vary between participants to a greater extent than the corresponding regions in the right hemisphere. Classification of face identity was successful in both left hemisphere regions, but the pattern of correlations with the model-based representations seemed to be less temporally stable than in the right hemisphere. In the right hemisphere regions, all significant differences between the models favored the image-based representation before 200 ms and favored the identity-based representation after 200 ms. In contrast, differences in the left hemisphere

fluctuated between the two representations after 200 ms, a result suggesting that the transition toward the identity-based representation after 200 ms may be less stable in the left hemisphere. Similar to lV1, significant differences between the two representations in rV1 consistently favored the image-based model, with the exception of a very early period during which classification performance in rV1 had only just begun to exceed chance. This pattern suggests that V1 primarily encoded relatively low-level image-based information in each hemisphere.

Comparison of Neural Data to Deep Neural Network. To understand the nature of the information represented in each region of interest we compared the similarity structure of the neural representations to those learned by a deep neural network trained to recognize versions of the experimental face stimuli. Input to the network consisted of 8,918 48 × 65 grayscale images of faces. Forty-nine versions of each of 182 original face images (91 individuals × 2 expressions) were created by jittering the position of the image up to ±3 pixels in both the *x* and *y* directions. For each such input the network was trained to activate a particular 1 of 91 output units.

The architecture of the network had four intermediate or "hidden" layers between the input and output, connected in a feedforward manner. The first hidden layer (H1) had 2,684 (44 × 61) units with 5 × 5 rectangular receptive fields (RFs) from the input with a stride of 1 (i.e., RFs were positioned every one input unit in both the *x* and *y* directions). The second hidden layer (H2) had 580 (20 × 29) units with 7 × 7 RFs from H1 with a stride of 2, and the third (H3) had 84 (7 × 12) units with 9 × 9 RFS from H2 with a stride of 2. The fourth and final hidden layer (H4) consisted of 20 units that received connections from all H3 units and sent connections to all 91 output units. In total, the network had 108,570 connections (including bias connections to all noninput units). Unlike a convolutional neural network (29, 48), positional invariance was not imposed within layers (by using weight-sharing among features and having separate "pooling" layers); rather, it was left to the network to learn to be sensitive or insensitive to positional information at each level of representation to the degree it supports effective recognition. Hidden units used a sigmoid activation function to their net inputs; the output group consisted of normalized or "soft-max" units whose activities were constrained to sum to 1.0 (49).

The network was trained with back-propagation (50) using momentum descent with accumulated gradients clipped at 1.0, with a learning rate of 0.05 and momentum of 0.8. After 6,000 presentations of each image, performance had reached near-asymptote: The network produced virtually no error (mean cross-entropy error per image of 0.00013) and recognition performance was perfect.

Representational similarity analysis was then carried out for each layer of the network by computing the degree of similarity (correlation) of the activation patterns produced by each pair of face images (without jitter). These similarities were then correlated with analogous similarities for the neural representations in various brain regions across time. We focused on H1 and H4, because the former largely reflects visual (image) similarity whereas the latter reflects primarily identity information. Indeed, for H4, the mean correlation is 0.097 for different identities and 0.939 for same identities. Although H1 and H4 seem to primarily represent image and identity information, respectively, the representations from these regions are likely to be more closely related to each other than the image-based and identity-based representations described in the main text, because layers H1 and H4 are both trained to perform the same task. Indeed, the correlation between H1 and H4 (Pearson $r = 0.46$) was slightly higher than that between the identity- and image-based representations in the main text (Pearson $r = 0.40$).

Fig. S2 shows the correlations over time of each hidden layer with the two face-selective regions (rLO-faces and rFG-faces) and the control region (lV1), considering only image pairs that differ in expression. For rLO-faces the correlation with H4 was stronger than the correlation with H1 at a series of time points between 50 and 100 ms, and a later series between 200 and 300 ms. In contrast, lV1 responses were more strongly correlated with H1 than H4 between 350 ms and 450 ms, with no significant differences in the opposite direction. In rFG, the correlation with H4 was stronger at only one early period (60–90 ms). Hence, as in the comparison with image-based and identity-based representations in the main text, responses in rLO-faces seem to be more similar to a higher level representation, whereas responses in lV1 seem to be more similar to a lower level, more image-based representation. rFG seemed to show an intermediate response in the current analyses, as it did for most time points in the analyses presented in the main text.

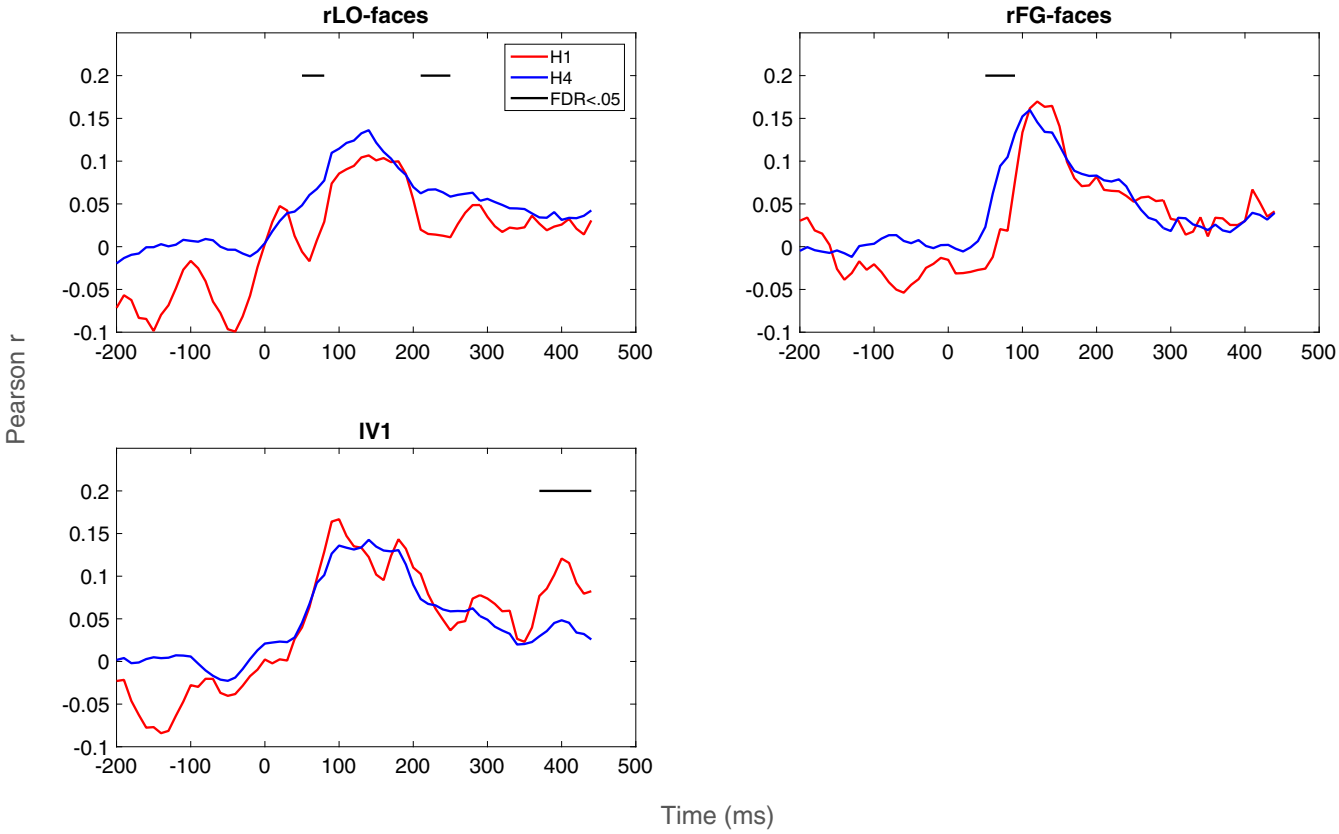

Fig. S1. Correlations between neural data and hidden layers 1 (H1) and 4 (H4) of a deep neural network trained on our stimuli, as a function of time (milliseconds). Separate plots are given for each region of interest (rLO-faces, rFG-faces, and lV1). The horizontal black line indicates postbaseline time points at which correlations differed significantly (FDR < 0.05) between H1 and H4.

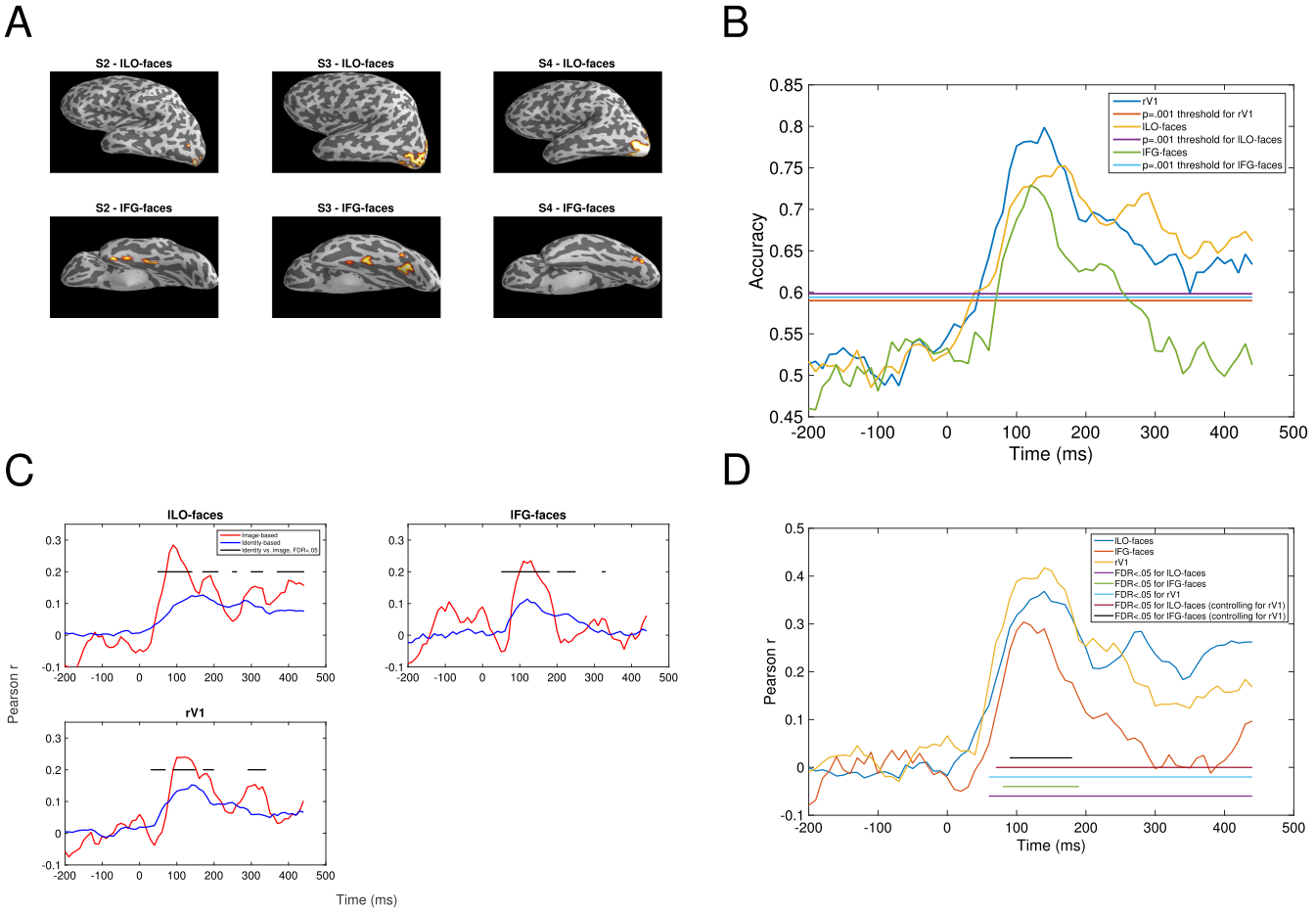

Fig. S2. Analyses for left hemisphere. (*A*) Face-selective regions in left hemisphere, for each participant in which regions could be identified. (*B*) Classification accuracy as a function of time (milliseconds) and region (lV1, lLO-faces, and lFG-faces). All other details as in Fig. 4. (*C*) Correlations between neural data and image-based and identity-based representations of our stimuli, as a function of time (milliseconds). All other details as in Fig. 5. (*D*) Correlations between behavioral and neural data, as a function of time (milliseconds). All other details as in Fig. 6*B*.