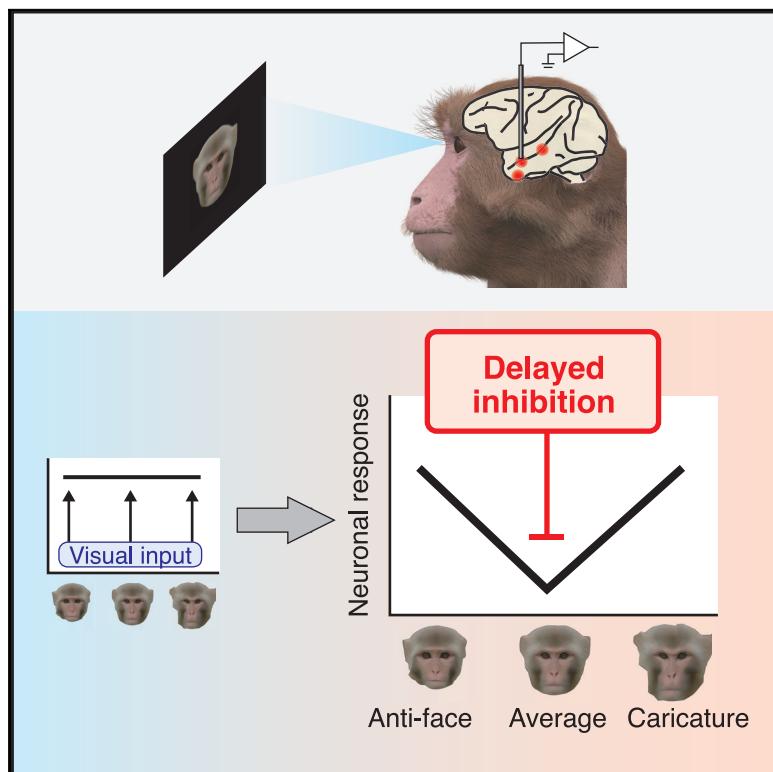


## Dynamic Suppression of Average Facial Structure Shapes Neural Tuning in Three Macaque Face Patches

### Graphical Abstract



### Authors

Kenji W. Koyano, Adam P. Jones,  
David B.T. McMahon,  
Elena N. Waidmann, Brian E. Russ,  
David A. Leopold

### Correspondence

kenji.koyano@nih.gov (K.W.K.),  
leopoldd@mail.nih.gov (D.A.L.)

### In Brief

Koyano et al. use morphed faces to examine neural tuning to facial identity in three macaque face patches. Neurons respond with a prominent tuning around average facial structure, expressed as a delayed suppression to low-identity faces. The findings suggest a mechanism by which the brain efficiently extracts individuals' unique facial features.

### Highlights

- Neurons in face patches showed robust identity tuning around average facial structure
- The tuning emerged as delayed suppression to low-identity faces
- The coincidence of the suppression and spike synchrony suggests active inhibition
- Control experiments confirmed that the tuning was not due to short-term adaptation



## Article

# Dynamic Suppression of Average Facial Structure Shapes Neural Tuning in Three Macaque Face Patches

Kenji W. Koyano,<sup>1,5,\*</sup> Adam P. Jones,<sup>1,5</sup> David B.T. McMahon,<sup>1,2</sup> Elena N. Waidmann,<sup>1,6</sup> Brian E. Russ,<sup>1,3</sup> and David A. Leopold<sup>1,4,7,\*</sup>

<sup>1</sup>Section on Cognitive Neurophysiology and Imaging, National Institute of Mental Health, 49 Convent Dr., Bethesda, MD 20892, USA

<sup>2</sup>Neuronal Networks Section, National Eye Institute, 49 Convent Dr., Bethesda, MD 20892, USA

<sup>3</sup>Center for Biomedical Imaging and Neuromodulation, Nathan S. Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA

<sup>4</sup>Neurophysiology Imaging Facility, National Institute of Mental Health, National Institute of Neurological Disorders and Stroke, National Eye Institute, 49 Convent Dr., Bethesda, MD 20892, USA

<sup>5</sup>These authors contributed equally

<sup>6</sup>Present address: The Rockefeller University, 1230 York Avenue, Box 386, New York, NY 10065, USA

<sup>7</sup>Lead Contact

\*Correspondence: kenji.koyano@nih.gov (K.W.K.), leopoldd@mail.nih.gov (D.A.L.)

<https://doi.org/10.1016/j.cub.2020.09.070>

## SUMMARY

The visual perception of identity in humans and other primates is thought to draw upon cortical areas specialized for the analysis of facial structure. A prominent theory of face recognition holds that the brain computes and stores average facial structure, which it then uses to efficiently determine individual identity, though the neural mechanisms underlying this process are controversial. Here, we demonstrate that the dynamic suppression of average facial structure plays a prominent role in the responses of neurons in three fMRI-defined face patches of the macaque. Using photorealistic face stimuli that systematically varied in identity level according to a psychophysically based face space, we found that single units in the AF, AM, and ML face patches exhibited robust tuning around average facial structure. This tuning emerged after the initial excitatory response to the face and was expressed as the selective suppression of sustained responses to low-identity faces. The coincidence of this suppression with increased spike timing synchrony across the population suggests a mechanism of active inhibition underlying this effect. Control experiments confirmed that the diminished responses to low-identity faces were not due to short-term adaptation processes. We propose that the brain's neural suppression of average facial structure facilitates recognition by promoting the extraction of distinctive facial characteristics and suppressing redundant or irrelevant responses across the population.

## INTRODUCTION

Humans and other primates rely on vision to perceive important social information, such as the identity of individuals in a group, their actions, and their emotional states.<sup>1</sup> Faces play a particularly important role in primate social perception, which is reflected in the multiple face-specialized regions in the temporal cortex of all primates studied to date.<sup>2–4</sup> Recognizing an individual by their unique facial structure is a notoriously difficult visual problem, in part because it requires a robust analysis of subtle facial features, including their volumetric shapes, colorations, and geometric configuration. Human psychophysical experiments indicate that the brain, in approaching this problem, utilizes an internally stored norm approximated by the average face. Theories of face recognition posit that the brain compares a given face's structural characteristics with this internal reference, which facilitates recognition through the efficient extraction of facial characteristics that are unique to an individual.<sup>5–8</sup> According to this view, the brain builds its average prototype

gradually, particularly during early life, by sampling the natural statistical variation in facial structure. Adherents to this framework typically make reference to a theoretical “face space,” where individual identity is conceived as a radial trajectory away from the average face in the center.<sup>9</sup> This perspective on face encoding has been important for understanding key elements of human face recognition revealed through psychophysical experiments.<sup>6,10,11</sup> The formalization of face space has also provided a straightforward framework for generating stimuli, using morphing or computational algorithms, to investigate the brain's encoding of identity.

One particularly important category of trajectories in face space are the “identity trajectories,” which radiate outward from the average face in the center and pass through the points corresponding to real individual faces. Identity trajectories are so named because each represents a unique set of structural transformations corresponding to an individual. Moving outward along an identity trajectory exaggerates a face's unique features, eventually creating a caricature for positions more eccentric than



the original face. Moving inward toward the average diminishes an individual's distinctive features. Extrapolating an identity trajectory to the opposite side the average face reverses the identity transformation and creates a face identity with nominally opposite features, sometimes called an anti-face.<sup>12</sup> Identity trajectories within face space all intersect at the average face, located at the center. Making an analogy to color space, faces along different identity trajectories correspond to different hues, and their eccentricities correspond to different saturation levels. Importantly, identity trajectories have been identified as critical psychophysical dimensions for recognition, for example, underlying the efficacy of caricatures and the expression of face identity aftereffects.<sup>6,10,11</sup>

While much psychophysical evidence supports the brain's utilization of an average for face recognition, researchers are divided on its potential implementation at the neural level.<sup>13–19</sup> An early electrophysiological study from our group found that most face-selective neurons in the macaque inferior temporal cortex exhibited graded responses to morphed human faces and responded most weakly to the average face.<sup>17</sup> This finding was interpreted as evidence that the brain codes face identity relative to an internal representation of average facial structure. Human fMRI experiments have shown broadly similar patterns of responses to face identity,<sup>18</sup> as well as other categories of high-level sensory stimuli such as human voices<sup>16</sup> and learned shapes.<sup>19</sup> However, more recent work investigating single-unit coding of facial identity in fMRI-defined face patches did not find evidence that the average face plays a special role in the tuning of face-selective neurons.<sup>13,14</sup> While those studies found similarly graded responses for extreme facial features, they did not find evidence that neurons responded most weakly for the average value of a face or its features. The basis of this discrepancy is an open and important question that likely stems from differences in the specific stimuli, presentation paradigm, recording locations, or analysis methods. For example, our earlier macaque study<sup>17</sup> only tested tuning along four identity trajectories, did not record from explicitly identified face patches, and left open a possibility that the observed tuning could have arisen from within-session adaptation. Critically, our previous study<sup>17</sup> did not systematically test identity trajectories that traversed the average face into anti-face territory, where a critical reversal in neural responses might be expected if the average face holds a special role.

In the present study, we investigate the responses of neurons in three macaque face patches to human and macaque face stimuli that were created along multiple intersecting identity trajectories. We report that average facial structure strongly shapes neural responses of face-selective neurons in each of the areas and is expressed as the dynamic suppression of sustained responses to low-identity faces. We discuss the implications of these findings for face recognition, as well as the potentially broader role of internalized, normative stimuli for high-level perceptual processing.

## RESULTS

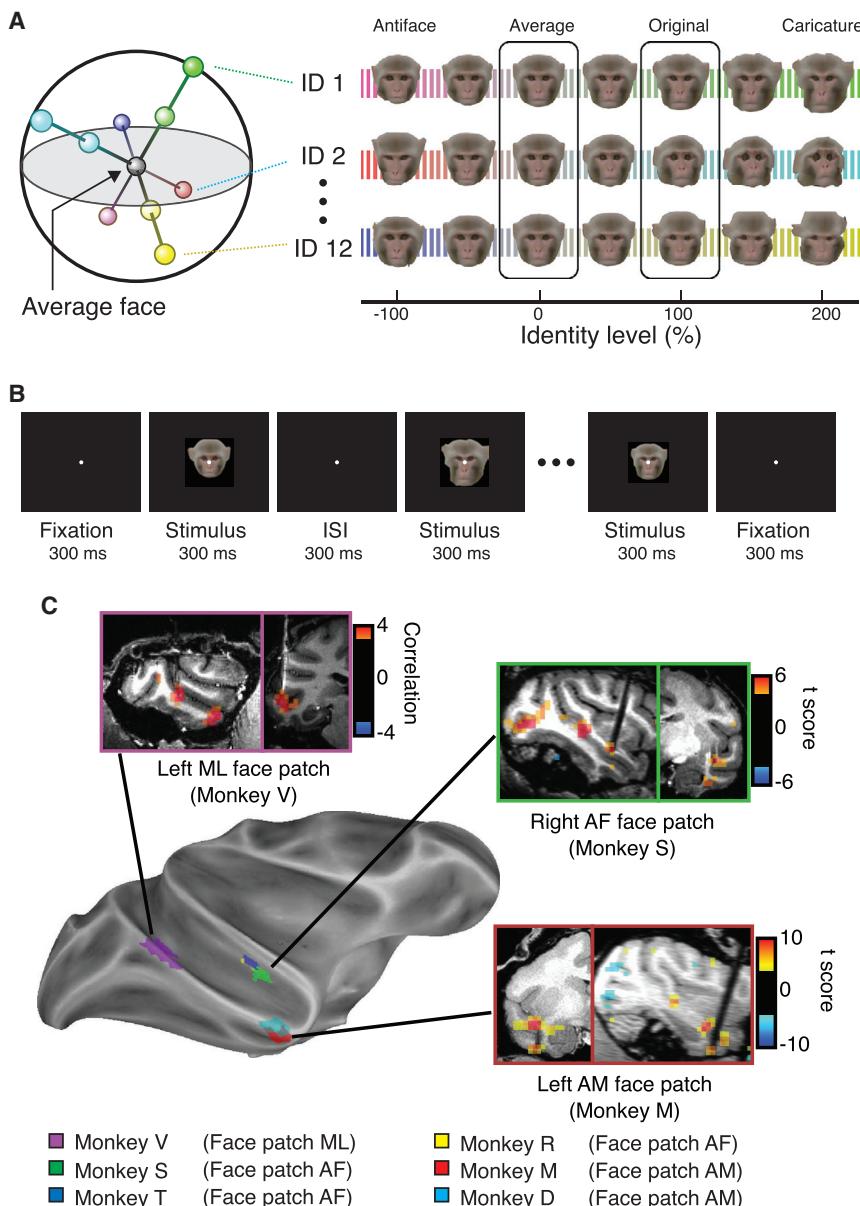
We assessed spiking responses to photorealistic face stimuli morphed along multiple continuous identity trajectories spanning the average face at the center of face space (Figure 1A).

For each trajectory, stimuli included diminished, full, and exaggerated identity faces, as well as anti-faces. In total, we recorded the response of each neuron to 145 monkey face stimuli derived from 12 original face identities and 289 human face stimuli derived from 24 original identities (Figures 1A, 1B, and S1). The species, identity, and position along the identity trajectory were randomly interleaved in passive presentation as monkeys maintained fixation. We recorded from a total of 279 neurons in three face patches (158 neurons from anterior medial, AM; 102 neurons from anterior fundus, AF; 19 neurons from middle lateral, ML) of six rhesus macaques (Figure 1C). We restricted analyses to 186 neurons (103 in AM, 67 in AF, and 16 in ML) that showed selectivity for monkey facial identity ( $p < 0.05$ , two-way ANOVA) during a window of 200–500 ms after stimulus onset.

In the following sections, we describe and quantify face identity tuning in these areas, focusing on the prominent and dynamic contribution of the average face. Throughout the paper, we feature the responses to the monkey face stimuli and refer to the corresponding human face responses in *Supplemental Information*.

### Prominent Tuning around the Average Face

Average facial structure strongly shaped neural responses in face patches AF, AM, and ML, with the faces nearest the average commonly eliciting the lowest spiking responses. The resulting V-shaped tuning function was often evident as a conspicuous dip in responses within individual identity trajectories (Figures 2A, 2B, S1B, and S1D) and less often took the form of an inverted V, or  $\Lambda$ , shape (Figure S2B). More broadly, neural tuning along an identity trajectory tended to show one of the following types of tuning around the average face: (1) a slope reversal at the average face (V or  $\Lambda$  shape), (2) an inflection point (knee shape), or (3) neither (linear ramp tuning). Different neurons exhibited varied combinations of identity tuning profiles (Figures S2A–S2C S3A, and S3B). Considering all trajectory/neuron combinations separately, 21.8% (486/2232) of all neuron/trajectory combinations exhibited a V shape, whereas 10.8% (241/2232) of all neuron/trajectory combinations exhibited a knee shape (Figures S3C–S3E). For the majority of neurons across the population, averaging across all 12 trajectories resulted in a V-shaped profile (Figures 2C and 2D). Likewise, averaging across the population resulted in a V-shaped profile for most of the individual identity trajectories (Figures S2D and S2E). Importantly, this tuning profile could not arise as the sum or average of linear ramp responses to different identities, which results in flat tuning, but instead depended upon the slope discontinuity at low-identity faces that could be extracted by averaging across identity trajectories (Figures S3F–S3K). Quantifying the slope changes on either side of the average face (Figures 2D–2G), we found that the slope reversal was most commonly expressed as a minimum response at or near the average face (see Figure S4 for corresponding quantification of responses to human faces). A control experiment demonstrated that morphing trajectories between two individuals, rather than along an identity trajectory, did not result in V-shaped tuning and exhibited no clear tuning along that dimension (Figures S5A–S5L). An additional control experiment revealed that contour-only stimuli derived from the morphed faces did not elicit a V-shaped tuning pattern (Figures S5M–S5R). Together, these



**Figure 1. Experimental Design**

(A) Schema of face space (left) and stimulus set (right). For both macaque and human faces, varying identity levels were created along multiple identity trajectories, created using standard morphing methods with reference to the average face (0% identity level) derived from the same stimulus set. The morphed faces varied in identity strength from average (0%) beyond the original face (100%) to create caricatures of exaggerated features (200%). By inverting the morphing transformation, we also created anti-faces on the same morphing trajectory (negative values), which have nominally opposite facial features.

(B) Sequence of passive-viewing task.

(C) Location of the fMRI-defined ML, AF, and AM face patches, from which neurons were recorded in the six monkeys participating in the present study.

See also Figure S1.

observations demonstrate the highly prominent role played by average facial structure in shaping the responses of face-selective neurons in face patches AF, AM, and ML.

#### Tuning Average Facial Structure Not Due to Short-Term Adaptation

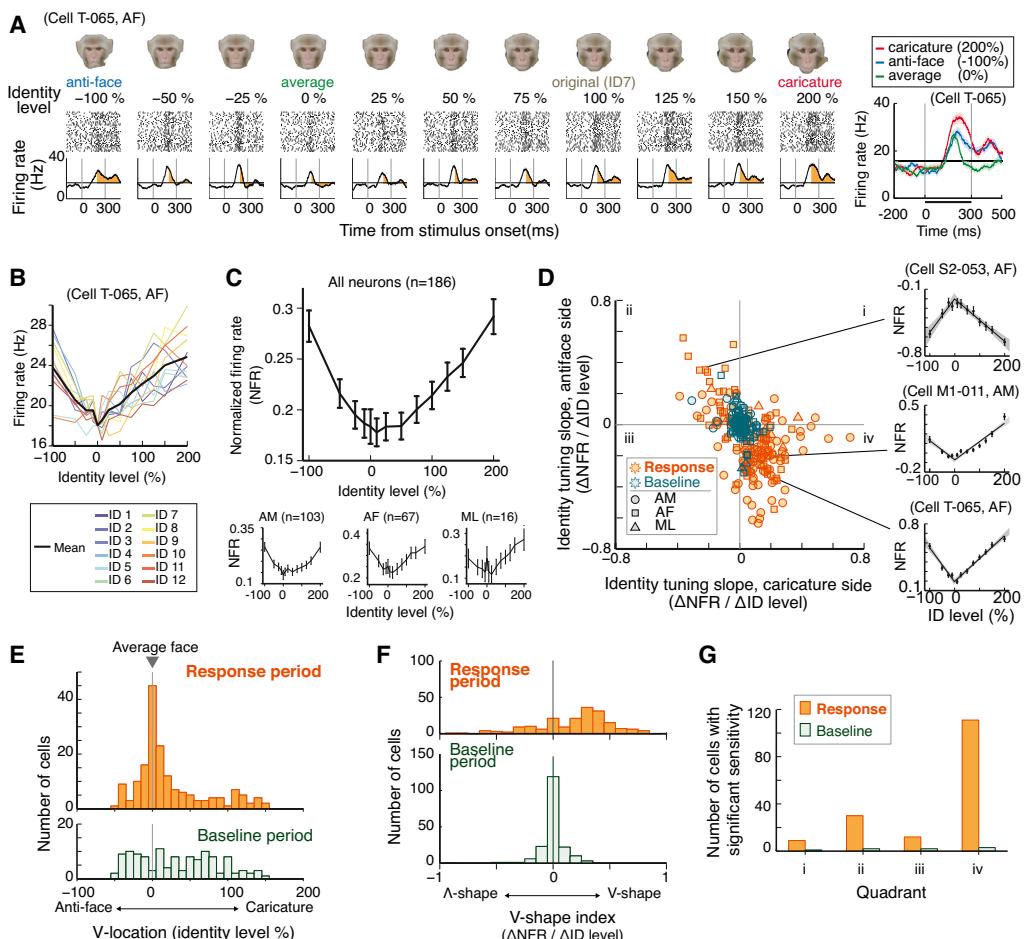
It has previously been suggested that short-term adaptation processes might underlie decreased responses to low-identity faces.<sup>15</sup> Here, we tested this possibility in four ways. First, we divided the main dataset into two segments to test whether the tuning for face identity level might evolve gradually within a session. The tuning computed from the early and late trials were nearly identical in each of the three face patches (Figure 3A). Second, we tested whether the observed tuning might be influenced by the immediately preceding stimulus.<sup>20</sup> Taking advantage of our intermixed stimulus sets, we found that trials preceded by human

faces and those preceded by monkey faces both exhibited V-shaped tuning (Figure 3B). Third, we utilized a unique dataset of stable longitudinal recordings to investigate the response to the first presentation of each stimulus during each session. This stimulus set involved daily, repeated recordings of stably isolated neurons in the AF face patch across 75 sessions spanning 3 months (Figure 3C). During this period, 13 neurons were stably monitored and showed an unchanging tuning for face identity. To examine the possibility of short-term adaptation, we performed analysis to just the first presentation of each stimulus on each day. This analysis revealed that the basic tuning profile was present from the very beginning of each session and was thus present from the first block of stimulus presentations (Figure 3D). Fourth, to address the remote possibility that adaptation to the

average face might develop with just a very small number of stimulus presentations, we analyzed responses to only the absolute first few trials each day (Figures 3E and 3F). This analysis revealed an expected nonspecific visual adaptation, likely stemming from the retina or retinotopic visual cortex, but no adaptation related to face identity. High-identity faces were, from the first absolute presentations during each session, larger than those to low-identity faces ( $1.35 \pm 0.16$  [high-identity faces] versus  $1.05 \pm 0.17$  [low-identity faces], paired *t* test,  $p < 0.001$ ). Taken together, these analyses rule out the possibility that the low responses to average faces arise through a form of within-session adaptation.<sup>14,15</sup>

#### Late Emergence of V-Shaped Tuning

We next examined the temporal evolution of the observed tuning profile, tracking the time course of V-shaped tuning using a



**Figure 2. V-Shaped Tuning by Face Patch Neurons**

(A) Response of a neuron to face stimuli along a single identity trajectory. This neuron exhibited its smallest sustained response to the low-identity faces in and around the average face. The orange shaded regions indicate responses above the baseline during 200–500 ms.

(B) Response of the neuron in (A) for all 12 macaque identity trajectories. This neuron exhibited a V-shaped tuning pattern along most of the individual identity trajectories during 200–500 ms after stimulus onset.

(C) Mean response of all face-selective neurons in the three face patches showing characteristic V-shaped tuning. Insets: mean population response shown separately for AM, AF, and ML patches.

(D) Quantification of tuning shape by two linear regression lines. During response period, the slopes on the caricature side tended to be anti-correlated with slopes on the anti-face side, indicating V-shaped or  $\Lambda$ -shaped tuning.

(E) V-location (identity level of crossing point of two regression lines, vertex of the V) distributions during response (top) and baseline (bottom) period. Many neurons had a V location around 0% identity level (average face) during the response period: 106 (57.0%) neurons showed a V location within  $\pm 25\%$  identity level, while only 35 (18.8%) did during the baseline period.

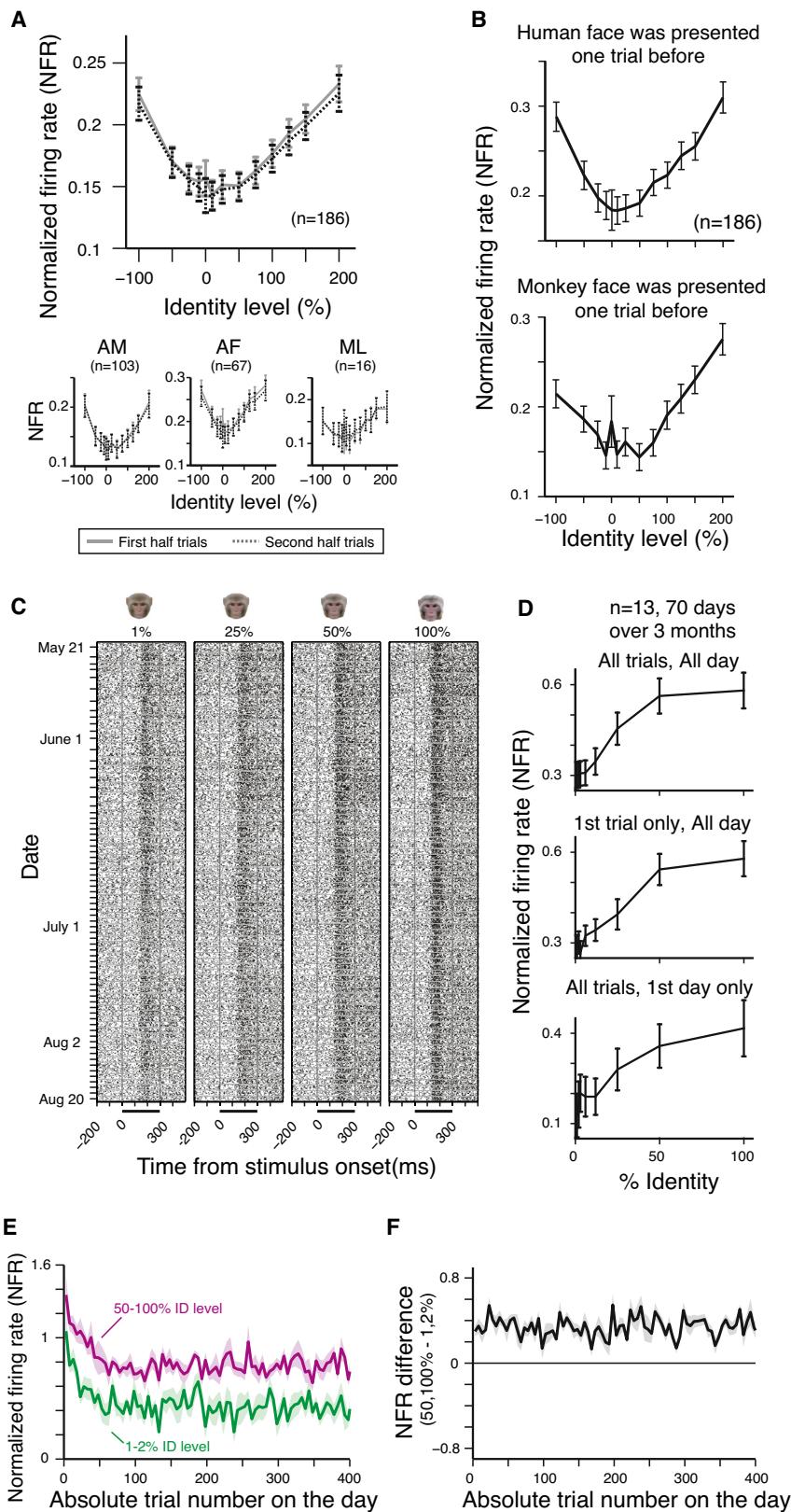
(F) V-shape index (difference of regression slope between two lines) distributions during response (top) and baseline (bottom) period. V-shape index was significantly biased toward positive values during the response period ( $p < 0.001$ , Wilcoxon's sign-rank test). During the response period, 147 (79.0%) neurons showed significantly different regression slopes, while only five (2.7%) did during the baseline period.

(G) Number of neurons with significant sensitivity (significant regression slopes in either side), counted in each quadrant of scatterplot in (D). During the response period, the majority (111, 59.7%) of neurons were in the fourth quadrant (V shape), and the second highest proportion (30, 16.1%) were the second quadrant ( $\Lambda$  shape).

Data are represented as mean  $\pm$  SEM. See also Figures S1–S5.

sliding-window analysis. We found that the V-shaped profile was not present immediately following the stimulus onset, but rather emerged during the later, sustained portion of the response. Figure 4A shows the time course of a typical face-selective neuron, with a large transient response evident around 100 ms after stimulus onset, followed by the later sustained response. For this and most other neurons, the V-shaped profile emerged after a delay, reaching a plateau at  $\sim 200$  ms after stimulus presentation. This

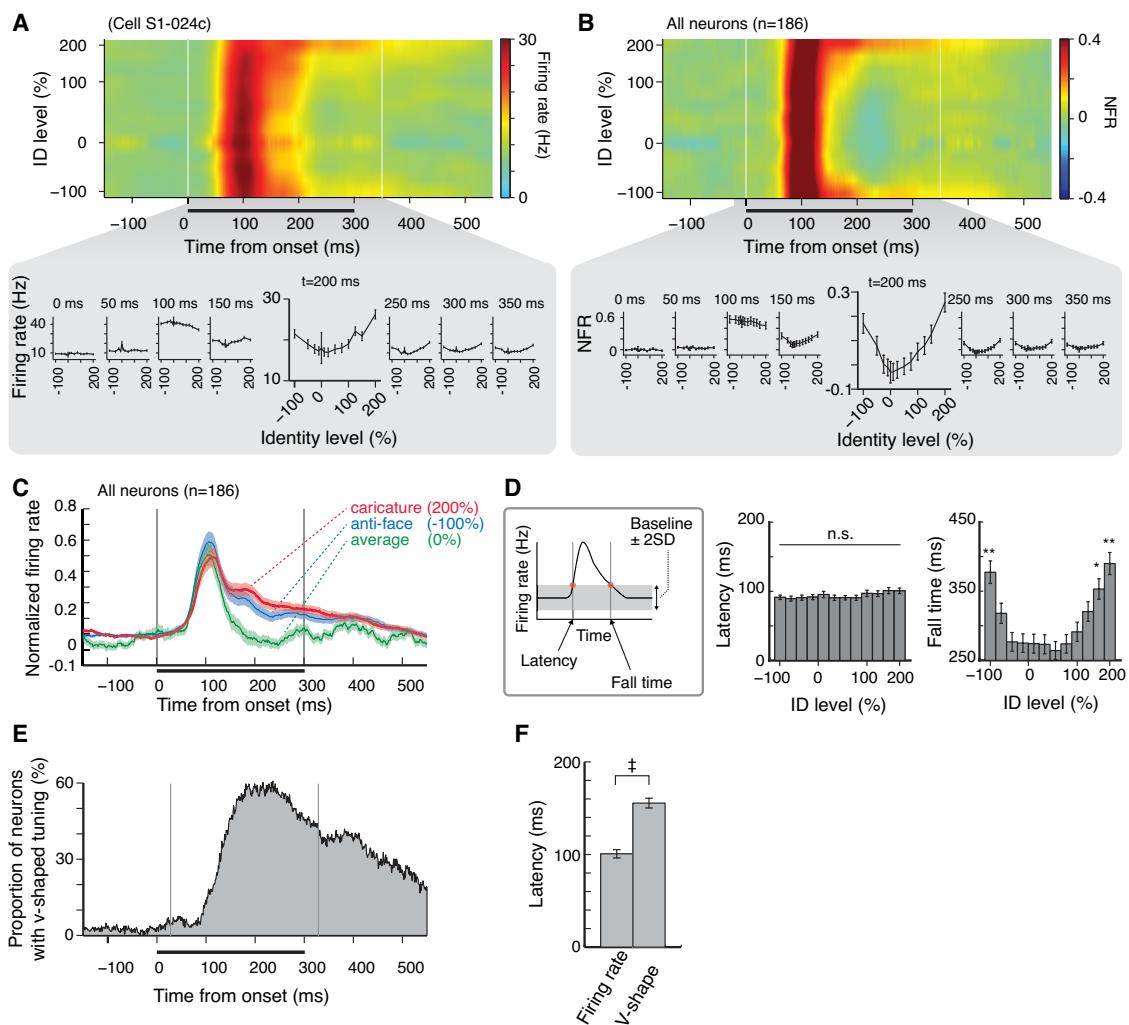
temporal tuning profile was a shared feature across neurons and was clearly evident in the population-averaged response (Figure 4B). Late tuning for facial identity is consistent with some previous studies reporting higher face identity tuning of neurons in inferotemporal cortex later in the neural response period.<sup>21,22</sup> It is also reminiscent of recent results in which the late responses of inferior temporal neurons are interpreted to be modulated by a hierarchical error signal.<sup>23</sup> Although there were no differences



**Figure 3. Stable Representation of the Average Face within and across Sessions**

- (A) The V-shaped tuning pattern was consistently observed throughout each recording session, as shown by comparing trials from the first half of each session with those from the second half of each session. Insets, response of each face patch.
- (B) V-shaped tuning did not depend upon the immediately preceding stimulus. The top panel shows the responses of all neurons only for trials in which the macaque face was preceded by a human face, whereas the bottom panel shows the responses of the same neurons for trials in which the macaque face was preceded by a macaque face.
- (C) Longitudinal recording from a neuron over 3 months, showing stable tuning pattern along identity levels.
- (D) Tuning pattern consistently observed through the longitudinal recording session, comparing the tuning of all responses (top) to those recorded from the first trial of each stimulus on each day (middle) to those recorded from all trials on the first day (bottom).
- (E) Responses to the first absolute trial numbers from each day, sorted according to high (50%–100%) versus near-average (1%–2%) identity levels. The first few absolute trials of each day exhibit a nonspecific adaptation, which equally affected the high and near-average identities.
- (F) There was a constant difference in the response to the high- and near-average identity levels analyzed in (E).

Data are represented as mean  $\pm$  SEM.

**Figure 4. Late Emergence of V-Shaped Tuning**

- (A) The time course of the V-shaped tuning pattern in an example neuron.
- (B) Average time course across the population of 186 face-selective neurons. The V-shaped tuning pattern was not evident during the large transient response around 100 ms but slowly emerged during the later sustained period that started around 150 ms and peaked at approximately 200 ms.
- (C) The mean response time courses of the average-face, caricatures, and anti-face responses, which diverge strongly during the sustained response period.
- (D) The response latency and fall time, presented as a function of face identity level. The response dropped earlier for faces of lower identity levels. \* $p < 0.05$ ; \*\* $p < 0.01$  (Tukey's post hoc test against faces between -25% to 25% identity level after one-way ANOVA).
- (E) The number of cells exhibiting significant V-shaped tuning as a function of time.
- (F) The latency of V-shape tuning and initial response, each computed for individual cells. The V-shaped tuning emerged much later than initial firing rate increases. ‡,  $p < 0.01$  (paired t test with Bonferroni's correction).

Data are represented as mean  $\pm$  SEM.

in the initial response latency across identity level, the response magnitude to low-identity faces was abruptly attenuated during the visual response, while that to the high positive and negative identity faces persisted (Figures 4C and 4D). Across the population, the proportion of neurons exhibiting V-shaped tuning peaked at around 200 ms ( $55.8\% \pm 3.5\%$  between  $200 \pm 50$  ms; Figure 4E), on average, emerging  $54.8 \pm 6.9$  ms (mean  $\pm$  SE) after the onset of the visual response (Figure 4F). These results indicate that V-shaped tuning is largely absent during the early transient response to faces but emerges during the transition to the later sustained period, driven by a drop in the responses to low identities.

### Superposition of V-Shaped Tuning and Linear Tuning

Previous studies reported linear or ramp tuning for facial features in face patches.<sup>13,14</sup> Our analysis demonstrated that late V-shaped tuning was superimposed upon an initial linear ramp tuning (Figure 5). To estimate the relative contributions of these two different processes over time, we first performed a regression analysis with a model that combined linear tuning and V-shaped tuning (Figures 5A and S6). Across the population, the late responses of face-selective neurons reflected a mixed contribution of both components (Figures S6C–S6G). Figure 5B is a graphic depiction of the linear tuning and V-shaped tuning components in early versus late response time windows. The center of the plots

corresponds to the average face, and identity levels become larger at the periphery. The plots are averaged across the population, computed after a rank ordering, for each cell, of the identities eliciting the highest to lowest responses. The tangential gradation in firing rate shown in the linear tuning component reflects neurons' typical preference for a subset of face identities, and this component did not change significantly over the course of the presentation (Figure 5B, left; tangential tuning in Figure 5C,  $p = 0.85$ , paired t test). In contrast, the concentric V-shaped tuning component was absent in the beginning but emerged during the sustained response period (Figure 5B, middle; see also Figures S6F and S6G), after which time the suppressed response to average facial structure strongly shaped neural tuning (radial tuning in Figure 5C,  $p < 0.001$ , paired t test; see also Figure 5D).

### Average Suppression Marked by Synchronous Spiking

Given the simultaneity of our recordings from local populations within each face patch, we asked whether a detailed look at spiking might offer clues about the nature of the observed tuning. We found that delayed suppression of low-identity faces coincided in time with heightened synchrony of spike timing events. All pairs of neurons were analyzed for their temporal coherence of their spike times, which is often taken as an index of circuit operation related to the efficient transmission of information.<sup>24–26</sup> Focusing on the late visual response, we found that the spike-spike coherence between pairs of neurons was notably stronger for low-identity faces than for more extreme faces (Figures 6A, 6B, and S7A–S7G), suggesting a higher level of local circuit coordination during the processing of low-identity faces. Coherent spikes most often showed coherence with a zero phase, indicating simultaneity in their firing (Figures 6C, S7D, and S7F). Spike-LFP (local field potential) coherence was similarly heightened during the presentation of low-identity faces (Figure S7B and S7D–S7G), with spikes most commonly issued in the troughs of the low- $\gamma$  LFP, when oscillatory inhibition should be weakest<sup>27</sup> (Figure S7F). Both spike-spike and spike-LFP coherence was graded as a function of identity level, falling off for higher identity levels in a pattern opposite to that of firing rate responses (Figures 6D and S7H). Like the firing rate suppression, this elevated coherence emerged during the sustained response, matching the temporal dynamics of the V-shaped firing rate profile (Figures 6E and 6F; see Figures S7M–S7Q for corresponding human face data; see also Figures S7R and S7S for time course of AM and AF face patches). Importantly, these results could not have arisen from a trivial coupling of firing rate and coherence, as we matched the spiking rates in the coherence calculation (see *Method Details* for details; see also Figures S7I–S7L). The inverted relationship between spiking synchrony and identity tuning suggests a mechanism of broad, synchronous inhibition for the average face during the late response period (Figure 7). Increases in spike timing synchrony among excitatory neurons have previously been linked to circuit inhibition.<sup>27–31</sup>

### DISCUSSION

In this study, we demonstrate that tuning around the average face is a shared feature of face-selective neurons in three face patches. This tuning took the form of a delayed suppression in

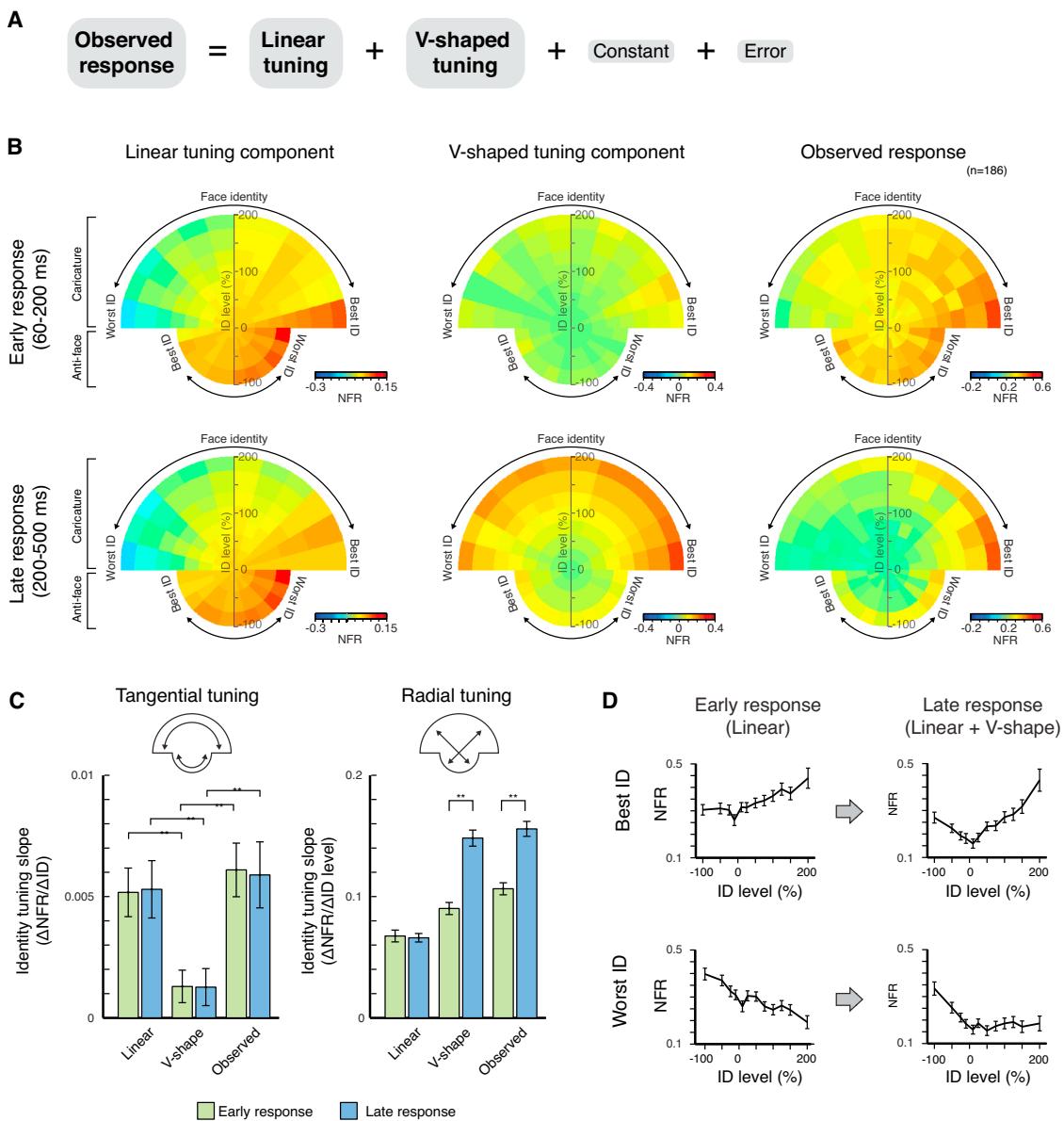
the spiking responses to low-identity faces, with the resulting V-shaped pattern superimposed upon a linear ramp tuning that was apparent from the earliest elevation in firing rate. In the sections below, we discuss the results in relation to previous studies, the implications for face recognition, and the possibility that similar normative principles might apply more broadly to the learning of complex structure.

### Relation to Previous Studies

One shared finding between this and previous electrophysiological studies is the tendency of neurons in face patches to respond most strongly to extreme facial structure, whether presented in the form of morphed faces or parameterized internal features.<sup>13,14</sup> That consistent observation is in broad agreement with norm-based models of face encoding and suggests that face-selective neurons are particularly sensitive to facial features that might be informative for individual recognition. However, an important difference with the previous studies has centered on the specific role of the average face in shaping these graded identity responses, with the previous studies finding no or minimal evidence that average facial structure played a critical role.

One potentially important methodological detail, particularly in light of the delayed emergence of V-shaped tuning, is the longer duration of face presentation times in the current study (300 ms) compared to previous studies (117–150 ms).<sup>13,14</sup> The importance of this factor is supported by our comparison of tuning during the early and late response periods. In agreement with the previous studies, we found a roughly linear ramp tuning during the early transient response, which was then met with a roughly equally weighted V-shaped tuning during the later sustained response (Figure 5). These two profiles might reflect different aspects of face processing, such as the difference between coarse versus detailed analysis of facial structure. This progression is consistent with previous reports suggesting that neural information corresponding to facial identity was most pronounced during the later sustained response period.<sup>21,22,32</sup> As none of these studies measured the monkey's performance recognition, but instead focused on the principles of high-dimensional tuning, the specific relationship to the animal's identity perception to the graded responses and late suppression of the average is only suggested but not proven. When tested on their individual recognition of morphed faces, macaques did show similar performance to humans,<sup>17,33</sup> though the degree of overlap in macaque and human face perception is a point of controversy.<sup>34</sup>

The stimulus generation and analysis methods may also contribute to the observed differences with previous studies. It is unlikely that the present use of macaque faces, in contrast to previous studies' use of human faces, is the basis of the difference between the present and previous studies, since in this study we also presented human faces and found qualitatively similar effects for both macaque and human faces, and since our previous findings used human faces.<sup>17</sup> While other stimulus-related details may be factors in the observed differences, such as the choices for parametrizing cartoon facial features<sup>14</sup> or the principles of linear feature combination in the context of a principal components space,<sup>13</sup> these factors seem less likely to explain the observed differences than the factors related to the temporal dynamics of V-shaped tuning. On the analysis side, we examined potential differences, such as our averaging across

**Figure 5. Identity Tuning as a Superposition of V-Shaped and Linear Tuning**

(A) Decomposition of the observed response. See also Figure S6 for the detail of the regression model.

(B) Heatmap showing the responses of linear tuning component (left), V-shaped tuning component (middle), and observed response (right). The identity trajectories were first rank-ordered according to the observed response and then averaged. See Figure S6 for the procedure of the decomposition.

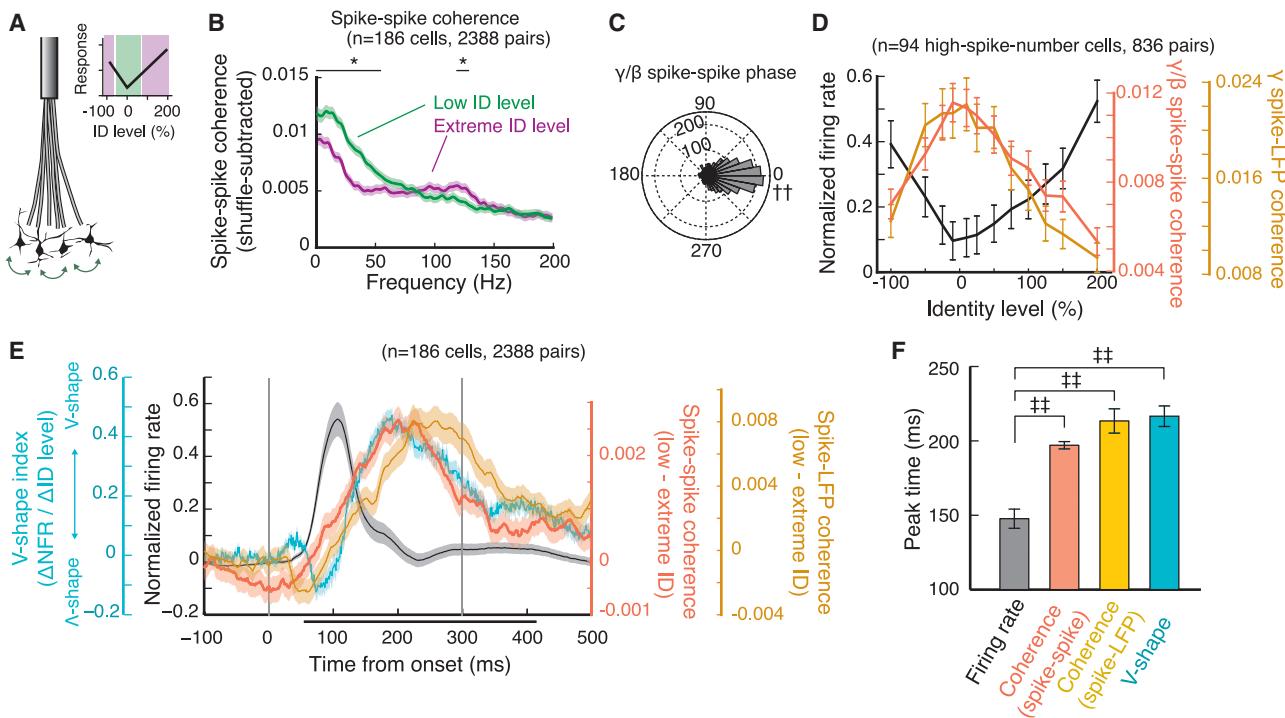
(C) Tuning for face identity (tangential tuning, left) and identity level (radial tuning, right). The linear tuning component showed high tangential tuning during both early and late response, while the V-shaped tuning component showed higher radial tuning later in the response (\*\* $p < 0.01$ , paired t test with Bonferroni's correction).

(D) Radial identity-level tuning for the best (top) and worst (bottom) identity trajectories, according to the initial rank ordering. Both the highest- and lowest-responding face identities (Best ID and Worst ID, respectively) elicited mostly linear tuning during the initial response but were then impacted by the later suppression of low-identity responses, resulting in a discontinuity (V shape or knee shape) at the average face. Note that in the case of the worst ID, the anti-faces, with nominally opposite features, showed the highest response to the initial tuning.

Data are represented as mean  $\pm$  SEM. See also Figure S6.

multiple identity trajectories and our presentation of many low-identity faces in the same session, in great detail to demonstrate that the observed role of the average face in neural tuning did not arise for trivial or irrelevant reasons (Figures 2, 3, and S3). Averaging together of V-shaped trajectories, knee-shaped trajectories, and linear ramp-shaped trajectories typically

revealed a minimum response at the average face, even though some of the trajectories appeared with no discontinuity there (Figure S3). Moreover, the observed tuning around the average face was present from the beginning of each recording session, ruling out its emergence due to short-term adaptation effects (Figure 3).



**Figure 6. Maximal Neuronal Synchrony for the Average Face**

- (A) Cartoon illustrating simultaneous recording of face patch neurons with implanted microwire bundles. Inset, classification of low (green) and extreme (magenta) identity levels.
- (B) Frequency spectrum of spike-spike coherence between simultaneously recorded neuron pairs. The coherence was higher in low-identity level than extreme identity level at <50 Hz. \* $p < 0.05$  (paired t test with Bonferroni's correction).
- (C) Phase relationships between neuronal pairs, showing coherent firing of spikes. ††,  $p < 0.01$  (Rayleigh test).
- (D) Highest coherence for the average face, showing opposite tuning pattern to firing rate.
- (E) Coherence time course following time course of V-shaped tuning rather than that of the firing rate changes.
- (F) Time to peak of coherence is similar to that of the V-shaped tuning and much delayed relative to the firing rate. ‡‡,  $p < 0.01$  (paired t test with Bonferroni's correction).

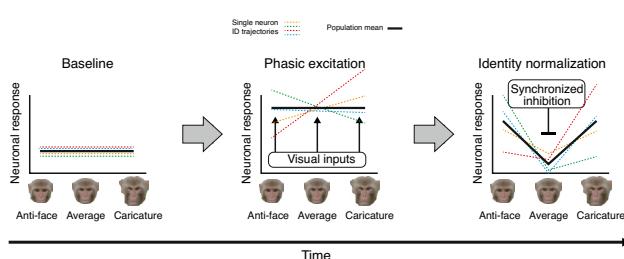
Data are represented as mean  $\pm$  SEM. See also Figure S7.

Future studies may employ other approaches to investigate how the average face dynamically shapes the cortical processing of face identity. For example, human electrophysiological studies have recently used an image reconstruction approach to characterize the neural dynamics of face processing,<sup>35–37</sup> which has been claimed to reconstruct face and to model memory for facial identity.<sup>38</sup> It is likely that this direction will soon become more powerful with the advancement of deep convolutional network,<sup>39,40</sup> which will allow for a more principled sampling of stimulus space and may reveal new surprises about how high-level visual neurons support aspects of face perception, including identity recognition. While the present study provides insight into the encoding principles of single cells, the passive viewing paradigm cannot establish the link with recognition performance. In fact, as most studies of face patches to date have used passive presentations, there is currently not much information about whether the activity of such tuned neurons plays a direct role in trial-by-trial recognition performance. Recent work has, however, demonstrated that local, time-resolved perturbation of face patch activity in macaques can produce psychophysical disruption of face perception during identity matching<sup>41</sup> and gender discrimination.<sup>42</sup> Additional work is needed to better understand the role of

face patch neurons in active and mnemonic aspects of face perception.

#### The Brain's Acquisition and Application of Norms

The average face has long been seen to have a computational role as a norm, against which the structural details of a face might be internally compared<sup>5,7,9</sup>. An interesting recent perspective on face processing places this operation in the context of predictive coding and emphasizes the temporal dynamics of potentially normative processes.<sup>23</sup> That work, which focused on typicality rather than individual identity, may provide clues about the nature of the delayed suppression to the average face. It is interesting to construe the brain's internal analysis of facial identity as a form of predictive coding, where a stable representation of face averageness serves as the prediction. Mechanistically, such a process might correspond to an identity normalization of sorts, with discounting of average facial structural components across the population for each viewed face in order to improve the internal signal for distinctive facial features. Like other examples of normalization in the nervous system, this operation would adjust a population's dynamic range to be most sensitive to dimensions that are important for behavior.<sup>43</sup>



**Figure 7. Schematic of Proposed Mechanism for Norm-Based Tuning**

An initial phasic response, in which linear identity-specific tuning is unaffected by the average face, is met by synchronous inhibition that suppresses the population responses to average facial structure. The active suppression of average facial structure could serve to hone neural responses to distinctive critical facial characteristics while suppressing redundant or irrelevant responses across the population.

How the brain initially acquires and maintains the internal representations of average faces is likely to become a critical question moving forward. The extended childhood of primates, and particularly that of humans, affords a vast sampling of facial variation, whose gradual influence on the brain may gradually embed a set of perceptual expectations, expressed through specialized neural tuning functions. In face recognition, the most conspicuous component of these expectations seems to be tuning around average facial structure. The principles of plasticity by which such tuning might arise, wherein the developing brain effectively tailors itself to optimally analyze its encountered statistical variation of important sensory stimuli, is a fascinating and poorly understood topic.

For face processing, some clues about development come from recent work tracking the progression of basic face selectivity in the visual cortex of infant macaques,<sup>44</sup> including its dependence on visual experience,<sup>45</sup> and its relationship to other forms of developmental plasticity.<sup>46</sup> In general, the basic organization of face-selective areas and corresponding responses of individual neurons emerge early in the infant brain.<sup>44,47</sup> However, the more elaborate tailoring of the brain's responses to meet the challenges of individual identity is likely to proceed much more gradually, certainly based on the time course of identity recognition performance in human children.<sup>48,49</sup> How this process impacts the responses of neurons in face patches is an open question. One possibility is that there is a cascade of local changes that follow a face processing hierarchy, which might be suggested by electrophysiological studies showing a more robust identity tuning in face patch AM than in other face patches tested.<sup>21</sup> In our study, however, we did not find obvious differences in the V-shaped tuning profiles of the face patches we tested. Thus, a second possibility that may be most consistent with the present results is that face patches experience a parallel, local modification of their visual selectivity that results in an embedding of average facial structure. Finally, a third possibility is the critical learning about facial structure is not applied to the visual system proper but perhaps instead takes place in other brain regions. The perirhinal cortex, for example, shares dense reciprocal connections with the inferior temporal cortex<sup>50</sup> and has recently been shown to respond selectively to familiar faces.<sup>51</sup> It is conceivable that a normalizing signal, expressed as the observed pattern of delayed suppression, represents

the input from the perirhinal cortex or other area external to the face patches.

Finally, the brain's approach to face recognition may be representative of its approach to other types of visual and nonvisual problems. As animals actively sample important details about stimuli and events in a continually changing natural environment, the brain is tasked with distilling and retaining useful information to aid in future life. The developing brain is particularly able to gather and assimilate new sensory, spatial, and social experiences. For the problem of face recognition, one product of this experience appears to be the embedding of average facial structure, almost certainly together with natural dimensions of facial variation that together prompt neurons to respond in a graded fashion to increasingly extreme features.<sup>13,14,17</sup> Similar principles appear to hold in the case of voice recognition<sup>16</sup> and the coding of recently learned geometric shapes.<sup>19</sup> It is well possible that the gathering, embedding, and real-time application of norms is a fundamental aspect of how we perceive the world, for example, shaping the brain's visual responses to bodily postures and movements,<sup>52</sup> object coloration,<sup>53</sup> volumetric structure,<sup>46,54</sup> and spatial layout.<sup>55</sup> Such normative principles and their implementation in the brain may also extend beyond sensory processing into more abstract domains, serving as the basis for understanding categories, social relationships, object value, and many other cognitive variables.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- [METHOD DETAILS](#)
  - Subjects
  - Behavioral task and visual stimuli
  - Electrophysiology
  - Long duration recordings to investigate role of visual adaptation
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cub.2020.09.070>.

## ACKNOWLEDGMENTS

We would like to thank David Yu, Charles Zhu, and Frank Ye for assistance with fMRI scanning; Julie Hong, Katy Smith, George Dold, and David Ide for technical assistance and development; Olga Dal Monte and Carlos E. Thomaz for providing face stimuli; and Vasileva Liubov, Igor Bondar, David Godlove, Bruno Averbeck, Teppei Matsui, and Toshiyuki Hirabayashi for helpful discussions. Functional and anatomical MRI scanning was carried out in the Neurophysiology Imaging Facility Core (NIMH, NINDS, NEI). This work utilized the computational resources of the NIH HPC Biowulf cluster, <http://hpc.nih.gov>. This work was supported by the Intramural Research Program of the National Institute of Mental Health (ZIAHM002838, ZIAHM002898, and ZIAHM002899).

to D.A.L. K.W.K. was supported by the fellowship from the Uehara Memorial Foundation.

### AUTHOR CONTRIBUTIONS

A.P.J. and D.A.L. designed the experiments. A.P.J. prepared stimulus set. K.W.K., D.B.T.M., E.N.W., B.E.R. and D.A.L. performed animal surgeries. K.W.K., D.B.T.M., and B.E.R. ran fMRI experiments and analysis. K.W.K., A.P.J., and E.N.W. collected electrophysiology data. K.W.K., A.P.J., and D.A.L. performed analysis of neuronal data. K.W.K. and D.A.L. wrote the paper. All the authors contributed to editing of the manuscript. D.A.L. supervised the study.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 15, 2020

Revised: August 14, 2020

Accepted: September 22, 2020

Published: October 15, 2020

### REFERENCES

- Leopold, D.A., and Rhodes, G. (2010). A comparative view of face perception. *J. Comp. Psychol.* 124, 233–251.
- Hung, C.-C., Yen, C.C., Ciuchta, J.L., Papoti, D., Bock, N.A., Leopold, D.A., and Silva, A.C. (2015). Functional mapping of face-selective regions in the extrastriate visual cortex of the marmoset. *J. Neurosci.* 35, 1160–1172.
- Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.
- Tsao, D.Y., Freiwald, W.A., Knutson, T.A., Mandeville, J.B., and Tootell, R.B. (2003). Faces and objects in macaque cerebral cortex. *Nat. Neurosci.* 6, 989–995.
- Benson, P.J., and Perrett, D.I. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *Eur. J. Cogn. Psychol.* 3, 105–135.
- Leopold, D.A., O'Toole, A.J., Vetter, T., and Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat. Neurosci.* 4, 89–94.
- Rhodes, G., and Leopold, D.A. (2011). Adaptive Norm-Based Coding of Face Identity. In *The Oxford handbook of face perception*, G. Rhodes, A. Calder, M. Johnson, and J.V. Haxby, eds. (Oxford: Oxford University Press), pp. 263–286.
- Webster, M.A., Kaping, D., Mizokami, Y., and Duhamel, P. (2004). Adaptation to natural facial categories. *Nature* 428, 557–561.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol.* A 43, 161–204.
- Rhodes, G., Carey, S., Byatt, G., and Proffitt, F. (1998). Coding spatial variations in faces and simple shapes: a test of two models. *Vision Res.* 38, 2307–2321.
- Lee, K., Byatt, G., and Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: testing the face-space framework. *Psychol. Sci.* 11, 379–385.
- Blanz, V., O'Toole, A.J., Vetter, T., and Wild, H.A. (2000). On the other side of the mean: the perception of dissimilarity in human faces. *Perception* 29, 885–891.
- Chang, L., and Tsao, D.Y. (2017). The code for facial identity in the primate brain. *Cell* 169, 1013–1028.e14.
- Freiwald, W.A., Tsao, D.Y., and Livingstone, M.S. (2009). A face feature space in the macaque temporal lobe. *Nat. Neurosci.* 12, 1187–1196.
- Kahn, D.A., and Aguirre, G.K. (2012). Confounding of norm-based and adaptation effects in brain responses. *Neuroimage* 60, 2294–2299.
- Latinus, M., McAleer, P., Bestelmeyer, P.E., and Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex. *Curr. Biol.* 23, 1075–1080.
- Leopold, D.A., Bondar, I.V., and Giese, M.A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* 442, 572–575.
- Loffler, G., Yourganov, G., Wilkinson, F., and Wilson, H.R. (2005). fMRI evidence for the neural representation of faces. *Nat. Neurosci.* 8, 1386–1390.
- Panis, S., Wagemans, J., and Op de Beeck, H.P. (2011). Dynamic norm-based encoding for unfamiliar shapes in human visual cortex. *J. Cogn. Neurosci.* 23, 1829–1843.
- Vinken, K., Op de Beeck, H.P., and Vogels, R. (2018). Face repetition probability does not affect repetition suppression in macaque inferotemporal cortex. *J. Neurosci.* 38, 7492–7504.
- Freiwald, W.A., and Tsao, D.Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330, 845–851.
- Sugase, Y., Yamane, S., Ueno, S., and Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400, 869–873.
- Issa, E.B., Cadieu, C.F., and DiCarlo, J.J. (2018). Neural dynamics at successive stages of the ventral visual stream are consistent with hierarchical error signals. *eLife* 7, e42870.
- Dean, H.L., Hagan, M.A., and Pesaran, B. (2012). Only coherent spiking in posterior parietal cortex coordinates looking and reaching. *Neuron* 73, 829–841.
- Gregoriou, G.G., Gotts, S.J., Zhou, H., and Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* 324, 1207–1210.
- Koyano, K.W., Takeda, M., Matsui, T., Hirabayashi, T., Ohashi, Y., and Miyashita, Y. (2016). Laminar module cascade from layer 5 to 6 implementing cue-to-target conversion for object memory retrieval in the primate temporal cortex. *Neuron* 92, 518–529.
- Sohal, V.S. (2016). How close are we to understanding what (if anything)  $\gamma$  oscillations do in cortical circuits? *J. Neurosci.* 36, 10489–10495.
- Hasenstaub, A., Otte, S., and Callaway, E. (2016). Cell type-specific control of spike timing by gamma-band oscillatory inhibition. *Cereb. Cortex* 26, 797–806.
- Hasenstaub, A., Shu, Y., Haider, B., Kraushaar, U., Duque, A., and McCormick, D.A. (2005). Inhibitory postsynaptic potentials carry synchronized frequency information in active cortical networks. *Neuron* 47, 423–435.
- Sohal, V.S., Zhang, F., Yizhar, O., and Deisseroth, K. (2009). Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature* 459, 698–702.
- Whittington, M.A., Traub, R.D., Kopell, N., Ermentrout, B., and Buhl, E.H. (2000). Inhibition-based rhythms: experimental and mathematical observations on network dynamics. *Int. J. Psychophysiol.* 38, 315–336.
- Ghuman, A.S., Brunet, N.M., Li, Y., Konecky, R.O., Pyles, J.A., Walls, S.A., Destefino, V., Wang, W., and Richardson, R.M. (2014). Dynamic encoding of face information in the human fusiform gyrus. *Nat. Commun.* 5, 5672.
- Leopold, D.A., and Bondar, I. (2005). Adaptation to complex visual patterns in humans and monkeys. In *Fitting the mind to the world: Adaptation and after-effects in high-level vision*, C.W.G.C.G. Rhodes, ed. (Oxford: Oxford University Press), pp. 189–211.
- Rossoni, B., and Taubert, J. (2019). What can we learn about human individual face recognition from experimental studies in monkeys? *Vision Res.* 157, 142–158.
- Nemrosov, D., Niemeier, M., Patel, A., and Nestor, A. (2018). The neural dynamics of facial identity processing: insights from EEG-based pattern analysis and image reconstruction. *eNeuro* 5, <https://doi.org/10.1523/ESEURO.0358-17.2018>.

36. Nemroff, D., Behrmann, M., Niemeier, M., Drobotenko, N., and Nestor, A. (2019). Multimodal evidence on shape and surface information in individual face processing. *Neuroimage* 184, 813–825.
37. Roberts, T., Cant, J.S., and Nestor, A. (2019). Elucidating the Neural Representation and the Processing Dynamics of Face Ensembles. *J. Neurosci.* 39, 7737–7747.
38. Zhan, J., Garrod, O.G.B., van Rijsbergen, N., and Schyns, P.G. (2019). Modelling face memory reveals task-generalizable representations. *Nat. Hum. Behav.* 3, 817–826.
39. Ponce, C.R., Xiao, W., Schade, P.F., Hartmann, T.S., Kreiman, G., and Livingstone, M.S. (2019). Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. *Cell* 177, 999–1009.e10.
40. O'Toole, A.J., Castillo, C.D., Parde, C.J., Hill, M.Q., and Chellappa, R. (2018). Face space representations in deep convolutional neural networks. *Trends Cogn. Sci.* 22, 794–809.
41. Afraz, A., Boyden, E.S., and DiCarlo, J.J. (2015). Optogenetic and pharmacological suppression of spatial clusters of face neurons reveal their causal role in face gender discrimination. *Proc. Natl. Acad. Sci. USA* 112, 6730–6735.
42. Moeller, S., Crapse, T., Chang, L., and Tsao, D.Y. (2017). The effect of face patch microstimulation on perception of faces and objects. *Nat. Neurosci.* 20, 743–752.
43. Carandini, M., and Heeger, D.J. (2011). Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62.
44. Livingstone, M.S., Vincent, J.L., Arcaro, M.J., Srihasam, K., Schade, P.F., and Savage, T. (2017). Development of the macaque face-patch system. *Nat. Commun.* 8, 14897.
45. Arcaro, M.J., Schade, P.F., Vincent, J.L., Ponce, C.R., and Livingstone, M.S. (2017). Seeing faces is necessary for face-domain formation. *Nat. Neurosci.* 20, 1404–1412.
46. Srihasam, K., Vincent, J.L., and Livingstone, M.S. (2014). Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nat. Neurosci.* 17, 1776–1783.
47. Rodman, H.R. (1994). Development of inferior temporal cortex in the monkey. *Cereb. Cortex* 4, 484–498.
48. Gemine, L.T., Duchaine, B., and Nakayama, K. (2011). Where cognitive development and aging meet: face learning ability peaks after age 30. *Cognition* 118, 201–210.
49. Weigelt, S., Koldewyn, K., Dilks, D.D., Balas, B., McKone, E., and Kanwisher, N. (2014). Domain-specific development of face memory but not face perception. *Dev. Sci.* 17, 47–58.
50. Lavenex, P., Suzuki, W.A., and Amaral, D.G. (2002). Perirhinal and parahippocampal cortices of the macaque monkey: projections to the neocortex. *J. Comp. Neurol.* 447, 394–420.
51. Landi, S.M., and Freiwald, W.A. (2017). Two areas for familiar face recognition in the primate brain. *Science* 357, 591–595.
52. Nelissen, K., Luppino, G., Vanduffel, W., Rizzolatti, G., and Orban, G.A. (2005). Observing others: multiple action representation in the frontal lobe. *Science* 310, 332–336.
53. Lafer-Sousa, R., and Conway, B.R. (2013). Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nat. Neurosci.* 16, 1870–1878.
54. Hung, C.-C., Carlson, E.T., and Connor, C.E. (2012). Medial axis shape coding in macaque inferotemporal cortex. *Neuron* 74, 1099–1113.
55. Vaziri, S., and Connor, C.E. (2016). Representation of gravity-aligned scene structure in ventral pathway visual cortex. *Curr. Biol.* 26, 766–774.
56. Quiroga, R.Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput.* 16, 1661–1687.
57. Mitra, P.B., and Hemant. (2007). Observed brain dynamics (Oxford University Press).
58. Mitra, P.P., and Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophys. J.* 76, 691–708.
59. Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. *J. Stat. Softw.* 31, 1–21.
60. Russ, B.E., and Leopold, D.A. (2015). Functional MRI mapping of dynamic visual features during natural viewing in the macaque. *Neuroimage* 109, 84–94.
61. McMahon, D.B., Jones, A.P., Bondar, I.V., and Leopold, D.A. (2014). Face-selective neurons maintain consistent visual responses across months. *Proc. Natl. Acad. Sci. USA* 111, 8251–8256.
62. McMahon, D.B., Bondar, I.V., Afuwape, O.A., Ide, D.C., and Leopold, D.A. (2014). One month in the life of a neuron: longitudinal single-unit electrophysiology in the monkey visual system. *J. Neurophysiol.* 112, 1748–1762.
63. Asaad, W.F., Santhanam, N., McClellan, S., and Freedman, D.J. (2013). High-performance execution of psychophysical tasks with complex visual stimuli in MATLAB. *J. Neurophysiol.* 109, 249–260.
64. Hwang, J., Mitz, A.R., and Murray, E.A. (2019). NIMH MonkeyLogic: Behavioral control and data acquisition in MATLAB. *J. Neurosci. Methods* 323, 13–21.
65. Kleiner, M., Brainard, D., Pelli, D., Ingling, A., and Murray, R. (2007). What's new in Psychtoolbox-3? *Perception* 36, 14.
66. Langlois, J.H., and Roggman, L.A. (1990). Attractive faces are only average. *Psychol. Sci.* 1, 115–121.
67. Bondar, I.V., Leopold, D.A., Richmond, B.J., Victor, J.D., and Logothetis, N.K. (2009). Long-term stability of visual pattern selective responses of monkey temporal lobe neurons. *PLoS ONE* 4, e8222.
68. Mitchell, J.F., Sundberg, K.A., and Reynolds, J.H. (2009). Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* 63, 879–888.

**STAR★METHODS****KEY RESOURCES TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
Rhesus macaque ( <i>Macaca mulatta</i> )	NIMH/NIH	N/A
Software and Algorithms		
MATLAB 2014a-2019b	MathWorks	<a href="https://www.mathworks.com/">https://www.mathworks.com/</a>
Wave_clus	<sup>56</sup>	<a href="https://github.com/csn-le/wave_clus">https://github.com/csn-le/wave_clus</a>
Chronux toolbox	<sup>57,58</sup>	<a href="http://chronux.org">http://chronux.org</a>
CircStat toolbox	<sup>59</sup>	<a href="https://www.mathworks.com/matlabcentral/fileexchange/10676-circular-statistics-toolbox-directional-statistics">https://www.mathworks.com/matlabcentral/fileexchange/10676-circular-statistics-toolbox-directional-statistics</a>

**RESOURCE AVAILABILITY**

The raw data supporting the current study have not been deposited in a public repository because of the complexity of the customized data format and the amount in size but are available from the corresponding author on request.

**Lead Contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, David Leopold ([leopoldd@mail.nih.gov](mailto:leopoldd@mail.nih.gov)).

**Materials Availability**

This study did not generate new unique reagents.

**Data and Code Availability**

The raw data supporting the current study have not been deposited in a public repository because of the complexity of the customized data format and the amount in size but are available from the corresponding author on request.

**METHOD DETAILS****Subjects**

Six rhesus monkeys (*macaca mulatta*, four females and two males, weighing 5.5 - 11.3 kg) were used in this study. All animals were surgically implanted with an MRI-compatible head post, and with chronic microwire electrode bundles in a face patch (Figure 1C; AM face patch for two animals, AF face patch for three animals and ML face patch for one animal) which was functionally localized using a standard fMRI block design<sup>60,61</sup> and/or naturalistic movie watching paradigm.<sup>60</sup> The apparatus and surgical implantation protocol have been described in detail previously.<sup>62</sup> All surgeries were performed under aseptic conditions and general anesthesia under isoflurane, and animals were given postsurgical analgesics and prophylactic antibiotics. During participation in the recording experiment, the animals were on water control and received their daily fluid intake during their testing (see below). Each subject's weight and hydration level were monitored closely and maintained throughout the experimental testing phases. All the experimental procedures and animal welfare were in full compliance with the Guidelines for the Care and Use of Laboratory Animals by U.S. National Institutes of Health and approved by the Animal Care and Use Committee of the U.S. National Institutes of Mental Health / National Institute of Health.

**Behavioral task and visual stimuli**

The animals were not required to perform a behavioral task that need cognitive efforts for distinguishing identities. With the passive viewing paradigm, neurons were examined for sensory processing of visual face stimuli. The monkeys sat in a primate chair in front of an LCD/OLED monitor with their head position stabilized by means of an implanted head post. They were required to maintain their gaze on a fixation point of 0.2° × 0.2° at the center of the monitor through a trial. In each trial, visual stimuli of 7° × 7° monkey face or 4° × 6.4° human face were presented for 300 ms in pseudo-random order followed by a 300 ms inter-stimulus interval (Figure 1B). The monkeys were rewarded with fruit juice for successfully maintaining fixation within a window of 1.5° - 2°, while their eye position was monitored using an infrared video-tracking system (EyeLink II; SR Research). Stimulus presentation, eye position monitoring, and reward delivery were controlled by MonkeyLogic software,<sup>63</sup> NIMH MonkeyLogic Software<sup>64</sup> or custom software courtesy of

David Sheinberg (Brown University, Providence, RI) running on a QNX computer in combination with another machine that run the psychtoolbox<sup>65</sup> in MATLAB (Mathworks). The monitor (either a ViewSonic 18" LCD monitor or LG 55" OLED monitor) was placed 91 or 90 cm in front of the monkey. Timing of stimulus presentation was recorded by a photodiode sensor that received signal from a small white square displayed on a corner of the screen at the same time of stimulus presentation.

The photo-realistic face stimuli were generated by morphing between face images (Figures 1A, S4, and S5) using a face-morphing software (Fantamorph, Absoft). Monkey faces were created based on twelve monkey face photographs of the NIH colony, prepared and provided by Dr. Olga Dal Monte, and human faces were created based on twenty four human faces selected from the FEI face database (<https://fei.edu.br/~cet/facedatabase.html>). Both monkey faces and human faces were not familiar to the subjects.

We first created the average face that resides at the center of the face space and approximates a norm,<sup>9</sup> by a morphing algorithm using point by point correspondence of the individual face. Then the stimuli were created by morphing along face-space trajectories between the average and original faces, that is to say an identity trajectory.<sup>12</sup> By extending the identity trajectories past the original faces, caricature faces that had exaggerated features were created. We also extended the identity trajectories away from the average, creating so-called anti-faces that had “opposite” features of the original faces. The morphing percentage was taken as the “identity level,” in which 0% identity level corresponded to the average face and 100% identity level corresponded to the original full identity faces. We assigned negative identity levels of < 0% to anti-faces. Stimuli were generated at identity levels of -100, -50, -25, -10, 0, 10, 25, 50, 75, 100, 125, 150 and 200%. In total, 145 monkey faces and 289 human faces were created and used. In this study, we morphed the faces based on shape information and applied average texture to all the images to exclude confounding effects of the texture smoothness which could be maximized at the norm stimulus. For the experiment in Figures 3C–3F, we generated stimuli along 8 identity trajectories at the identity levels of 1, 2, 5, 10, 25, 50, 75 and 100%, based on both shape and texture information. For the experiment in Figures S5A–S5H, we morphed between pairs of original identities at the identity levels of 0, 10, 25, 50, 75, 90 and 100% (for this morphing, we defined 0 and 100% identity levels as one of paired original faces). For the control experiment in Figures S5M–S5R, we generated contour-only stimuli based on the 145 monkey faces and 289 human faces. The inside of the contour-only stimuli was uniformly painted with the average color of original textures to eliminate all the internal face structures. In addition to the face stimuli, 12 object and pattern images were also shown to ensure that the same neurons were held across experimental sessions on different days. The area of the face stimuli ranged between 76.3% to 127.5% (human faces) and between 87.3% to 127.0% (monkey faces) with compared to the area of the average face. The stimulus size was balanced across different identities and there was no significant bias of the size with neither absolute identity strength ( $p = 0.17$  (human faces) and  $p = 0.60$  (monkey faces), one-way ANOVA) nor identity level ( $p = 0.40$  (human faces) and  $p = 0.85$  (monkey faces), one-way ANOVA).

It would be unwise to presume that the averaged monkey and human faces in this study were identical to the statistical averages of the faces that the monkeys had previously seen. On the other hand, it may be even more unwise to presume that they are unrelated. In the study of human face composites, it is long known that separate populations of randomly selected faces can appear similar, as long as there are around 16-32 faces being averaged (e.g.,<sup>66</sup>). Putting aside the issue of attractiveness, which as a central issue in much of this research on face composites, this suggests that there is good reason to expect that the average rhesus monkey face (composed of 12 individuals) and the average Caucasian human face (composed of 24 individuals) bears similarity to the monkey subjects’ previously experienced statistical average of monkey and human faces. But we did not test this systematically in our study, nor did we systematically explore the creation of face morphs using different individuals.

### Electrophysiology

Extracellular neuronal signals were recorded with 32, 64 or 128 chronically implanted NiCr wires that permitted tracking of individual neurons over multiple recording sessions.<sup>61,62</sup> The microwire electrodes were designed and initially constructed by Dr. Igor Bondar (Institute of Higher Nervous Activity and Neurophysiology, Moscow, Russia) and subsequently manufactured commercially (Microprobes). The recorded neuronal signals were amplified and digitized at 24.4 kHz in a radio frequency-shielded room by PZ5 NeuroDigitizer (Tucker-Davis Technologies), and then stored to an RS4 Data Streamer controlled by an RZ2 BioAmp Processor (Tucker-Davis Technologies). A gold wire inserted into a skull screw was used for ground. Broadband signals (2.5-8 kHz) were collected, from which individual spikes were extracted offline using the WaveClus software<sup>56</sup> after filtering between 300 and 5000 Hz. Local field potentials (LFPs) were extracted offline from the same broadband signals by zero-phase, bi-directional fourth-order Butterworth band-pass filtering between 3-200 Hz. Event codes, eye positions and a photodiode signal were also stored to a hard disk using OpenEX software (Tucker Davis Technologies).

The method for longitudinal identification of neurons across days was described in detail previously.<sup>61,62,67</sup> Briefly, spikes recorded from the same channel on different days routinely had closely matching waveforms and interspike interval histograms, and were provisionally inferred to arise from the same neurons across days. This initial classification based purely on waveform features and spike statistics was tested against the pattern of stimulus selectivity and temporal structure of the neurons’ firing evoked by visual stimulation. Guided by our previous observations that neurons in inferotemporal cortex respond consistently to statically presented visual stimuli across days and even months,<sup>62,67</sup> we used the distinctive visual response pattern generated by isolated spikes as a neural “fingerprint” to further disambiguate the identity of single units over time. The longitudinal aspect of the recording allowed us to collect neuronal responses for a large number of stimuli. By spending  $2.96 \pm 0.89$  days, we recorded neuronal

responses  $19.91 \pm 5.65$  times for each of 536 stimuli, resulting in  $10673.3 \pm 3026.2$  trials in total. In addition, we recorded responses from single neurons for 75 days, which enabled us to evaluate within-session adaptation for the stimuli (Figures 3C–3F).

### Long duration recordings to investigate role of visual adaptation

We exploited the stable recordings of our microwire electrodes to track a population of 13 neurons in the AF face patch across 75 sessions spanning three months beginning with the first presentation of each stimulus. As above, we used the selectivity “fingerprint” to establish the identity of individual neurons for tracking during this period, a method that has been used effectively previously.<sup>61</sup> For this data collection, only positive face identities, between 0 and 100% identity were used, with no anti-faces or caricatures. Collecting data with the same stimulus set across days, from the first presentation of each stimulus, allowed for the evaluation of changes both within and between sessions. The 75 sessions allowed us to draw selectively from the first presentation of each stimulus during the day, averaging this first presentation across sessions, to determine whether neural responses to the first presentation differed critically from those to later presentations. This allowed us to test the hypothesis that the V-shaped tuning of face-selective neurons might arise due to short-term visual adaptation.

### QUANTIFICATION AND STATISTICAL ANALYSIS

Stored neuronal response data were analyzed offline with MATLAB software (Mathworks, MA). All the data in the text was expressed as mean  $\pm$  SD unless otherwise stated. Error bars in figures are standard error unless otherwise stated. To evaluate V-shaped tuning, firing rate responses of neurons to each stimulus were calculated from 200 to 500 ms after the stimulus onset. Stimulus selectivity of the responses were evaluated for each neuron by two-way analysis of variance (ANOVA) [two factors: identity trajectories (identity of 12 original monkey faces or 24 original human faces of the trajectory) and identity level (12 identity levels ranging from –100% to 200%, excluding 0% identity face that is shared across all the trajectories)]. Neurons which showed one or more main effect and/or interaction between factors ( $p < 0.05$  after Bonferroni correction) were considered as face-selective and further analyzed.

Mean tuning for identity level were calculated by averaging across all the identity trajectories for each neuron (Figure 2B). Population-averaged tuning was calculated by averaging normalized firing rate response (NFR), which was calculated from the maximum and baseline responses of each neuron (Figure 2C). Baseline response was calculated from 200 ms before the stimulus onset to 50 ms after the stimulus onset across all the trials of each neuron. Maximum response was defined as the response to the stimulus that elicited largest response during 200 to 500 ms after stimulus onset, either excitation or suppression, as compared to the baseline response. Median differences between the maximum response and baseline response were 7.01 and 6.33 Hz for the monkey and human faces, respectively. The median response to the non-face images was 1.93 Hz (AM: 1.91 Hz, AF: 2.52 Hz and ML: 1.02 Hz), which was significantly smaller than the maximum response to monkey and human faces ( $p < 0.001$ , Wilcoxon’s sign-rank test). For comparison, we also calculated the response during earlier time window (60–200 ms after stimulus onset), and the median maximal firing rate in this period was 14.97 Hz, and the median difference from the baseline was 9.27 Hz. The response for each stimulus was normalized by subtracting the baseline response and then dividing it by the absolute difference between the baseline and maximum response of the neuron, resulting in normalized firing rate value that ranges from –1 to 1 (–1 or 1 corresponded to the maximum response and 0 corresponded to the baseline response). Neurons whose mean response was less than the baseline were considered as suppressive neurons ( $n = 62$  and 74 for monkey and human faces, respectively), and the sign of their normalized firing rates was inverted before calculating population-averaged response, since the tuning shape tended to be inverted for those neurons ( $R = 0.31$  between normalized mean response and V-shape index). Also, there was a significant correlation between the V-shape index and the maximum firing response change from the baseline ( $R = 0.30$ ,  $p < 0.001$ ).

For each neuron, the V-shaped tuning pattern was quantified by modeling the tuning with two regression lines, which included four parameters: anti-face side (left side) slope, caricature-face side (right side) slope, identity level of the vertex (V-location, x axis value of a crossing point of two regression lines) and NFR of the vertex (y axis value of the crossing point). The optimal modeling parameters were estimated by least-squares technique with maximum of 2,000 iterations. If parameters were not estimated after 2,000 iterations, or if the estimated vertex was located too right or left where not enough (fewer than two) identity levels were available to estimate either the right or left side slope, the initial value was changed and the estimation process was repeated again. If the parameters were not successfully estimated after a total of 100,000 iterations, we dismissed the two-line model and adopted a single-line linear regression for that neuron. Of 186 face-selective neurons, response activities of 172 (92.5%) neurons were modeled with the two-line model, while baseline activities of 121 (65.1%) neurons were modeled with the two-line model. The estimated tuning slopes were represented as a change of NFR per 100% identity level. The tuning pattern of each identity trajectory was quantified by modeling with two regression lines using three parameters: anti-face side (left side) slope, caricature-face side (right side) slope and NFR of the vertex. Identity level of the vertex in each trajectory was fixed at 0% to prevent overfitting due to fewer data points. The optimal modeling parameters were estimated by least-squares technique as modeling for each neuron. The average face, whose identity level was 0%, was not included in the regression to avoid duplicated contribution to multiple identity trajectories since it was shared across different identity trajectories. In Figures 5 and S6, each identity trajectory was also modeled by combination of linear ramp tuning and V-shaped tuning (see Figure S6A for the model function). The basis functions of the linear and V-shaped tuning were both designed to cross the origin [0, 0] and to cover a range of  $\pm 0.5$  of normalized firing rate in the y axis. V-shape index was defined as the difference of slopes between caricature and antiface sides. To evaluate tuning across different identity trajectories (“tangential tuning” in Figure 5C), for each neuron, we first rank-ordered trajectories by the response to 200% identity level. Then the rank-

ordered-responses were averaged across neurons and displayed as circular heatmaps in Figure 5B. The identity trajectories of the linear ramp tuning and V-shaped tuning components were also sorted by the observed raw response to 200% identity level before averaging for Figure 5B. The “tangential” tuning pattern across identity trajectories were quantified at each identity level by modeling with two regression lines using three parameters: slope of worst-identity side, slope of best-identity side and NFR of the vertex. The optimal modeling parameters were estimated by least-squares technique as modeling for identity trajectories. Identity tuning slope of each neuron across tangential direction was defined as the average of all the tangential tuning slopes. Before the averaging, sign of the slopes was inverted for antifaces which had opposite feature of original faces. Identity tuning slope across radial direction was defined as the average of the modeled slopes of each identity trajectory. After averaging antiface-side slopes and caricature-side slopes separately, absolute values were taken to measure the sensitivity for the identity level.

Spike trains were smoothed by convolution with a Gaussian kernel ( $\sigma = 30$  ms) to obtain spike density functions (SDF) for each stimulus (Figure 2A). To evaluate the temporal dynamics of firing rate response and regression slopes (Figure 4), firing rates were calculated with a sliding window of  $\pm 25$ -ms width moving in 1-ms steps. To evaluate the temporal dynamics of linear and V-shaped tuning for each identity trajectory, sliding window of  $\pm 50$ -ms width was used (Figures S6D–S6G). Firing rate responses at each time point were normalized in a similar manner to the normalized firing rate as described above, with the baseline response which was calculated between 200 ms before and 50 ms after the stimulus onset and the maximum response which was calculated between 200 and 500 ms after the stimulus onset. The onset latency was defined as the time after the stimulus onset when the firing rate exceeded  $\pm 2$  SD of the firing rate during the 250-ms baseline period for 10 consecutive bins and reached to 3 SD of the baseline firing rate at least once. The fall time was defined as the time from the stimulus onset when the firing rate decreased below  $\pm 2$  SD of the firing rate during the baseline period for 10 consecutive bins and reached to 1 SD of the baseline firing rate at least once. The regression slopes were estimated at each time point of the normalized firing rate time course in the same manner of the two-line modeling described above. To evaluate dynamics across firing rate response, neuronal tuning and coherence (Figure 6E), we calculated peak time since it depends less on time window width, which needed to be relatively large to ensure a high enough signal-to-noise ratio for coherence computation. Peak time was also used for comparing  $\beta$  coefficients and correlations between linear and V-shaped tuning (Figures S6E and S6G) which needed relatively large time window for regression of each identity trajectory from fewer number of trials. The peak time was defined as the time point after the stimulus onset when the time course reached its maximum value between 0 to 400 ms after the stimulus onset.

We calculated spike-spike and spike-LFP coherence to evaluate synchronized computations that may be processed across neural circuits. Coherence spectrum was estimated with multi-taper spectral methods<sup>57,58</sup> using Chronux toolbox (<http://chronux.org>) and CircStat toolbox.<sup>59</sup> For comparisons between low and extreme identity faces, five Slepian taper functions were used with a  $\pm 75$  ms (150-ms width) time window with  $\pm 10$  Hz (20-Hz bandwidth) resolution. Faces whose absolute identity level was equal to or less than 50% were classified as low identity faces, and faces whose absolute identity level was equal to or larger than 75% were classified as extreme identity faces. Eleven Slepian taper functions were used with a slightly larger window,  $\pm 100$  ms and  $\pm 15$  Hz resolution, for comparison across identity levels which had fewer trials per data point. In addition, we excluded neurons whose mean total spike number, which summed across trials and faces at the same identity level, was less than 200 (this corresponds to  $0.91 \pm 0.30$  spikes/trial (mean  $\pm$  SD)) for the comparison across identity levels. For human faces in Figure S7N and for non-spike-remove control analysis in Figure S7I, we excluded neurons whose spikes per identity level was less than 400. Coherence was calculated during the later sustained response phase, starting at 150 ms after stimulus onset. Coherogram and coherence time course were computed with sliding window of  $\pm 75$ -ms width moved at 1 ms steps. Spike-spike coherence was calculated between all the possible neuronal pairs that were simultaneously recorded in the same session. For calculation of spike-LFP coherence, LFP signals were averaged across all the recording channels of the microwire bundles except for noisy or dead channels that couldn't picked up any unit activities. To avoid potential contamination of LFP signals by spike waveform, the channel that recorded the spikes for a given coherence calculation was also excluded from the averaging of LFPs. Based on coherograms (Figure S7G) and frequency spectrums of coherence (Figure 6B), we averaged coherence and phase in low- $\gamma/\beta$  range (15–50Hz, for spike-spike coherence) and low- $\gamma$  (30–50 Hz, for spike-LFP coherence) for comparisons across identity levels and between low and extreme identities. The number of trials was balanced across conditions: the difference in number of trials was  $1.52 \pm 0.7\%$  for comparisons between low and extreme identity levels and  $2.9 \pm 2.2\%$  for comparisons across identity levels. Thus potential effect of bias from different sample sizes was minimized.

We examined the effect of identity levels on the coherence, which also affected firing rate responses of each neuron. Although coherence is designed to be invariant over differences of firing rate by normalization,<sup>57,58</sup> some aspects of the coherence measure, like signal-to-noise ratio, could be influenced by firing rate. To rule out any potential contribution of firing rate differences to coherence, we matched firing rate across conditions before coherence calculation by randomly dropping spikes.<sup>24,25,68</sup> During the comparisons between low and extreme identities, for each neuron, we randomly dropped spikes of one condition that showed higher firing rate until it matched with the other condition. During the comparisons across identity levels, firing rates were matched to an identity level that showed minimum firing rate response. For calculating coherograms and coherence time courses, we separated time course into 200-ms-width bins and matched firing rates in each bin. Matching of firing rate did not affect essential aspects of coherence of our data; even without any firing rate matching, low identity faces elicit higher  $\gamma/\beta$  coherence during late phase of the response (Figures S7I–S7L). Matching of firing rate reduced absolute coherence values, especially at high- $\gamma$  range over 100Hz in higher identity level conditions which elicit larger firing response (Figure 6B compared with Figure S7J), suggesting that the observed coherence at high- $\gamma$  range might be partially reflecting firing rate response.

To control coherence arising from stimulus-locked response, trial-shuffled coherence was also calculated. For each pair of neurons and each neuron-LFP pair, the order of trials was randomly shuffled within each stimulus in a neuron while maintaining the trial order of the other signal (the other neuron or LFP), and then coherence was computed based on the trial-shuffled spike trains. The trial-shuffled coherence was created 1000 times, and their mean was subtracted from the original coherence to produce trial-shuffle-subtracted coherence. For controlling coherogram and coherence time course data, one-trial shifted control was subtracted from the original to produce a trial-shift-subtracted coherogram and coherence time course. The trial-shuffled or trial-shifted control were subtracted in all the coherence/coherogram data displayed in Figures, except for the “Raw” coherence in [Figures 7B, 7H, 7I, and 7N](#).