nature neuroscience

A face feature space in the macaque temporal lobe

Winrich A Freiwald[1–4,6], Doris Y Tsao[1–3,5,6] & Margaret S Livingstone[1]

The ability of primates to effortlessly recognize faces has been attributed to the existence of specialized face areas. One such area, the macaque middle face patch, consists almost entirely of cells that are selective for faces, but the principles by which these cells analyze faces are unknown. We found that middle face patch neurons detect and differentiate faces using a strategy that is both part based and holistic. Cells detected distinct constellations of face parts. Furthermore, cells were tuned to the geometry of facial features. Tuning was most often ramp-shaped, with a one-to-one mapping of feature magnitude to firing rate. Tuning amplitude depended on the presence of a whole, upright face and features were interpreted according to their position in a whole, upright face. Thus, cells in the middle face patch encode axes of a face space specialized for whole, upright faces.

Viewing the world, we are confronted by myriad visual objects. How does the brain extract these objects from the incoming bits and pieces of information? The representation of an object (as opposed to a spot, edge or smear of color) must involve a mechanism for representing its gestalt. What is the neural mechanism by which curves and spots are assembled into coherent objects? And how does the brain preserve fine distinctions between individual objects throughout this process?

We have a good understanding of how edges, a form common to all objects, are coded by cells in area V1 (ref. 1), but the mechanisms by which the brain analyzes shapes at the next level are less understood. One major experimental difficulty is that there are so many different forms and no clear approach to choosing one set of forms over another for testing each cell. It is clear, however, that any study of object recognition must employ a restricted set of all possible forms. The challenge, then, is to find a way to constrain the stimulus space by incorporating prior knowledge about the cells' stimulus preferences.

Functional magnetic resonance imaging (fMRI) provides a solution to this challenge[2]. Using fMRI in macaque monkeys, we found a cortical area in the temporal lobe that is activated much more by faces than by nonface objects[3]. Subsequent single-unit recordings showed that this area, the middle face patch, consists almost entirely of face-selective cells[4]. Targeting single-unit recordings to this area provides a powerful strategy for dissecting the mechanisms of high-level form coding in a homogeneous population of cells that are selective for a single type of complex form.

The space of faces still contains an infinite variety of particular forms (as it must for face perception to be useful). An effective strategy to further reduce the stimulus space is to represent faces as cartoons[5]. This approach has several justifications. First, the nameable features making up a cartoon (eyes, nose, etc.) correspond to the brightness discontinuities of real faces[6], and thus approximate the representation relayed by early visual areas. Second, a cartoon face is clearly perceived as a face, and cartoons therefore effectively convey the overall gestalt of a face. Thus, cartoons constitute appropriate stimuli for studying the neural mechanisms of face detection. Third, cartoons convey a wealth of information about individual identity and expression through both the shape of individual features (for example, mouth curvature), and the configuration of features (for example, inter-eye distance). Therefore, cartoons constitute appropriate stimuli for studying the neural mechanisms of face differentiation. Finally, cartoon shapes can be completely specified by a much smaller set of parameters than would be required to specify individual pixel values of images of real faces, thus simplifying analysis. For all these reasons, cartoons provide a powerful and effective way of simplifying the space of faces.

We asked how cells in the middle face patch detect and differentiate faces. We used fMRI to localize the middle face patch and then targeted it for single-unit recording. We first measured responses of cells to photographs of faces and other objects; the results of this test confirmed the selectivity of the middle face patch for faces. We next measured the responses of these cells to pictures of both real and cartoon faces and found that responses to cartoon faces were comparable to those of real faces. Armed with this knowledge, we then probed cells with systematically varying cartoon faces to address three fundamental questions: what is the mechanism for face detection, what is the mechanism for face differentiation and what role, if any, does facial gestalt have in face differentiation.

RESULTS

Selectivity for real and cartoon faces

We determined the locations of the middle face patches in the temporal lobes of three macaque monkeys with fMRI (**Fig. 1a**) and then targeted one middle face patch in each monkey for electrophysiological recordings. For every cell that we recorded (286 total), we first determined the face selectivity of the cell by measuring its response to images of 16

[1]Department of Neurobiology, Harvard Medical School, Boston, Massachusetts, USA. [2]Centers for Advanced Imaging & Cognitive Sciences, Bremen University, Bremen, Germany. [3]Division of Biology, California Institute of Technology, Pasadena, California, USA. [4]The Rockefeller University, New York, New York, USA. [5]Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, Massachusetts, USA. [6]These authors contributed equally to this work. Correspondence should be addressed to W.A.F. (winrich.freiwald@googlemail.com).

Received 17 April; accepted 3 June; published online 9 August 2009; doi:10.1038/nn.2363

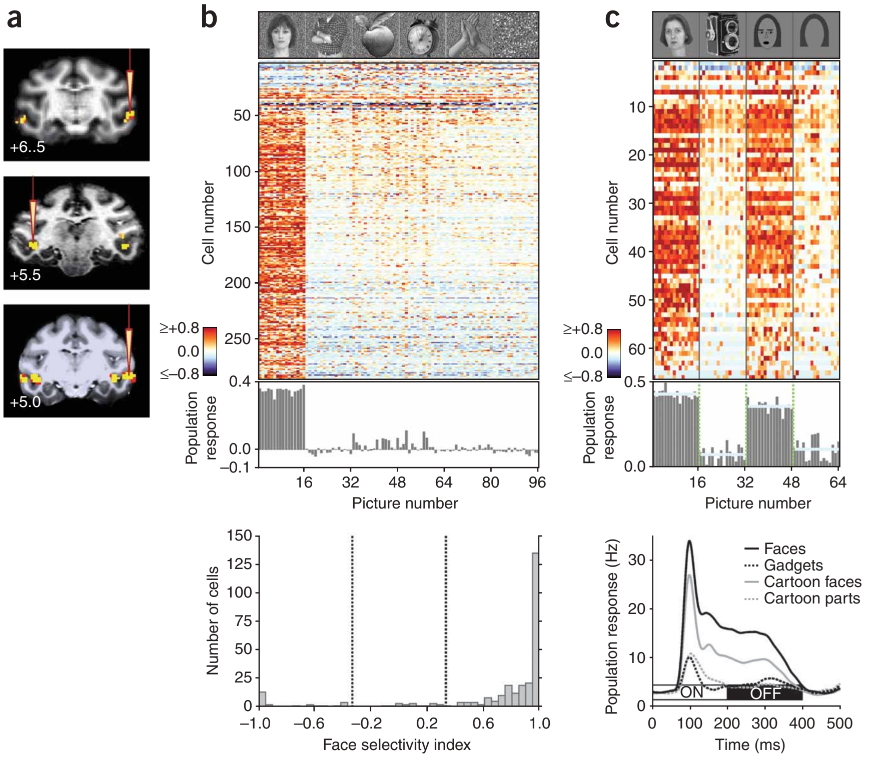

Figure 1 Selectivity of the middle face patch for real and cartoon faces. (**a**) fMRI-defined middle face patches ($P < 10^{-4}$) shown on coronal slices (millimeters anterior to inter-aural line indicated in bottom left) for the three monkeys used in this study (monkeys A, T and L from top to bottom), with recording sites marked by electrode icons. (**b**) Top, response profiles of all 286 cells tested with 96 pictures of faces, bodies, fruits, gadgets, hands and scrambled patterns (16 images per category, one example per category is shown) and average normalized population responses (bar graph). Bottom, distribution of face-selectivity indices. 268 of 286 neurons (94%) were face selective (that is, face-selectivity index larger than $\frac{1}{3}$ or smaller than $-\frac{1}{3}$, dotted lines). Out of the subset of 241 neurons that met the stringent criterion for visual responsiveness (see Online Methods), 230 (95%) were face selective. (**c**) Top, face and cartoon selectivity of 66 cells tested with 64 images of faces, gadgets, cartoons and cartoon face parts (data are presented as in **b**). Bottom, time course of population response to four stimulus categories: faces, gadgets, cartoons and cartoon parts. Stimuli were presented for 200 ms and separated by 200-ms interstimulus intervals.

frontal faces, 64 nonface objects and 16 scrambled patterns (**Fig. 1b**; examples of stimuli are shown in **Supplementary Fig. 1**). Across the population, 94% of the cells were face selective (**Fig. 1b**).

We then compared the responses of middle face patch neurons to cartoon faces and to real faces. We recorded responses of 66 cells to images of 16 real faces, 16 nonface objects, 16 cartoon faces and 16 isolated parts of cartoon faces. The cartoon faces were constructed from seven elementary parts (hair, face outline, eyes, irises, eyebrows, mouth and nose) whose shape and position were matched to those in the real faces (examples are shown in **Supplementary Fig. 1**). Across the population, the mean response magnitude to cartoon faces was 83% of the response to real faces, whereas the mean response to nonface objects was 17% and the response to cartoon parts was 24% of the response to real faces (**Fig. 1c**). Response ranges to real and cartoon stimuli were largely overlapping; all but one cell responded more to at least one cartoon face than to one of the real faces, and a cartoon face elicited the best or second best response in 45% of the cells. Furthermore, the selectivity of cells for real and cartoon faces was correlated ($r = 0.62$, $P < 0.001$) and response time courses were similar (mean correlation across cells, $r = 0.90$, $P < 0.001$; **Fig. 1c**). Thus, although cartoon faces lack many of the details found in real faces, such as pigmentation, texture and three-dimensional structure, they constitute effective substitutes to middle face patch neurons. Therefore it is appropriate to use cartoon stimuli to probe the detailed mechanisms of face representation by these neurons (in addition, **Supplementary Fig. 1** provides psychophysical data that our cartoon stimuli successfully captured essential aspects of face identity).

Face detection: selectivity for face parts

Cartoon faces can easily be decomposed into parts (without introducing additional edges, as would be the case with cutting up images of real faces) and therefore are ideally suited for studying the mechanisms of face detection. We presented a set of all 128 (2^7) possible decompositions of a seven-part cartoon face to 33 middle face patch neurons (**Fig. 2a**; stimuli are shown in **Supplementary Fig. 1**). ANOVA revealed that, across the population, cells were directly influenced by at least one, and at most four, face parts (**Fig. 2b**). This first order effect explained half of the response variance (52%) on average. In addition, a majority of cells (78%) showed significant pair-wise interactions between their part responses ($P < 0.005$; **Fig. 2b**); these second order effects explained an additional 18% of the variance. This dependence of responses on multiple parts and part interactions shows that middle face patch neurons are not simple feature detectors. However, because 70% of the response variance was explainable by first and second order effects alone, middle face patch cells are not highly nonlinear holistic cells either.

Notably, middle face patch neurons did not have a single best stimulus that uniquely elicited the maximum firing rate. In particular, the response magnitude to whole cartoon faces was only 42% of that of the summed response to the seven face parts on average (**Fig. 2c**; see ref. 8). As a consequence, the same cell often fired at its maximum rate to both the whole face and to a variety of partial faces (**Fig. 2d**), a property that is useful for face detection.

Face differentiation: encoding of facial features

In the previous experiment, we determined selectivity for the presence of various face parts. We next investigated selectivity for the geometric shape of various face parts (for example, nose width). For this purpose, we used the same cartoon face stimulus described above (comprising seven parts), but now specified the geometry of parts and part relations by 19 different parameters, each with 11 values (six of these parameters are illustrated in **Fig. 3a** and all 19 parameters are listed in **Fig. 3b** and are defined in the Online Methods). The stimulus was presented in rapid serial visual presentation mode, with each parameter being updated randomly and independently every 117 ms (**Supplementary Video 1**). This approach allowed us to probe in detail the mechanisms by which cells distinguish different faces.

We first asked whether cells in the middle face patch are tuned to simple facial features. For each stimulus dimension, we computed a time-resolved tuning curve (see Online Methods and **Supplementary**

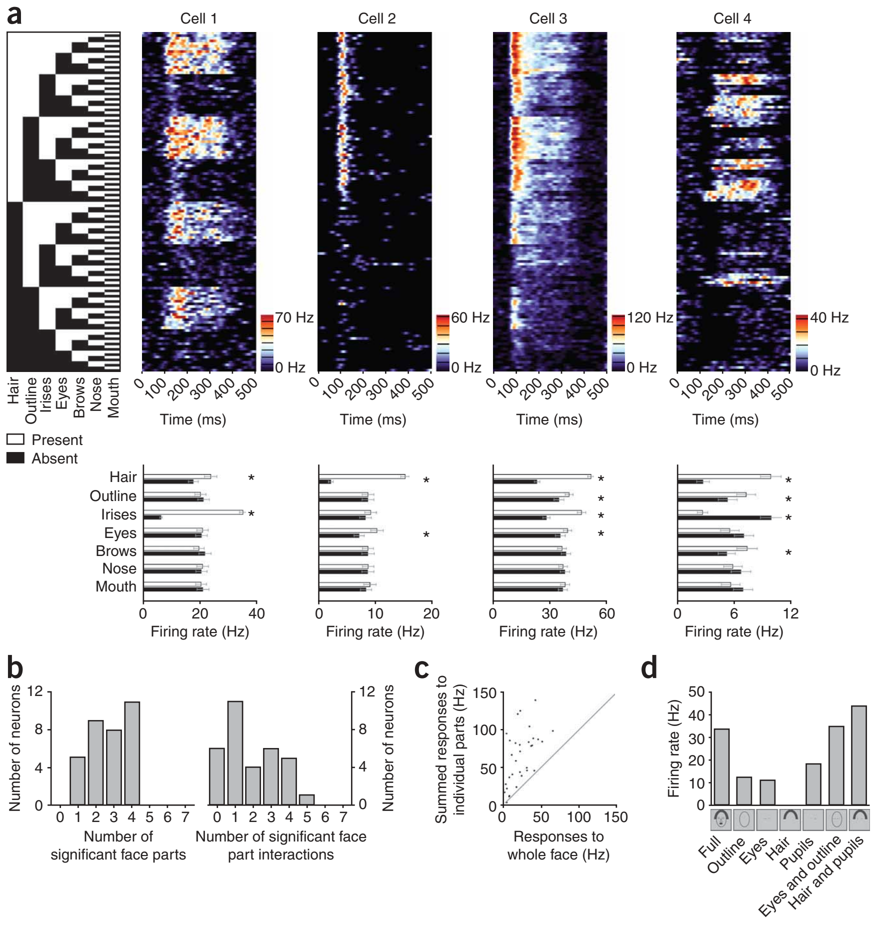

Figure 2 Selectivity for face parts. (**a**) Stimulus conditions for the cartoon face decomposition experiment (left) and responses of four example cells. All combinations of seven face parts (hair, outline, irises, eyes, eyebrows, nose and mouth) were shown, including the whole cartoon face with all features (top row) and a gray background without any face features (bottom row). Top right, responses are shown as a function of time and stimulus condition. Bottom, average responses in the presence (white bars) or absence (black bars) of a given face part. * indicates significant modulation ($P < 0.005$). Cell 1 fired significantly more strongly when irises were present and when hair was present ($P < 0.05$). Cell 2 was influenced by two parts, and cells 3 and 4 by four cell parts. Cell 4 responded more strongly when irises were absent than when they were present. In cell 4, interactions between face parts were stronger than in the other cells, giving rise to the less regular appearance of responses across conditions. (**b**) Distributions of the number of face parts (left) and the number of pair-wise interactions (right) that exerted a significant influence on cell firing for 32 cells ($P < 0.005$). At most 5 of the 21 possible feature interactions were significant ($P < 0.005$). (**c**) Scatter plot of responses to the whole face (abscissa) versus the sum of the responses to the seven parts (ordinate) for all 32 cells. (**d**) Responses of an example cell to the full cartoon stimulus (left), the four face parts that modulated activity significantly (outline, eyes, irises enhancing and hair suppressing) and two combinations of these two parts ($P < 0.005$).

Fig. 2) that captures how feature values (−5 to +5) influence the firing rate as a function of poststimulus time (0–400 ms). Comparing each time-resolved tuning curve to a distribution of shuffle predictors allowed us to determine whether a given feature dimension exerted a significant influence on the spiking of a cell ($P < 0.001$, see Online Methods). Tuning started as early as 75 ms after feature change. The time window over which tuning occurred overlapped with the time window of face selectivity, although the latter (**Fig. 1d**) was determined with a very different stimulus presentation regime (**Supplementary Fig. 3**).

Individual cells were modulated by different subsets of features (all 19 tuning curves for three example cells are shown in **Fig. 3b**). Across the population of 272 cells recorded in this experiment, 90% showed tuning to one or more stimulus dimensions (**Fig. 3c**). Individual cells were tuned to between one and eight stimulus dimensions (on average, 2.8 dimensions for cells that showed any tuning at all). Thus, each cell was specialized for a small number of feature dimensions. We did not find cells tuned to all aspects of a face.

Some features were represented more frequently in the population than others (**Fig. 3d**). The most popular parameter was face aspect ratio, to which more than half the cells (59%) were tuned, followed by iris size (46%), height of feature assembly (39%), inter-eye distance (31%) and face direction (27%). Thus, the most important feature categories were facial layout geometry and eye geometry; the feature categories mouth and nose were represented by five cells only. This pattern of results was robust and observed for different fixation conditions (**Supplementary Text 1** and **Supplementary Fig. 4**). The two most prominent dimensions were associated with the largest and the second smallest physical differences between feature values (face aspect ratio and iris size, respectively). Thus, the incidence of tuning was not directly related to the magnitude of physical change associated with each dimension, but the two were positively correlated overall (**Supplementary Text 2** and **Supplementary Fig. 5**). The low incidence of tuning for entire feature categories (mouth and nose) leads to a reduction in the dimensionality of the face space coded by the middle face patch.

Face differentiation: shape of feature tuning

In our example tuning curves (**Fig. 3b**), seven out of ten of the significantly modulated tuning curves were strictly ramp shaped, with a maximal response at one extreme of the feature range and a minimal response at the opposite extreme. Such ramp-shaped tuning dominated across the population as well. We sorted all of the significantly modulated tuning curves by the feature value that elicited the maximal response (**Fig. 4a**). Most tuning curves (62%) peaked at an extreme feature value. The distribution of maximal response values was so highly skewed to the extremes that tuning to the most extreme values was sixfold more frequent than tuning to the physically similar second-most extreme values. These extreme values extended to or even transgressed the limits of realistic face space. For example, inter-eye distances ranged from almost cyclopean to abutting the edges of the face, and the most extreme face aspect ratios were outside those of any living primate. Preference for ramp-shaped tuning was found in all feature dimensions (**Fig. 4b**). For most feature dimensions, both ends of the feature range were well represented, except for eyebrow slant and iris size; tuning to the popular parameter iris size was biased to the maximal iris size, with 33% of all recorded middle face patch neurons responding maximally to the largest irises.

Extreme feature values also suppressed activity more than intermediate ones did; 76% of all tuning curves exhibited minima at extremes, 14 times as many as for other feature values (**Fig. 4a**). Because each feature extreme elicited maximal responses in some cells and minimal ones in others, the population response to a set of face characteristics was amplified for extreme compared with intermediate feature values (**Fig. 4c**). Such amplification predicts that caricatured representations should be distinguished more easily than average faces, near the center of face space[7,8].

The bias for response minima and maxima at extreme feature values was characteristic not only of the population of cells, but also of individual cells' tuning curves. Two thirds (67%) of tuning curves with a maximum at one extreme showed the minimum at the opposite extreme (**Fig. 4d**). These tuning curves were, on average, almost linear

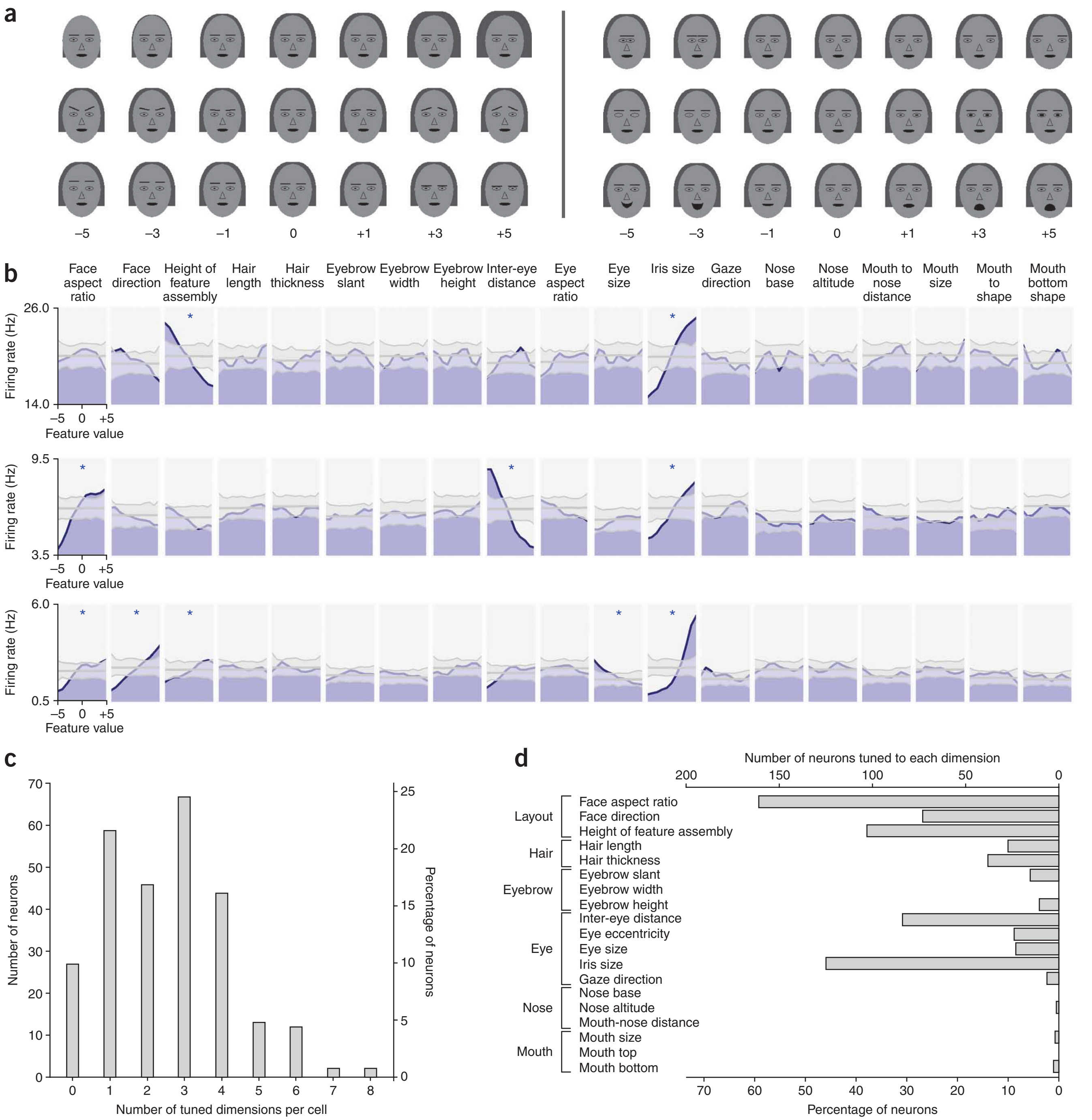

Figure 3 Cartoon tuning. (**a**) Example cartoon stimuli for six dimensions (hair width, eyebrow slant, eyebrow height, inter-eye distance, iris size and mouth top shape) with seven feature values each, spanning the entire range of values. (**b**) Tuning curves of three example cells. For each of the 19 feature dimensions, the tuning curve (blue) is shown at a delay corresponding to maximal modulation. Maximal, minimal and mean values from the shift predictor are shown in gray. Asterisks mark significant modulation ($P < 0.001$). (**c**) Distribution of the number of tuned dimensions per cell. (**d**) Distribution of the number and fraction of cells tuned to each of the 19 feature dimensions (together with definition of six feature categories).

in shape (**Fig. 4d**), thereby establishing a one-to-one mapping between the entire range of feature values and firing rate. Figuratively speaking, these cells measure feature dimensions.

Because response minima only rarely occurred near the midpoint of the cartoon face space, adaptation to the average face cannot account for the shape of these tuning curves. This is in contrast with the case of face-responsive (not necessarily face selective) cells anterior to the middle face patch, which have been reported to respond minimally (or maximally) to the average face[9]. We also tested for possible short-term adaptation effects or coding of feature changes rather than coding of features *per se*. It would be possible that a response is maximal to a feature extreme because extremes are preceded, on average, by the largest physical changes and not because extremes are special shapes. We performed a two-way ANOVA to test for interactions between responses to successive feature values of the same dimension. Of the 514 dimensions tested, only seven showed a significant interaction between subsequent feature values (ANOVA, $P < 0.05$). Thus, we do not find evidence for adaptation or change magnitude being involved in generating feature tuning.

Face differentiation: frequency of feature combinations

The typical middle face patch neuron is tuned to approximately three feature dimensions. Such conjunctive feature representation can aid in solving the binding problem[10], as cells tuned to overlapping feature constellations can uniquely represent a particular face. To achieve this computational goal, the population of neurons should create all

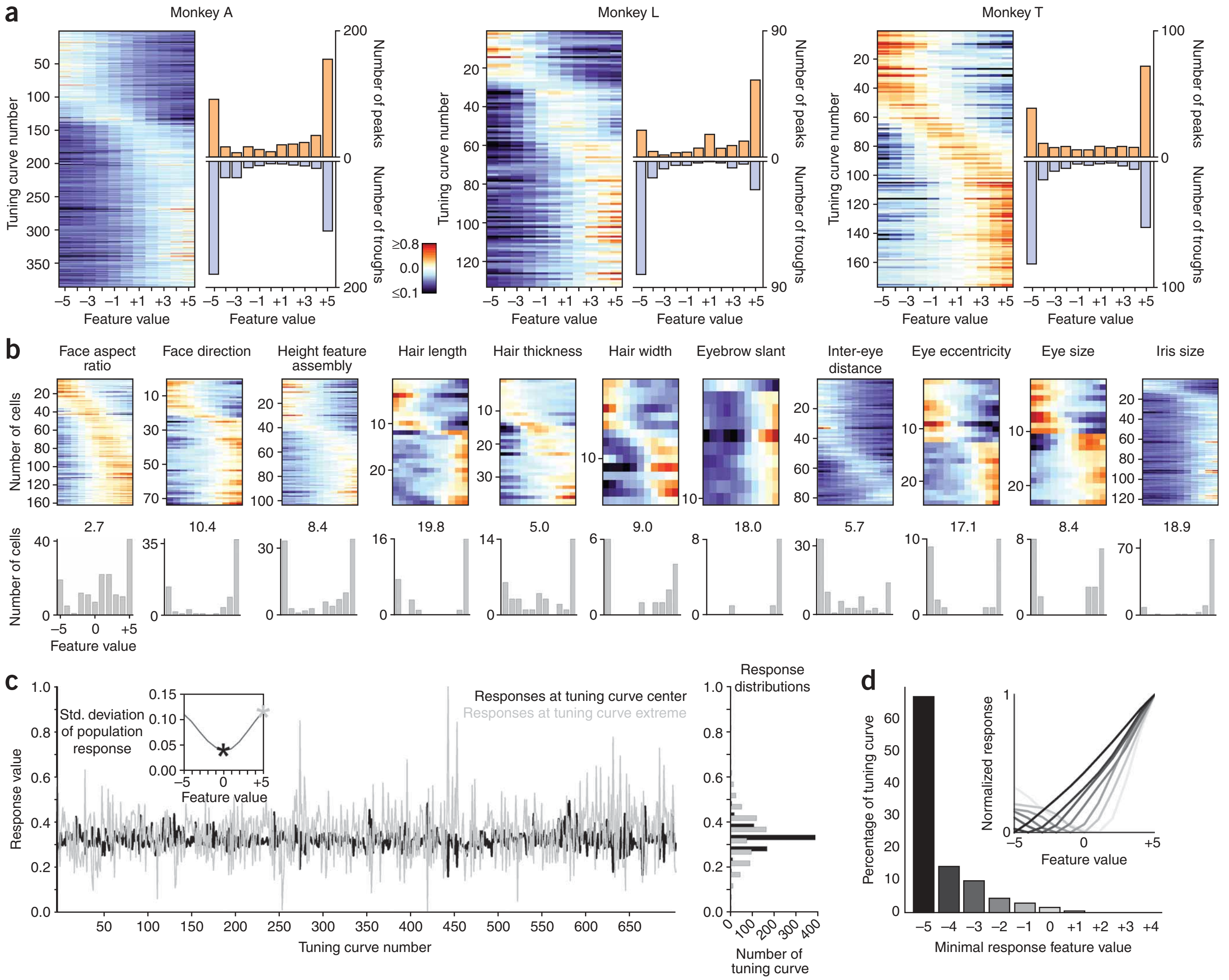

Figure 4 Preference for extreme feature values. (**a**) For each monkey, (left) all significantly modulated tuning curves are shown (area normalized, see Online Methods), sorted from top to bottom by feature value eliciting maximal response, (right) frequency distributions of feature values eliciting maximal (top) and minimal (bottom) responses. (**b**) Tuning curves (color code ranges from 0 to 1) and frequency distributions as shown in **a** for all features for which ≥10 cells were tuned. On top of each histogram, the relative over-representation of feature extremes is shown, quantified as the average number of cells with maximal response to an extreme (–5 and +5) divided by the average number of cells with maximal response to an intermediate feature value (–4 through +4). (**c**) Responses in all significantly modulated tuning curves to the average feature (0, black) and to one extreme (+5, gray). Frequency distributions as marginals (right) and s.d. over population responses to all feature values (inset, asterisks indicate values of the two marginal distributions). The s.d. at extremes was threefold larger than that at the center of face space (0.12 versus 0.04). (**d**) Distribution of tuning curve shapes with preference for extremes (feature values –5 or +5; tuning curves with preference for feature value –5 are horizontally flipped such that all curves for this analysis show a maximal response at feature value +5). Two thirds (67%) of these tuning curves had response minima at the opposite extreme. The inset shows average tuning curve shapes for the seven histogram bins of corresponding shade of gray.

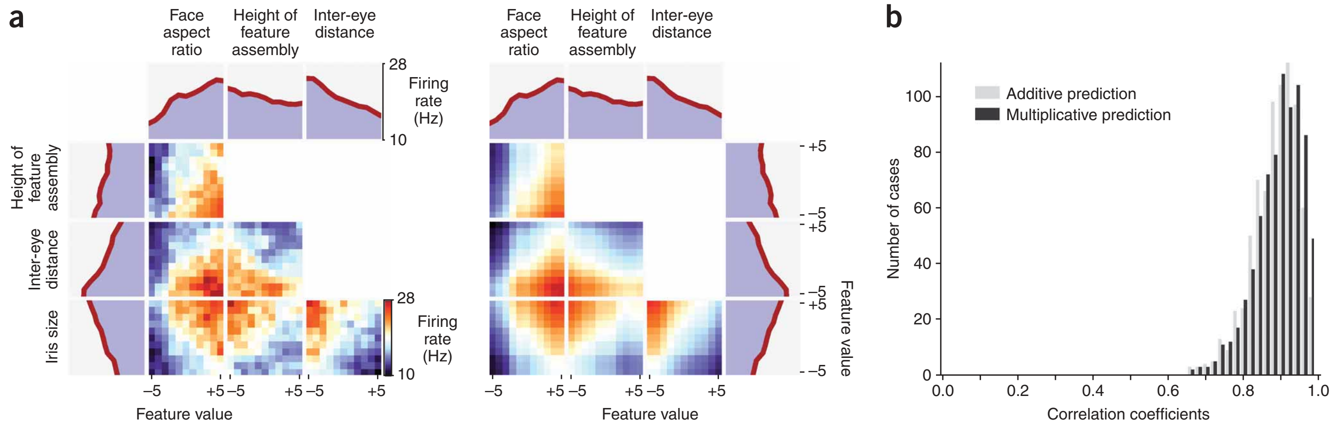

Figure 5 Joint tuning to feature dimension pairs. (**a**) Left, six joint tuning functions of an example neuron tuned to four features (tuning curves shown in marginals) and (right) predicted joint tuning functions based on a multiplicative model (correlation coefficients between actual and predicted joint tuning functions are indicated in the upper right corners). The full set of 19 × 19 joint tuning curves for this cell is shown in **Supplementary Figure 8**. (**b**) Correlation coefficients of all 771 joint tuning functions and their multiplicative (red) and additive (blue) predictors. Multiplicative predictors are slightly better (average correlation coefficients of 0.89 and 0.88, respectively, $P < 0.01$, U test).

possible conjunctions of feature tuning. Because cells were tuned to 14 of the 19 stimulus features, there were $C_2^{14} = 91$ feature combinations possible. We found only one feature combination that occurred more often than would be expected by chance combinations (8 times, compared with a $P = 0.01$ significance threshold of 7.7, details in Online Methods). Thus the population approaches maximal coverage of feature combinations even across different face parts. Furthermore, even neighboring cells often encoded different feature combinations (**Supplementary Text 3** and **Supplementary Fig. 6**).

We next asked how middle face patch neurons integrate different features. If feature integration is to preserve faithful measurement of individual features, then tuning to feature combinations should be separable into tuning to individual features[11]. Joint tuning functions computed by multiplying single-feature tuning curves were almost identical to the actual joint tuning functions (correlation coefficients between 0.90 and 0.97; **Fig. 5a**). In our sample of 771 pairs of significantly modulated marginal single-dimension tuning curves, the average correlation coefficient between the actual joint tuning functions and multiplicative predictors was 0.89 (an alternative analysis of separability is provided in **Supplementary Text 4** and **Supplementary Fig. 7**). This result does not imply, however, that feature combination was precisely multiplicative: prediction of joint tuning functions by additive predictors was almost as good as that by multiplicative predictors (average correlation coefficient of 0.88, which was significantly lower ($P < 0.01$) than multiplicative predictors, Wilcoxon rank sum test; **Fig. 5b**).

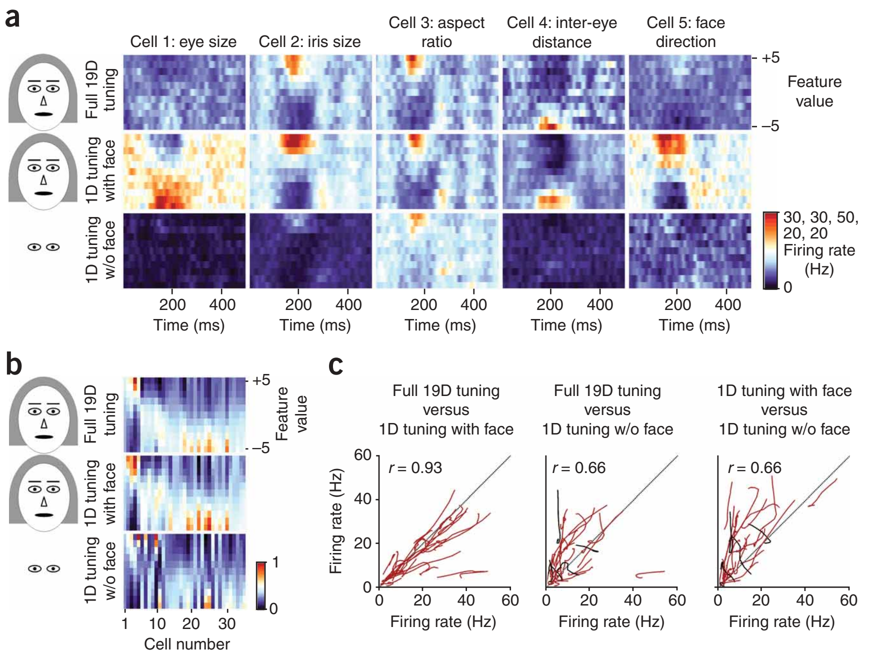

Figure 6 Integration of features and effects of face context. (**a**) Time-resolved tuning curves from five cells (left to right) showing the effect of three different contexts (top to bottom) on face feature tuning. The top row shows the responses to varying a single face feature in the full-dimension tuning experiment, in which all 19 features were simultaneously varied (19D). The middle row shows the responses to varying the same feature with the other 18 features fixed at their mean value. The bottom row shows responses to varying the same feature with all other features removed. Cells 1 and 4 lost tuning in the absence of the rest of the face, cells 2 and 5 weakened in the absence of the rest of the face, cell 3 showed stable tuning across conditions, cells 1 and 5 showed strengthened tuning when other features were static, and cells 2, 3 and 4 showed similar tuning strengths when other features were static as when they changed. (**b**) Area-normalized tuning curves in 35 cells in which significant tuning was observed in all three experiments ($P < 0.001$). Tuning curves were sorted by the position of maximal response in the full tuning experiment (left). Tuning curves were similar for single-dimension tuning with the face present and less so than when the rest of the face was absent. (**c**) Correlations between tuning curves in 35 cells in the three experiments. Positively correlated tuning curves are shown in dark red, negatively correlated ones in black and the average correlation coefficients across all curves are shown in the upper left of each curve.

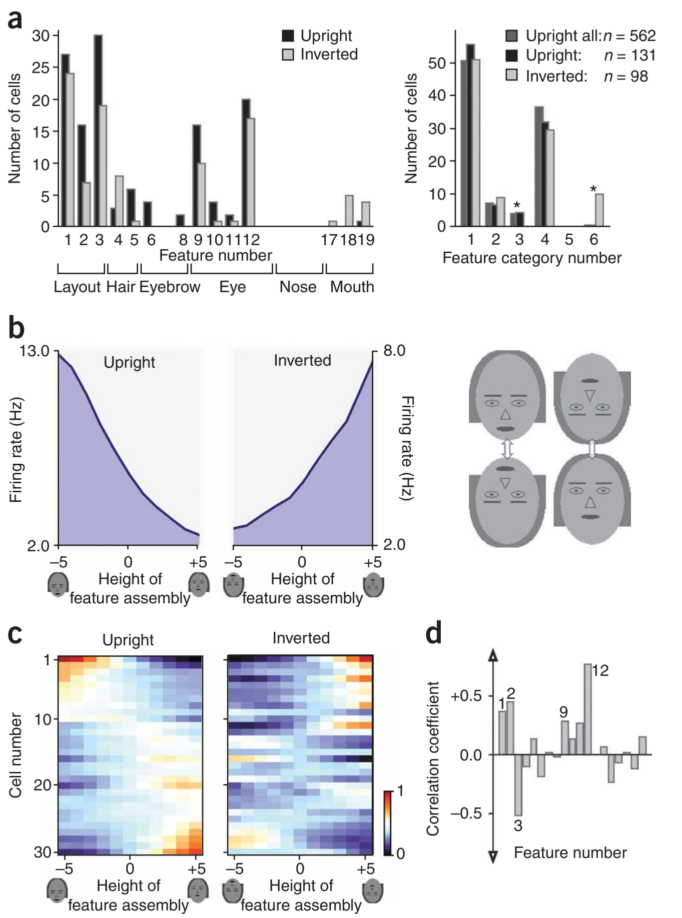

Figure 7 Face inversion and feature tuning. (**a**) Distributions of number of significantly tuned dimensions per cell for upright (black, dark gray) and inverted (light gray) cartoon stimuli (data are presented as in **Fig. 3d**). (**b**) Tuning curve of an example neuron to height of feature assembly in upright (left) and inverted (middle) cartoons. Example cartoons are shown below the tuning curves. Enlarged on the right, the upper row shows effective cartoon stimuli and the lower row shows ineffective cartoon stimuli. Effective cartoons had the feature assembly below the face center and ineffective ones had it above the face center. (**c**) Tuning curves of all 30 cells with significant tuning to the parameter height of feature assembly in upright faces. Data are presented as in **Figure 4a,b**. Tuning curves from stimulus trains of upright cartoons are shown on the left and those using inverted cartoons are shown on the right. Sorting of cells in both plots is identical and according to feature value in upright cartoon experiment eliciting the maximal response in a cell. (**d**) Correlations between tuning curve shapes to upright and inverted cartoon stimuli for all 19 feature dimensions. The most popular parameters are indicated by number: face aspect ratio (1), face direction (2), height of feature assembly (3), inter-eye distance (9) and iris size (12).

Separable joint tuning allows for faithful readout of individual feature values and confirms the importance of feature representation in the middle face patch.

Mechanisms for holistic processing: facial context

Face recognition has been characterized as holistic in that the face is processed as whole unit without breakdown into parts or part relations[12,13]. What role, if any, does the facial whole have in the representation of face identity in the middle face patch? We addressed this question in two experiments that are similar to two classical psychophysical tests of holistic processing. In the first, we tested how coding of individual facial features depended on the presence or absence of the rest of the face. This experiment was motivated by the finding that humans recognize parts of faces best when they are embedded in a whole face[13]. In the second experiment, we tested the effect of face inversion on tuning. This was motivated by the fact that humans recognize a facial feature embedded in an upright face better than the same feature embedded in an inverted face[14].

For the first experiment, we measured cartoon tuning in 49 neurons under three different contexts. In context 1 (the original tuning experiment), all 19 dimensions were simultaneously varied and a single dimension showing significant tuning was identified. In context 2, this one feature dimension was then varied while all others were kept constant (at their mean value). In context 3, the same feature dimension was varied, but all nonvarying face parts were removed (tuning of five example cells measured under these three contexts is shown in **Fig. 6a**; **Supplementary Videos 2–4**). The shape of the tuning curves determined in contexts 1 and 2 were similar ($r = 0.93$, $P << 0.001$). In the absence of the rest of the face (context 3), 14 cells (29%) lost any significant tuning. The tuning curves of the remaining 35 cells that kept their tuning were similar in shape across all three contexts (mean $r = 0.93$, 0.66 and 0.66 between contexts 1 and 2, 2 and 3, and 1 and 3, respectively, $P << 0.001$; **Fig. 6b**,**c**). The gain of tuning curves, however, was bigger when the whole face was present (average gain ratio of 2.2 for context 2 versus 3; **Fig. 6c**). In addition, firing rates were higher when the whole face was shown (14.7 Hz in context 2 versus 9.3 Hz in context 3). These results indicate that the overall gestalt of a face exerts an influence on tuning to individual feature dimensions by gain modulation; when the whole face is present, gain increases and tuning becomes more robust, whereas the shape of the tuning curve remains constant. A further prediction of a gain modulation model is the separability of joint tuning functions (for which we present evidence above).

Mechanisms for holistic processing: face inversion

What brings about this gain modulation? Is it the particular gestalt of a face or simply the presence of other features in close proximity? Face inversion offers a unique opportunity to address this question, as upright and inverted stimuli are similar at the feature level, but substantially different in their overall gestalt[14–16]. Our cartoon stimuli are especially well suited for studying inversion effects because face inversion does not physically alter facial layout, eye- and eyebrow-related features, but merely flips the order of feature values. For example, raised eyebrows turn into lowered eyebrows and vice versa. Nose, hair and mouth, on the other hand, are physically changed by inversion.

We tested 48 neurons to rapid serial visual presentation stimulus sequences of upright and inverted cartoon stimuli. We found a 25% reduction in the incidence of tuning with inversion (131 significantly modulated tuning curves for upright faces and 98 for inverted faces), which affected all important feature dimensions (**Fig. 7a**). However, there were two notable exceptions to the general reduction; tuning to eyebrow parameters was not just reduced, but was lost entirely with inversion; and substantial tuning to mouth related parameters emerged *de novo* (both significant at $P = 0.01$ using bootstrapping controls). Because mouth and eyebrow features remained physically identical on inversion, it must have been the change of their placement inside the face with inversion that caused loss of tuning to the former and emergence of tuning to the latter. If middle face patch cells match the incoming stimulus against an upright face template, this would parsimoniously explain the emergence of tuning to mouths (as they appear at locations expected for eye- and eyebrow-related features), the loss of tuning to eyebrows (resulting from mismatch between actual and expected position) and the general decrease in tuning to the other features (resulting from partial mismatch with expected position) on inversion. The template hypothesis makes a further prediction for how inversion should change the shape of tuning:

When inversion merely flips the order of feature values, the tuning curve to that feature should flip as well. We examined a neuron tuned to height of feature assembly (**Fig. 7b**). This neuron responded maximally to the feature assembly located near the chin inside an upright face and to the feature assembly located near the forehead inside an inverted face. In both cases, this neuron preferred the feature assembly toward the bottom (in gravitational terms) of the facial outline. In the population, we found the predicted negative correlation between the shape of tuning to height of feature assembly for upright and inverted faces ($r = -0.50$; **Fig. 7c**,**d**), On the other hand, when inversion leaves both the ordering and physical appearance of a feature unchanged, tuning curves to the feature should stay unchanged on inversion. Indeed, we found robust positive correlations between tuning curve shapes for upright and inverted faces to the popular features of face aspect ratio, face direction, inter-eye distance and iris size (**Fig. 7d**).

DISCUSSION

In this study, we took advantage of the rich, but simple, face code supplied by cartoon faces to probe strategies for face detection and differentiation in the middle face patch. By studying this problem in a small cortical volume, we identified new coding principles that may be of general importance to the extraction of complex form in inferotemporal cortex.

Cells in the middle face patch detect a wide range of faces, as evidenced by their vigorous responses to both real and cartoon faces compared with objects (**Fig. 1c**,**d**). However, different cells accomplish this by different means. No cell required the presence of a whole face to respond, indicating that the detection process is not strictly holistic. Instead, responses to systematically decomposed cartoon faces showed that different cells were selective for different face parts and interactions between parts (**Fig. 2a**,**b**), and even the same cell can respond maximally to different combinations of face parts (**Fig. 2d**). Thus, there is no single blueprint for detecting the form of a face in the middle face patch.

The mechanism for distinguishing between individual faces appears to rely on a division of labor among cells tuned to different subsets of facial features. This was revealed by dense parametric mapping[17]; responses were measured to a cartoon stimulus in which all of the face parameters were independently varied. Tuning to individual features was almost universal (found in 90% of cells) and each cell was tuned, on average, to only three feature dimensions (**Fig. 3c**), which were integrated in a separable manner (**Fig. 5a**). Together, cells in the middle face patch span a face space[18,19] with three salient characteristics. First, the axes of the space, represented by the tuning curves of individual cells, correspond to basic face features and not to holistic exemplars. Second, the dimensionality of the space is reduced compared with the physical face space[20] (**Fig. 3d**) and the population focused on features related to the eye and face layout geometry. Finally, the location in the space is coded predominantly by the firing rates of cells with broad, monotonic tuning curves (**Fig. 4d**). A majority of tuning curves peaked at one extreme and showed a minimal response at the opposite extreme, and the dynamic range of tuning often spanned or slightly exceeded the range of physical plausibility. Monotonic tuning allows for simple readout[21] and may be a general principle for high-level coding of visual shapes[22,23]. It may also aid in emphasizing what makes an individual face unique (that is, separates it from the standard face)[7,24–26], as population response variance is highest to such ‘unusual’ features (**Fig. 4c**). Finally, the breadth of tuning underscores the fact that cells in the middle face patch encode axes and not individual faces. This finding indicates that coding in the middle face patch is coarse, and is not sparse as it is at higher stages of the processing hierarchy[27], and substantiates theoretical proposals that coarse population codes are advantageous for representing high-dimensional stimulus spaces[28,29].

Psychophysicists have long proposed that perception of face identity has a holistic component[12,13,30], in which a face is obligatorily processed as a whole. We found two lines of evidence for holistic processing of face geometry. First, we found that the presence of a whole, upright face increased the gain of feature tuning curves by an average factor of 2.2 (**Fig. 6c**). Our finding that holistic coding uses gain modulation underscores the idea that gain modulation may be a computational mechanism of general importance to cortical function[31,32], even beyond coordinate transformations[33,34] and for attention[35,36]. Second, by comparing responses to upright and inverted cartoon faces, we found that the identity of individual features is interpreted according to the heuristics of an upright face template (**Fig. 7**). These two results demonstrate specific neural mechanisms by which the presence of an upright facial gestalt influences feature measurement in single cells. In our experiments, cartoons were presented rapidly, putting the system to a test in a feedforward mode[37–39] in the sense that no expectations about the upcoming feature values could be formed. In real-world face perception, top-down feedback[40] is important and may be necessary for additional effects of holistic processing.

We found a high incidence of tuning to some facial features, mostly to eyes and facial layout, and a paucity of tuning to others, mostly mouth and nose-related ones. It seems plausible that such a spatial bias of tuning preference in the face may be the result of attention or preferential looking rather than a computational strategy for face processing, as attention has been shown to augment feature tuning[35,36,41]. Several results are, however, incompatible with a spatial attention or preferential looking account of tuning biases. First, these accounts would predict stronger tuning to isolated face parts as a result of the absence of other potentially distracting visual stimuli (the rest of the face). However, we found the opposite (**Fig. 6a**). Second, preference for face aspect ratio over both internal features (nose, mouth, eyes and eye brows) and external ones (hair) can only be explained by a donut-shaped spotlight of attention and this would not cause preferential tuning for face direction or height of feature assembly, two popular parameters defined by the relative positioning of the internal features to the face layout. Third, the feature tuning bias occurred independently of slight gaze-direction biases above or below the fixation spot (**Supplementary Text 1**); for example, the preference for eye parameters remained even during fixations below the fixation spot. This rules out the possibility of a preferential looking account and renders a spatial attention confound unlikely, as spatial attention is tied to eye movements. Thus, preferential tuning for facial layout and eye parameters seems to be influenced little, if at all, by attention and eye positioning, but instead seems to be the result of computational mechanisms of shape analysis in the middle face patch.

For the same reasons, it seems unlikely that preferential representation of extreme feature values is a byproduct of attentional capture. Furthermore, the attentional capture account would predict response maxima for both extremes, that is, U-shaped tuning curves, because both ends of the shape spectrum are equally extreme shapes in most feature dimensions. Instead, we found that tuning curves were ramp shaped and even more response minima than maxima occurred at extremes. Similarly, the special status of extreme feature values cannot be explained by shape changes (whether attention capturing or not) rather than genuine shape preferences, as significant interactions between responses to successive feature values were found in less than 2% of all tuned feature dimensions ($P < 0.05$).

Our results expand existing conceptions about inferotemporal organization[42–44] in two major ways. First, it has been suggested that an IT cell can be characterized by its 'critical feature', defined as the simplest stimulus that still elicits a maximal response[42]. Our results suggest that such a characterization is incomplete and needs to be augmented by a description of the cell's feature tuning and its full selectivity for parts and part interactions. Cells in the middle face patch are not only selective for the presence of subsets of face parts (**Fig. 2**), but also show tuning to subsets of face features (**Fig. 3**). The critical feature for a cell would be a face optimized along all dimensions to which the cell is tuned. However, knowing this single best image would not allow one to distinguish between features to which the cell is tuned, and parts that are simply required to be present (in whatever shape). Furthermore, the predominance of broad, ramp-shaped tuning suggests that all levels of response to a tuned feature, including minimal responses, are important (minimal responses are just as informative about what feature is present as maximal responses, see ref. 8 for a related idea). This notion, that all levels of response to a tuned feature are informative, is not included in the critical feature account of IT. On average, the response to a full face was less than the sum of the responses to each part and cells often fired maximally to different combinations of face parts (**Fig. 2d**). Therefore, an IT cell, at least in the middle face patch, is only incompletely characterized by a single critical feature; instead, it is necessary to describe all of the parts and part combinations for which the cell is selective.

The second major insight from our findings concerns the functional organization of IT. It has been suggested that cells selective for visually similar critical features are grouped into columns. Our results indicate that what cells in the middle face patch have in common is a strong preference for faces over other objects, but this preference is a true form selectivity that cannot be captured by common selectivity to any fixed visual feature. There was a marked diversity in part selectivity and feature tuning in the middle face patch, and the tuned features of two neighboring face cells often shared no visual similarity at all (for example, hair width versus eyebrow slant). This diversity of feature tuning provides the brain with a rich vocabulary to describe faces and shows how a high-dimensional parameter space may be encoded even in a small region of IT. The macaque temporal lobe contains three face patches anterior to the middle face patch, and future experiments may reveal how the vocabulary of the middle face patch is used by the anterior face patches.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We are grateful to the late D. Freeman and N. Nallasamy for stimulus programming; R. Tootell, W. Vanduffel and members of the Massachusetts General Hospital monkey fMRI group for assistance with scanning; A. Dale and A. van der Kouwe for allowing us to use their multi-echo sequence and undistortion algorithm; T. Chuprina and N. Schweers for animal care and Guerbet for providing Sinerem contrast agent. This work was sponsored by US National Institutes of Health grant EY16187 and German Science Foundation grant FR 1437/3-1. D.Y.T. was supported by a Sofja Kovalevskaja Award from the Alexander von Humboldt Foundation.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://www.nature.com/reprintsandpermissions/.

1. Hubel, D.H. & Wiesel, T.N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol. (Lond.)* **148**, 574–591 (1959).
2. Kanwisher, N., McDermott, J. & Chun, M.M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).
3. Tsao, D.Y., Freiwald, W.A., Knutsen, T.A., Mandeville, J.B. & Tootell, R.B.H. Faces and objects in macaque cerebral cortex. *Nat. Neurosci.* **6**, 989–995 (2003).
4. Tsao, D.Y., Freiwald, W.A., Tootell, R.B.H. & Livingstone, M.S. A cortical region consisting entirely of face-selective cells. *Science* **311**, 670–674 (2006).
5. Brunswik, E. & Reiter, L. Eindruckscharaktere schematisierter Gesichter. *Zeitschrift Fuer Psychologie* **142**, 67–135 (1937).
6. Biederman, I. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* **94**, 115–147 (1987).
7. Rhodes, G., Brennan, S. & Carey, S. Identification and ratings of caricatures: implications for mental representations of faces. *Cognit. Psychol.* **19**, 473–497 (1987).
8. Benson, P.J. & Perrett, D.I. Perception and recognition of photographic quality facial carricatures: implications for the recognition of natural images. *Eur. J. Cogn. Psychol.* **3**, 105–135 (1991).
9. Leopold, D.A., Bondar, I.V. & Giese, M.A. Norm-based face encoding by single neurons in monkey inferotemporal cortex. *Nature* **442**, 572–575 (2006).
10. Mel, B. & Fiser, J. Minimizing binding errors using learned conjunctive features. *Neural Comput.* **12**, 247–278 (2000).
11. Grunewald, A. & Skoumbourdis, E.K. The integration of multiple stimulus features by V1 neurons. *J. Neurosci.* **24**, 9185–9194 (2004).
12. Young, A.W., Hellawell, D. & Hay, D.C. Configural information in face perception. *Perception* **16**, 747–759 (1987).
13. Tanaka, J.W. & Farah, M.J. Parts and wholes in face recognition. *Q. J. Exp. Psychol. A.* **46**, 225–245 (1993).
14. Thompson, P. Margaret Thatcher: a new illusion. *Perception* **9**, 483–484 (1980).
15. Yin, R.K. Looking at upside-down faces. *J. Exp. Psychol.* **81**, 141–145 (1969).
16. Bartlett, J.C. & Searcy, J. Inversion and configuration of faces. *Cogn. Psychol.* **25**, 281–316 (1993).
17. Brincat, S.L. & Connor, C.E. Underlying principles of visual shape selectivity in posterior inferior temporal cortex. *Nat. Neurosci.* **7**, 880–886 (2004).
18. Valentine, T. A unified account of the effects of distinctiveness, inversion and race in face recognition. *Q. J. Exp. Psychol. A.* **43**, 161–204 (1991).
19. McKone, E., Aitkin, A. & Edwards, M. Categorical and coordinate relations in faces, or Fechner's law and face space instead? *J. Exp. Psychol. Hum. Percept. Perform.* **31**, 1181–1198 (2005).
20. Fraser, I.H., Craig, G.L. & Parker, D.M. Reaction time measures of feature saliency in schematic faces. *Perception* **19**, 661–673 (1990).
21. Guigon, E. Computing with populations of monotonically tuned neurons. *Neural Comput.* **15**, 2115–2127 (2003).
22. Kayaert, G., Biederman, I., Op de Beeck, H. & Vogels, R. Tuning for shape dimensions in macaque inferior temporal cortex. *Eur. J. Neurosci.* **22**, 212–224 (2005).
23. De Baene, W., Premereur, E. & Vogels, R. Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. *J. Neurophysiol.* **97**, 2900–2916 (2007).
24. Rhodes, G. *Superportraits: Caricatures and Recognition* (Psychology Press Publishers, East Sussex, UK, 1996).
25. Webster, M.A. & MacLin, O. Figural aftereffects in the perception of faces. *Psychon. Bull. Rev.* **6**, 647–653 (1999).
26. Leopold, D.A., O'Toole, A.J.O., Vetter, T. & Blanz, V. Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat. Neurosci.* **4**, 89–94 (2001).
27. Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C. & Fried, I. Invariant visual representation by single neurons in the human brain. *Nature* **435**, 1102–1107 (2005).
28. Zhang, K. & Sejnowski, T.J. Neuronal tuning: to sharpen or broaden? *Neural Comput.* **11**, 75–84 (1999).
29. Eurich, C.W. & Wilke, S.D. Multidimensional encoding strategy of spiking neurons. *Neural Comput.* **12**, 1519–1529 (2000).
30. Galton, F. Composite portraits, made by combining those of many different persons into a single, resultant figure. *J. Anthropol. Inst.* **8**, 132–144 (1879).
31. Salinas, E. & Thier, P. Gain modulation: a major computational principle of the central nervous system. *Neuron* **27**, 15–21 (2000).
32. Liu, Y. & Jagadeesh, B. Neural selectivity in anterior inferotemporal cortex for morphed photographic images during behavioral classification or fixation. *J. Neurophysiol.* **100**, 966–982 (2008).

33. Andersen, R.A., Bracewell, R.M., Barash, S., Gnadt, J.W. & Fogassi, L. Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of macaque. *J. Neurosci.* **10**, 1176–1196 (1990).
34. Zipser, D. & Andersen, R.A. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature* **331**, 679–684 (1988).
35. Treue, S. & Martínez Trujillo, J.C. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* **399**, 575–579 (1999).
36. Koida, K. & Komatsu, H. Effects of task demands on the responses of color-selective neurons in the inferior temporal cortex. *Nat. Neurosci.* **10**, 108–116 (2007).
37. Thorpe, S., Fize, D. & Marlot, C. Speed of processing in the human visual system. *Nature* **381**, 520–522 (1996).
38. Keysers, C., Xiao, D.-K., Földiák, P. & Perrett, D.I. The speed of sight. *J. Cogn. Neurosci.* **13**, 90–101 (2001).
39. Fabre-Thorpe, M., Richard, G. & Thorpe, S.J. Rapid categorization of natural images by rhesus monkeys. *Neuroreport* **9**, 303–308 (1998).
40. Cox, D., Meyers, E. & Sinha, P. Contextually evoked object-specific responses in human visual cortex. *Science* **304**, 115–117 (2004).
41. McAdams, C.J. & Maunsell, J.H.R. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* **19**, 431–441 (1999).
42. Tanaka, K., Saito, H.-A., Fukada, Y. & Moriya, M. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.* **66**, 170–189 (1991).
43. Fujita, I., Tanaka, K., Ito, M. & Cheng, K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* **360**, 343–346 (1992).
44. Wang, G., Tanaka, K. & Tanifuji, M. Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* **272**, 1665–1668 (1996).
45. Bruce, C., Desimone, R. & Gross, C.G. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* **46**, 369–384 (1981).

ONLINE METHODS

All procedures conformed to local and US National Institutes of Health guidelines, including the US National Institutes of Health Guide for Care and Use of Laboratory Animals as well as regulations for the welfare of experimental animals issued by the German federal government.

Animals. Three male rhesus macaques were implanted with ultem headposts, trained via standard operant conditioning techniques to maintain fixation on a small spot for a juice reward and then scanned in a 3T Allegra (Siemens) horizontal bore magnet to identify face-selective regions using MION/Sinerem contrast agent (further details are provided in refs. 3,4). In all monkeys, a prominent face-selective region was located ~6-mm anterior to the inter-aural line. This middle face patch was targeted for recordings (details in ref. 4). Monkey A had a middle face patch located on the lip of the superior temporal sulcus, monkey T in the fundus and monkey L had two middle face patches, one on the lip (which we targeted) and one in the fundus.

Single-unit recording and eye-position monitoring. We recorded extracellularly with electropolished tungsten electrodes coated with vinyl lacquer (FHC). Extracellular signals were amplified, bandpass filtered (500 Hz to 2 kHz) and fed into a dual-window discriminator and an audio monitor (Grass). Spike trains were recorded at 1-ms resolution. Only well-isolated single units were studied. Cells that were visually responsive to the screening stimulus set (**Fig. 1b**) by ear were further tested with cartoon stimuli; in addition, some cells that were unresponsive to the screening stimuli by a formal criterion (see below) were also tested. Eye position was monitored with an infrared eye tracking system (ISCAN) at 60 Hz with an angular resolution of 0.25°, calibrated before and after each recording session by having the monkey fixate dots at the center and four corners of the monitor.

Visual stimuli. The monkey sat in a dark box with its head rigidly fixed and was given a juice reward for keeping fixation for 3 s in a 2.5° fixation box. Visual stimuli were presented using custom software (written in Microsoft Visual C/C++) and presented at a 60-Hz monitor refresh rate and 640 × 480 resolution on a BARCO ICD321 PLUS monitor. The monitor was positioned 53 cm in front of the monkey's eyes. Pictures subtended a 7° × 7° region of the visual field and cartoons subtended a 5.4° × 7.6° region on average, with both being presented at the center of the screen.

Pictures were presented for 200 ms, separated by 200-ms blank intervals in three experiments. In the first, 96 pictures from six different image categories (faces, human bodies, produce, technical objects, human hands and scrambled images) were shown (**Supplementary Fig. 1**). In the second, images of 16 real faces, 16 fitted cartoons, 16 technical objects and 16 cartoon face parts were shown (**Supplementary Fig. 1**). In the third, all 128 (2^7) decompositions of a cartoon stimulus were shown (**Supplementary Fig. 1**).

In contrast, cartoon stimuli were shown continuously, updated every seven frames (117 ms). Cartoon faces were defined by 19 parameters, each of which could take any of 11 values. The face defined by mean parameters ($p_1 = p_2 = \ldots = p_{19} = 0$) was specified by measurements taken from a photograph of Tom Cruise. Face aspect ratio defined the eccentricity of a solid ellipse constituting the face outline. Face direction defined the horizontal offset of the feature assembly (that is, eyes, eyebrows, nose and mouth) as a fraction of face width. Thus, the horizontal position of the feature assembly could range from the left edge of the face to the right. Height of feature assembly defined the vertical offset of the feature assembly as a fraction of face height. Hair was modeled as an inverted U of height, hair length and thickness, and hair width. Inter-eye distance was defined as the distance between iris centers, normalized by face width (ranging from almost cyclopean to abutting the edges of the face). Eye aspect ratio defined the aspect ratio and eye size the size of the ellipse surrounding the iris. Eye, iris and eyebrow were drawn only when the left (right) edge of the eye was to the right (left) of the left (right) edge of the face. Gaze direction defined 3 × 3 pupil positions in the eye as follows:

$$\begin{pmatrix} -4 & -3 & -2 \\ -1 & 0 & 1 \\ 2 & 3 & 4 \end{pmatrix},$$

where matrix position denotes iris position in the eye and matrix value denotes feature value. Parameter values 0, −5 and +5 all represented a straight gaze direction. Horizontal and vertical spacing between positions was fixed at 2 pixels. Iris size defined the size of a solid ellipse in the eye (with the same aspect ratio as the eye) as a fraction of eye size. The eyebrow was modeled as an angled line segment, with the angle defined by eyebrow slant, width defined by eyebrow width and height above eyes defined by eyebrow height, with the latter two being normalized by face width and face height, respectively. The nose was modeled as an outline of an isosceles triangle, with base width defined by nose base and altitude by nose altitude. The mouth was modeled as two half ellipses. In a smiling mouth, one half ellipse was black and the other was a gray mask, which served to carve out the curve of the upper lip; in a neutral/frowning mouth, both half ellipses were black and joined to form a convex mouth shape. The width of the mouth, expression of the mouth (smiling to frowning) and height of the mouth (open to closed) were defined by mouth size, mouth top and mouth bottom, respectively. The distance of the mouth below the nose was defined by the mouth-nose distance (this parameter only affected the vertical placement of the mouth; nose position was unaffected).

Picture data analysis. For each cell, we analyzed the poststimulus time histograms over 400 ms for all images shown (96 in first experiment, 64 in second and 128 in the third). Poststimulus time histograms were smoothed with a Gaussian kernel in time with $\sigma_t = 15$ ms. For experiments 1 and 2, we collapsed responses in each category to compute the cross-category response variance for each time bin. This variance had to be threefold higher than that of spontaneous activity (measured between σ_t and 80 ms after stimulus onset) for a cell to be classed as being visually responsive. The visual response period was then defined from the first to last point in time that exceeded the variance threshold. For cells that did not meet this strict criterion for visual responsiveness, a default response period was defined to last from 120 ms to 319 ms. Firing rates were computed as averages over this interval. In the case of the first experiment, the response magnitudes were determined for faces and objects relative to the baseline firing rate and normalized to the maximal response. A face selectivity index was then computed as the ratio between difference and sum of face- and object-related responses. For $|\text{face-selectivity index}| > 1/3$, that is, if the response to faces was at least twice (or at most half) that of nonface objects, a cell was classed as being face selective[45–47].

Face decomposition analysis. Because the 128 images in this experiment did not fall into distinct categories, the method for finding the response period deviated slightly from the procedure described above. The poststimulus response interval started when the firing rate exceeded a threshold equal to spontaneous activity plus two s.d. of spontaneous activity. The response interval ended when the response fell below this threshold value. The resulting 128-element response vector was subjected to a seven-way ANOVA with the presence/absence of each of the seven face parts (**Fig. 2a**) as factors.

Cartoon data analysis. All data analysis was performed using custom programs written in MATLAB (MathWorks).

Determining significance of tuning. For each cell and feature dimension, we computed time-resolved poststimulus tuning profiles (**Supplementary Fig. 2**) over three feature update cycles (351 ms of duration at 1-ms resolution) and 11 feature values. Profiles were subsequently smoothed with a two-dimensional Gaussian kernel of width $\sigma_t = 15$ ms in time and $\sigma_f = 1$ in the feature domain. We searched each profile for feature tuning, that is, increased diversity of response magnitudes, at each time delay. To minimize biases for tuning shape, we computed an entropy-related measure termed hetereogeneity[48]. Heterogeneity is derived from the Shannon-Weaver diversity index $H' = -\sum_{i=1}^{k} p_i \log(p_i)$, with k being the number of bins in the distribution (11 in our case) and p_i being the relative number of entries in each bin. Homogeneity is defined as the ratio of H' and $H_{\max} = \log(k)$; heterogeneity is defined as 1 – homogeneity. Thus, if all p_i values are identical, heterogeneity is 0, and if all values are zero except for one, heterogeneity is 1.

For each dimension and delay, we compared the heterogeneity value against a distribution of 5,016 surrogate heterogeneity values obtained from shift predictors. Shift predictors were generated by shifting the spike train relative to the stimulus sequence in multiples of the stimulus duration. This procedure preserved firing rate modulations by feature updates, but destroyed any systematic relationship between feature values and spiking. From the surrogate heterogeneity distributions, we determined significance using Efron's[49,50] percentile method; for an actual heterogeneity value to be considered significant, we required it to exceed 99.9% (5,011) of the surrogate values. Note that this method is exact only for the actual 5,016 surrogate values and that a different set of values may have generated a different threshold. Therefore, to get a more robust and even more stringent significance level, we took as our significance threshold the average of the fifth largest heterogeneity value and the average of the five largest heterogeneity values. We validated this method with simulations, larger surrogate datasets of selected cells and by estimating significance levels from gamma functions fitted to the surrogate distributions (**Supplementary Fig. 9**). In the vast majority of cases reported here, the heterogeneity value of a significant tuning curve was much higher than even the largest of the surrogates. For a dimension to be considered significantly tuned, the significance threshold had to be passed at least twice at a temporal separation of at least $2\sigma_t$. We further required the tuning curve's maximal value to be at least 25% larger than the minimal value (see **Supplementary Text 5** and **Supplementary Figs. 10–12** for a different method, Gaussian fitting, for finding significantly tuned dimensions).

Co-occurrence of significant tuning. We found 14 out of 19 feature dimensions to be represented by middle face patch neurons. We then asked for each of the $C_2^{14} = 91$ feature combinations whether the frequency of occurrence was larger than would be expected by a model of chance associations. This model took into account the number of features each cell was tuned to (**Fig. 3c**) and the number of cells tuned to each feature dimension (**Fig. 3d**). We generated surrogate data that exactly matched these two distributions, but in which the associations between cell and features was otherwise random. Generating a distribution of 5,000 such surrogates for each feature combination, we tested for significance at $P = 0.0055$, a significance level at which only half a false positive dimension is expected on average.

Joint tuning functions were computed for a temporal delay suitable for both feature dimensions considered. For the analysis of interactions between significantly tuned dimensions, we first computed the center of mass of the heterogeneity measures (functions of time) of all significantly modulated dimensions to derive a 'joint delay'. When the optimal delays of both tuning curves considered were either shorter or longer than this joint delay, the center of mass between the heterogeneity functions of these two dimensions was chosen instead (**Supplementary Text 4** contains additional analysis of joint tuning functions).

Normalization conventions. For each cell, responses were baseline subtracted and divided by the maximal response above baseline; normalized responses were then averaged across cells. All tuning curves were normalized to the same area (**Figs. 4**, **6b** and **7c**, and **Supplementary Fig. 8**). The maximal response in the whole population of tuning curves was then set to 1 and the minimum was set to 0. In the inset in **Figure 4d**, deviating from this convention, the maximal response of each of the seven average tuning curves was set to 1 and the minimal response was set to 0.

46. Baylis, G.C., Rolls, E.T. & Leonard, C.M. Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Res.* **342**, 91–102 (1985).
47. Perrett, D.I. *et al.* Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc. R. Soc. Lond. B* **223**, 293–317 (1985).
48. Zar, J.H. *Biostatistical Analysis* (Prentice Hall, Upper Saddle River, New Jersey, 1998).
49. Efron, B. Bootstrap methods: another look at the jackknife. *Ann. Stat.* **7**, 1–26 (1979).
50. Manly, B.F.J. *Randomization, Bootstrap and Monte Carlo Methods in Biology* (CRC Press, Boca Raton, Florida, 2007).

doi:10.1038/nn.2363