nature neuroscience

Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain

Pia Rotshtein[1], Richard N A Henson[1,2], Alessandro Treves[3], Jon Driver[1,2] & Raymond J Dolan[1]

How the brain represents different aspects of faces remains controversial. Here we presented subjects with stimuli drawn from morph continua between pairs of famous faces. In the paired presentations, a second face could be identical to the first, could share perceived identity but differ physically (30% along the morph continuum), or could differ physically by the same distance along the continuum (30%) but in the other direction. We show that, behaviorally, subjects are more likely to classify face pairs in the third paired presentation as different and that this effect is more pronounced for subjects who are more familiar with the faces. In functional magnetic resonance imaging (fMRI), inferior occipital gyrus (IOG) shows sensitivity to physical rather than to identity changes, whereas right fusiform gyrus (FFG) shows sensitivity to identity rather than to physical changes. Bilateral anterior temporal regions show sensitivity to identity change that varies with the subjects' pre-experimental familiarity with the faces. These findings provide neurobiological support for a hierarchical model of face perception.

Recognizing the face of a particular individual, such as Marilyn Monroe, can be achieved across different lighting, views, hairstyles, expressions and ages. How we achieve such robust recognition is not fully understood, despite much research and progress[1,2]. An emerging consensus is that occipito-temporal regions are particularly important for face processing[3–5], although other areas may contribute[6–9]. Neurons in occipito-temporal cortex that are responsive to faces have been observed by single-cell recording in primates[10] and intracranial electrophysiological recording in humans[11]. In addition, neuropsychological deficits in face processing (prosopagnosia) are associated with lesions to right occipito-temporal cortex[12,13]. Functional neuroimaging studies have identified regions in occipito-temporal cortex that show greater responses to faces than to other classes of visual stimuli in FFG, IOG and superior temporal sulcus[3,14,15].

An unresolved issue concerns the nature of representations of faces in such areas, for which many possibilities have been considered. In one account, faces are represented in terms of identity categories[16] that are acquired by experience. Other accounts suggest that representations are more continuous in terms of physical properties in 'face space'[17,18]. A behavioral approach for investigating whether representations are categorical or continuous involves judgments of stimuli from a physical continuum, such as color perception as a function of wavelength[19]. For faces, artificial continua can be generated by morphing between pairs of faces[16,20] (**Fig. 1a** and **Supplementary Fig. 1** online). This approach suggests that perception of face identity may be 'categorical,' involving abrupt transitions in perception of identity across a morph continuum[16]. This result seems to depend on prior experience with face identities and is not observed for a morph continuum between unfamiliar faces[16].

Here we have used morphs between famous faces in behavioral and fMRI studies. In fMRI, we used a variant on the 'paired-repetition' model[21] that has been shown to produce decreases in the blood oxygenation level–dependent (BOLD) signal in brain areas that are sensitive to repeated stimulus properties[15,21–23]. For 43 morph continua, three types of face-pair trial were generated. The second face could match the first in physical properties and identity ('identical'), could differ in physical properties without crossing an identity boundary ('within'), or could differ in both respects ('between')[20]. Notably, the within and between conditions always involved a 30% shift along the morph continuum (**Fig. 1b** and Methods), albeit in different directions that were counterbalanced.

We expected brain areas that are sensitive to physical facial properties to show BOLD repetition-related decreases for identical relative to within and between pairs (because only the latter two had physical differences), whereas areas that are sensitive to perceived identity should show greater repetition decreases for identical and within relative to between pairs (higher responses in the latter case owing to crossing a perceived identity boundary). Sensitivity to change for all three types of pair was confirmed in separate behavioral work. We also assessed each participant's self-rated level of pre-experimental familiarity with the famous faces used, given that 'categorical' perception of face identity is thought to depend on familiarity[16] and that this varies among individuals even for famous faces, such as those presented here.

Our fMRI findings show that there are different levels of face representation across distinct regions of the brain. We show that IOG is sensitive to physical changes between face pairs, regardless of whether these changes cross an identity boundary. Right FFG, by contrast, is primarily sensitive to changes that cross an identity boundary, even when the presence of physical change is controlled by comparing a 30%
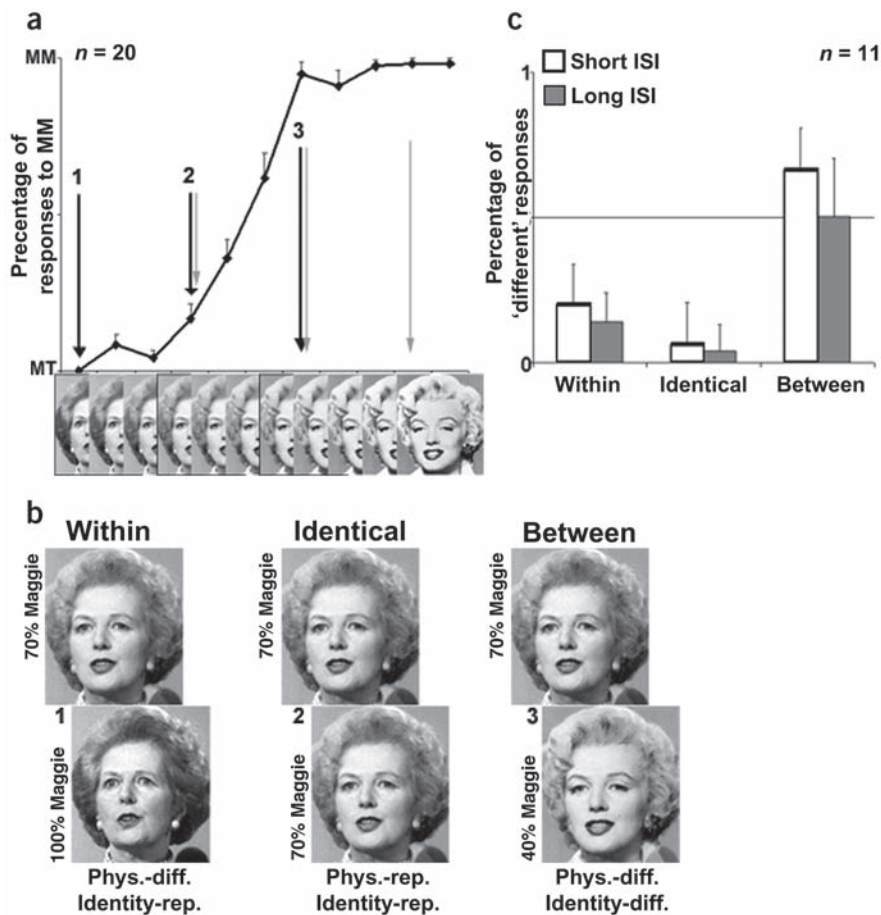
Figure 1 Example stimuli and behavioral results. (**a**) Results of the identity-classification experiment (two additional continua are shown in **Supplementary Fig. 1** online). The *x*-axis shows 11 levels from the morph continuum. The graph represents the averaged proportion of responses (across 20 subjects in the identity-classification task) that used the name shown at top of the *y*-axis. MM, Marilyn Monroe; MT, Margaret (Maggie) Thatcher. Black arrows and numerals 1–3 indicate the three morphing levels used as a stimulus set (the 'Maggie set') in the subsequent behavioral same-different and fMRI experiments (Methods). Gray arrows point to morphs used as a counterbalanced stimulus set (the 'Marilyn set') in those experiments, in which two of those morphs were reliably identified as MM, the other as MT. (**b**) The three types of face pair used to produce the three main conditions (within, identical and between) in the same-different and fMRI experiments. Note that the first face in a pair was equivalent in all three conditions. The second face was varied to create the different pair types. Phys, physical properties; rep, repeated; diff, different. Note also that a given morphing level (such as the fourth or seventh morph from the left along the *x*-axis in **a**) could appear in either the identical or the between condition, depending on which of the counterbalanced stimulus sets was used (Methods). (**c**) Results of the behavioral same-different experiment. The histogram shows the proportion of 'different' responses averaged across 11 subjects for the six different experimental conditions (Methods). Error bars in **a**,**c** represent the s.e.m.

shift within one side of the identity boundary with a 30% shift that crosses it. Finally, anterior and medial temporal regions show effects of changed identity that depend systematically on individual differences in self-rated pre-experimental familiarity with the face set.

RESULTS

Categorical perception of identity: behavioral results

In an initial 'familiarity rating' phase, 20 subjects rated 86 original (unmorphed) famous faces on a scale of 0 ('don't recognize') to 3 ('know very well'; Methods). This rating also exposed the subjects to the set of faces that was subsequently used in an 'identity-classification' phase.

For the identity-classification phase, subjects made two alternative forced-choice identity responses for each of 11 levels of morphing[16] (**Fig. 1a** and **Supplementary Fig. 1** online) presented in a random sequence for each of our 43 continua. Response profiles for each face pair and subject were fit by polynomial regression. In addition to a linear component, a highly significant (negative) cubic component ($t_{42} = -39.6$, $P < 0.001$) was observed (tested across continua), which captured a step-like (sigmoidal) function (**Fig. 1a** and **Supplementary Fig. 1** online). This nonlinear pattern seems to be consistent with categorical perception of identity[19].

A subsequent perceptual-discrimination experiment provided a more stringent criterion for categorical perception[19]. A further 11 subjects first provided familiarity ratings as above and then performed a 'same-different perceptual discrimination' task, in which the stimulus set comprised three morphs from each of 43 continua. These morphs were preselected on the basis of the location of the categorical bound-

ary estimated from the identity-classification experiment above. All three morphs differed by 30% along the physical morphing dimension (**Fig. 1a** and **Supplementary Fig. 1** online), with two on one side of the identity boundary and the third on the other side. Subjects made a same-different judgment on three types of sequential morph pair (corresponding to within, identical and between conditions; **Fig. 1b**). We also varied the interstimulus interval (ISI) between morphs in each pair (75 or 500 ms) in case a shorter ISI might emphasize differences (for example, because of any apparent motion between the pair). The design was a 3 (pair type) × 2 (ISI) factorial.

We found that judgments that stimulus pairs differed depended on pair type ($F_{1.4,13.9} = 80$, $P < 0.001$; **Fig. 1c**). Between pairs that straddled the categorical boundary were judged as different more often than were within pairs ($t_{10} = 7.9$, $P < 0.001$). An additional (smaller) effect of physical difference arose, because within pairs were judged as different more often than identical pairs ($t_{10} = 4.5$, $P < 0.005$). In the fMRI study, we therefore expected brain activity to be affected by perceived alterations in both identity change and physical change: in other words, the crucial issue was whether different brain regions were affected selectively by one or other of these manipulations. Note that similar behavioral patterns were observed at both ISIs (**Fig. 1c**), although an interaction was observed ($F_{1.7,17.3} = 7.3$, $P < 0.01$) with a tendency for between pairs to elicit even more 'different' judgments at a short ISI ($t_{10} = 4.7$, $P < 0.05$).

Self-rated pre-experimental familiarity with the face set correlated with performance in the behavioral same-different task, and mean familiarity per subject correlated positively ($r = 0.798$, $P < 0.005$) with the effect of between pairs versus within pairs on proportions of 'different' judgments.

Table 1 Physical and identity change effects

(a) Physical change effects: identical versus between plus within conditions

Anatomical description		No. of voxels[a]	Z score		MNI	
			$P < 0.001$	x	y	z
IOG	R	11	3.72	42	−69	−18
	L	12	3.68	−45	−78	−15
IFG	L	30	4.76	−39	30	−15
mOFC	L	6	3.76	−9	42	−21
SFG	L	7	3.67	−9	45	48

(b) Identity change effects: within plus identical versus between conditions

FFG	R	41	4.35	45	−57	−24
		6	3.26	33	−45	−27
LOS	R	8	3.39	33	−87	0
	L	5	3.35	−36	−90	−9
		8	3.28	−36	−78	−3
aSTS	R	16	4.25	48	−6	−21
IFG	R	10	3.97	48	36	3
	L	7	3.52	−33	36	−3
IFS	R	42	3.91	42	15	24

[a]Number of voxels, $3 \times 3 \times 3$ mm³. aSTS, anterior superior temporal sulcus; IFG, inferior frontal gyrus; IFS, inferior frontal sulcus; LOS, lateral occipital sulcus; mOFC, middle orbital frontal cortex; SFG, superior frontal gyrus; L, left; R, right hemisphere.

Subjects who were more familiar with the faces were thus more likely to judge morphs as 'different' when they crossed an identity boundary.

fMRI experiment

The stimuli and conditions used in the fMRI experiment were the same as those used in the behavioral same-different experiment, except that subjects no longer made explicit same-different judgments to avoid confounding repetition-decrease BOLD effects with response requirements[24]; instead, we used a target-monitoring task[23]. Subjects ($n = 14$) had to press a button on appearance of a prespecified rare (presented in ~15% of trials) 'target' face. Given findings of previous imaging studies on face processing[3–5], we focused on regions in occipito-temporal cortex, but for completeness **Tables 1** and **2** list all regions with an uncorrected value of $P < 0.001$. Because the ISI did not interact with the effects of interest in occipito-temporal cortex, all subsequent analyses collapsed across this factor. This makes it unlikely that the effects of physical change observed in occipito-temporal cortex (see below) reflect apparent motion between successive stimuli (or any other influences depending on the ISI).

We tested for sensitivity to 'physical change,' by comparing repetitions of the same physical stimulus with conditions with a physical change (that is, identical versus within plus between) in whole-brain analysis. Bilateral IOG showed less activation when physical properties were repeated (**Fig. 2a** and **Table 1a**). A similar outcome was observed for the simple comparison of identical versus between conditions (right, $t_{12} = 5.00$, $P < 0.001$; left, $t_{12} = 4.14$, $P < 0.005$), which controls for distance from the original unmorphed faces (hence from the 'ends' of the morph continua; Methods). Likewise, this region also showed less activity in a simple comparison of identical versus within conditions, both of which should fall on the same side of the identity boundary when physical aspects alone are varied (right, $t_{12} = 3.48$, $P < 0.005$; left, $t_{12} = 3.56$, $P < 0.005$).

For 8 of the 14 subjects, a separate 'functional localizer' was obtained to independently assess face-responsive regions in individuals. A blocked comparison of faces and non-faces (here, houses and scrambled faces; Methods), as described in many previous studies[3,25], was used to identify a right posterior 'occipital face area' (OFA)[5] and a right 'fusiform face

area' (FFA)[5]. The former could be reliably identified in seven of the eight subjects (Montreal Neurological Institute (MNI) brain coordinates: $x = 43 \pm 3.8$, $y = −61 \pm 4.8$, $z = −20 \pm 3.6$; $Z = 3.95 \pm 1.3$; mean ± s.d.). These individually defined OFAs showed less response for the identical than for the within condition ($t_6 = 2.99$, $P < 0.05$) and for the identical than for the between condition ($t_6 = 2.67$, $P < 0.05$), but no significant difference for the between versus the within condition (**Fig. 2**). Thus, voxel-based group analysis and a separate functional region of interest (fROI) analysis both suggest that activity in inferior occipital cortex relates primarily to physical change and not to perceived identity. A 30% change along the morph continuum was treated as similar here regardless of its direction.

We next tested for regions that showed sensitivity to the 'identity boundary,' initially by comparing conditions in which the pair of faces remained on the same side of the identity border against conditions in which they crossed it (between versus identical plus within). Right FFG showed less activation when identity was repeated (**Fig. 2b** and **Table 1b**). A simple comparison of the between versus the identical condition also showed this effect ($t_{12} = 4.28$, $P < 0.005$), while at the same time controlling for distance from the original unmorphed faces (and hence from the 'ends' of the morph continua; Methods). Similarly, a simple comparison of the between versus the within condition also showed this effect ($t_{12} = 3.48$, $P < 0.005$), while controlling for a 30% shift along the morph continuum.

The separate functional localizer scan (see above) identified the right FFA in seven of the eight subjects (MNI coordinates: $x = 45 \pm 6.7$, $y = −40 \pm 7.9$, $z = −24 \pm 5.4$; $Z = 3.4 \pm 1.2$). An fROI analysis showed greater activity for the between than for the within condition ($t_6 = 3.6$, $P < 0.05$) and for the between than for the identical condition ($t_6 = 2.75$, $P < 0.05$), but no difference for the identical versus the within condition (**Fig. 2c**). The effect of change or repetition in right FFG (including the independently and individually defined right FFA) thus depended on whether there was a shift across the identity boundary and not merely on physical change.
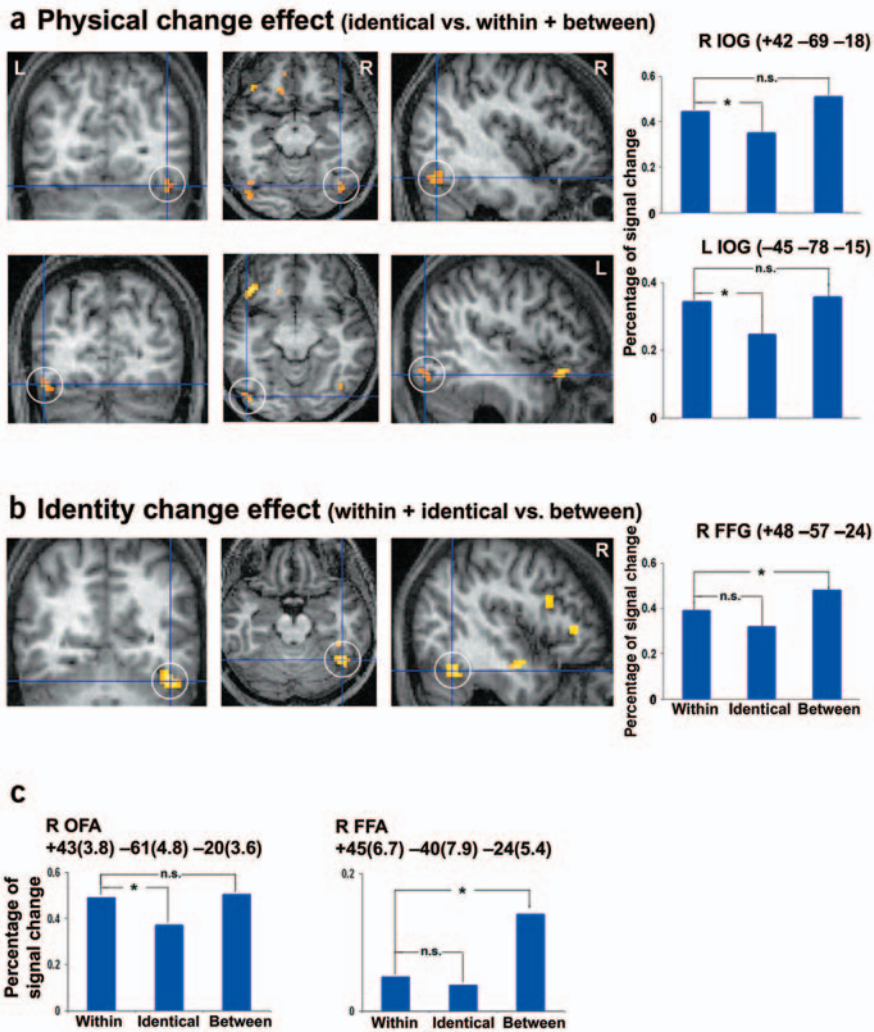
Taken together, these findings suggest that there is a functional dissociation between posterior (occipital) and more anterior (fusiform) face-responsive regions. To test this formally, we carried out an analysis of variance (ANOVA) with region (independently defined OFA and FFA) and condition (within and between) as factors. This confirmed an interaction ($F_{1,6} = 5.8$, $P = 0.05$), owing to the between condition differing from the within condition for FFA but not for OFA.

Finally, we used a whole-brain analysis to test whether intersubject differences in self-rated pre-experimental familiarity with the faces modulated sensitivity to identity change in any brain region (as assessed by the between minus within comparison, which controls for presence of physical change; it was this comparison that revealed a correlation with pre-experimental familiarity in the behavioral same-difference study).

Table 2 Effects related to pre-experimental familiarity

(a) Familiarity and identity change effects: between versus within conditions

Anatomical description		No. of voxels[a]	Z score		MNI	
			$P < 0.001$	x	y	z
aTP	L	26	5.27*	−45	0	−39
	R	1	3.09	63	−6	−33
aHipp	R	34	4.14	24	−12	−27
SFG	L	28	4.16	−6	18	69

(b) Familiarity and physical change effects: within versus identical conditions

Calcarine sulcus		46	4.47	18	−87	15
CG		8	3.77	12	18	48

[a]Number of voxels, $3 \times 3 \times 3$ mm³. CG, cingulate gyrus; SFG, superior frontal gyrus; L, left; R, right hemisphere. *$P < 0.01$, Family-wise error (FWE) corrected for whole brain.

a Physical change effect (identical vs. within + between)

b Identity change effect (within + identical vs. between)

c

Figure 2 Physical change and identity change in fMRI results. Left, statistical parametric maps showing physical change (**a**) and identity change (**b**) effects of interest across subjects (threshold $P < 0.001$, uncorrected), overlaid on a T1 image of one subject. Right, histograms showing the mean percentage signal change for the three conditions (collapsed across ISI) taken from the maxima voxels marked with blue crosshairs, for which the coordinates are given above each histogram. (**c**) Results of analysis on ROIs independently identified by a functional localizer ($n = 7$), showing the mean percentage signal change extracted from the functionally defined right OFA in the left histogram, and from the right FFA in the right histogram. Mean ($\pm$ s.d.) MNI coordinates are given for each of the two fROIs that were independently and functionally defined. Zero in these histograms is the averaged signal along the whole scan. A direct comparison of the between versus within effects in these two fROIs confirmed that different patterns were found for them. *$P < 0.05$; n.s., not significant.

along the morph continuum also has a small influence on behavioral judgments, because the within condition was judged to be 'different' slightly more often than was the identical condition. This finding suggests that face discrimination shows sensitivity to both physical properties and identity categories.

The fMRI experiment followed the logic that when hemodynamic responses are selectively affected by repeating or changing particular stimulus attributes, the nature of stimulus representations in the affected brain areas can be revealed[21–24]. We found less activity in IOG (and right OFA) when the physical properties of a face were repeated (identical) than when they were changed (within or between; **Fig. 2a,c**). This observation suggests that bilateral IOG and the right OFA are primarily sensitive to sensory attributes of faces, consistent with previous suggestions that these regions have a role in relatively early face processing[5,26]. Self-rated pre-experimental familiarity had no affect on these brain areas (**Table 2**).

An influential model of face processing[27] includes a 'structural encoding' phase that encodes view-specific face representations corresponding to variant aspects that can change between views and images and are considered to be unaffected by face familiarity[27]. This psychological stage seems to be consistent with our findings that subtle physical changes affect activation of IOG and OFA, whereas identity and self-rated familiarity do not. Thus IOG and OFA seem to be a plausible substrate for the structural encoding of faces.

In contrast to IOG, right FFG (and FFA) showed sensitivity to identity change (**Fig. 2b,c**) even when the presence of physical change along the morph continuum was controlled at 30%. This observation provides strong evidence in support of the previous proposals that FFG may have a role in differentiating between individual faces[5,23,25–29]. Our finding can be also reconciled with single-cell evidence for units tuned to facial identity in regions that may be homologous to human FFG[30]. It also fits with a recent report that epileptic activity in right FFG can cause inability to distinguish between faces[31]. The ability of intact right FFG to generalize to a degree across different physical

The difference in fMRI activity for the between minus the within condition correlated positively with intersubject differences in the extent of pre-experimental familiarity (that is, it was more pronounced in those who were more familiar with the faces) in bilateral anterior temporal pole (aTP) and in right anterior hippocampus (aHipp; **Fig. 3** and **Table 2b**). Note that these regions showed no reliable main effect of identity change when the familiarity of the subjects was not considered. The right FFG region (and FFA), which did show an overall effect of identity change (see above), showed no such evidence of modulation by the extent of pre-experimental familiarity (FFG, $P = 0.41$; FFA, $P = 0.6$).

DISCUSSION

By using computerized morphs generated between pairs of famous faces, we have shown that the behavioral identification responses of subjects can be described with a step-like function (**Fig. 1a** and **Supplementary Fig. 1** online) consistent with previous research[16]. In addition, the same-different judgments of the subjects confirm that two faces straddling a putative 'category boundary' for identity are more likely to be judged as different than are two faces that are from the same side of the boundary but separated equally along the morph continuum (**Fig. 1b,c**). The size of this effect correlates positively with the subjects' self-ratings of pre-experimental familiarity. These findings are in accord with previous evidence indicating that categorical perception of face identity is less apparent when the face pair is unknown[16]. Mere physical change

Figure 3 Correlation of self-rated familiarity with identity change effect. Left, statistical parametric maps for regions showing a positive correlation between intersubject pre-experimental familiarity and the BOLD effect of identity change when the presence of a physical shift along the morph continua was controlled (that is, between versus within) with a threshold at $P < 0.001$ (uncorrected) and overlaid on a T1 image of one subject. Right, scatter plots showing the subjects' mean pre-experimental familiarity ratings on the whole face set (x-axis) plotted against their identity change fMRI effect (y-axis) with a fitted regression line for left aTP and right aHipp.
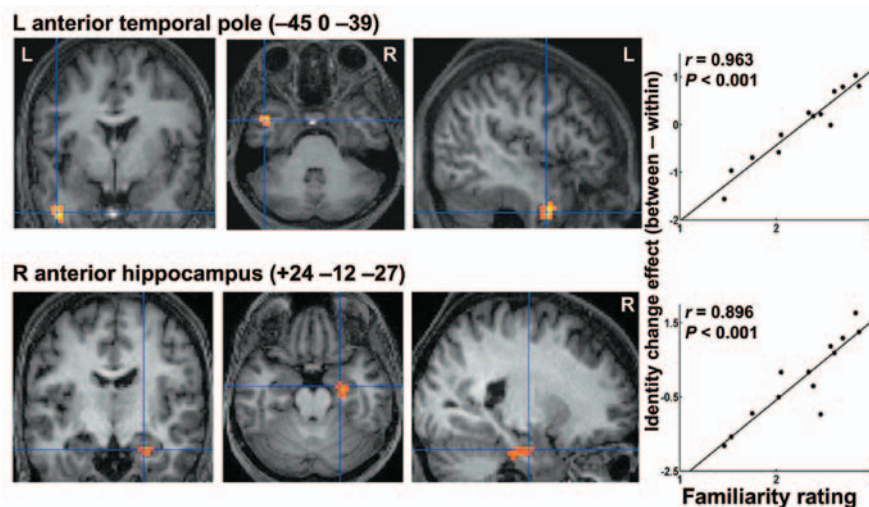
properties and to respect identity category boundaries more than physical properties *per se* (at least for the relatively subtle physical changes used here) also seems to be consistent with repetition-related effects on hemodynamic responses across other types of physical change, such as facial expression[32]; however, generalization across different viewing angles might be limited[22].

One neuroanatomical model of face processing[5] postulates a role for FFG and FFA in identity processing. More specifically, it suggests that FFG may encode relatively invariant aspects of a face and provide a possible anatomical substrate for face recognition units[27]. These units are considered to provide codes of known faces that are matched to input from an earlier structural encoding phase to achieve recognition. Here we found that right FFG and FFA were differentially affected by changing or repeating identity, even while controlling for a shift along the morph continuum, consistent with a role for face recognition units. By contrast, we found no evidence that the effects of identity change in right FFG, unlike those in more anterior temporal regions, were modulated by the extent of self-rated pre-experimental familiarity with the famous faces. Because all of these faces were famous and were shown to the subjects before fMRI, however, future variations on our model, including completely unknown faces, might in principle reveal an influence of prior knowledge on the effects of identity change in FFG.

Nevertheless, the observed effect of crossing identity boundaries on FFG and FFA does suggest that this region has a role in recognizing faces as the same or different despite physical change. Past work has suggested several ways in which this might be achieved. In 'face-space'[18,33], for example, learning about particular faces leads to a 'warping' of a multidimensional space (which might initially encode face parameters fairly linearly), such that each known identity becomes an attractor that can capture faces within its boundaries[34]. Information represented in FFG and FFA might thus provide a warped version of a multidimensional physical feature space that was initially coded by earlier stages in visual processing (such as IOG and OFA, as suggested here). Recordings from neurons in macaque inferior temporal cortex in a supervised categorization task[35] show that neurons tend to become more tuned to 'diagnostic' visual features (such as distance between eyes or length of nose) that distinguish between categories of cartoon faces. From this perspective, the FFG or FFA pattern observed here might reflect activity in a population of such 'diagnostic-feature' neurons. Future variations on our fMRI model, together with single-units recordings and information theoretical analysis[36], may elucidate further the nature of face identity codes in FFG and FFA.

In more anterior regions, namely aTP and right aHipp, we found a positive relationship between the extent of the subjects' self-rated pre-experimental familiarity with the faces on the one hand and the size of identity change effects on the other (**Fig. 3**). This finding is in agreement

with several neuroimaging studies and intracranial recordings of the activation of aTP[37–39] and right aHipp[7,40] in the context of processing familiar faces. We found a linear relationship between the identity change effects and different levels of self-rated pre-experimental familiarity for a set of famous faces with both the aTP results and the behavioral same-different results, thereby validating the self-ratings. One interpretation is that anterior temporal regions may provide access to semantic information regarding a person[5,37], perhaps via personal identity nodes[27]. The more familiar a person's face, the greater the amount of information that can be retrieved about them. This could explain why the effect of repeating versus changing identity was greater in these regions for individuals who were more familiar with the faces used. This proposed role for aTP in storing semantic information with regard to well-known faces seems to be consistent with some lesion studies[40] (but see also ref. 41).

In conclusion, our combination of face morphing, repetition-related BOLD measures and pre-experimental familiarity ratings has enabled us to test for different types of face representation in human occipito-temporal cortex. Our fMRI data indicate that representations in IOG and OFA primarily reflect physical properties of a face and provide evidence that right FFG and FFA are influenced by face identity, because these regions are sensitive to shifts across identity boundaries. Finally, representations in more anterior temporal cortex seem to be influenced by longer-term familiarity with a face, which is presumably derived through more extensive experience. Our findings provide neurobiological support for a classic hierarchical psychological model of face processing[5,27].

METHODS

Subjects. All participants had normal or corrected vision and no neurological or psychiatric history. Informed consent was obtained according to procedures approved by the Joint Ethics Committee of the National Hospital for Neurology and Neurosurgery and the Institute of Neurology, London.

Identity-classification task. Twenty volunteers (ten females; mean age, 28 years; range, 22–39 years) took part in the identity classification experiment. The initial stimulus set compromised 86 achromatic portraits of famous people. Using Photoshop 6.0, all faces were resized to 220 × 250 pixels. To optimize the morphing procedure, the faces were split into 43 pairs by selecting suitable pairs of original faces on the basis of matched gender and composition of the photograph. We then morphed between faces in each pair using Morpher 3.1 to generate a symmetrical continuum of 11 images (morphs) that represented gradual transitions from one original face to the other in steps of 10% (that is, 0 to 100%; ref. 42, **Fig. 1a** and **Supplementary Fig. 1** online).

Before the identity-classification task (and before all of the other behavioral or fMRI experiments), subjects gave familiarity ratings for each original face, by indicating how well they knew that person on a scale of 0 to 3 as follows: 0, 'I don't recognize this face at all'; 1, 'the face looks familiar'; 2, 'I have contextual information regarding that person' (for example, she is a movie star); and 3, 'I know this person well and can name him or her.' In the identity-classification task, each continuum was tested separately, starting with presentation of the two original (unmorphed) faces together with their names[16]. This was followed by 33 trials (three presentations of each of the 11 different morphs) presented in random order. Morphs were presented for 750 ms, followed by a fixation point that remained until the naming response.

We used polynomial regression for parametric characterization of the response function of the identity classification task. To test whether the response function had a categorical component, that is, a step-like function, the parameter estimates for the third order (cubic) function, averaged across subjects, were tested against zero by using a one-sample *t*-test across face pairs.

Same-different perceptual discrimination task. Eleven volunteers (six females; mean age, 30 years; range, 25–38 years) took part in the same-different experiment. Three morphs from each continuum were selected on the basis of performance in the identity classification task (above) that other subjects had undergone. Specifically, two morphs were used that had been reliably perceived as the same identity and were 30% apart along the continuum, in addition to a third that was perceived as a different identity but was also 30% distant from one of the other morphs (**Fig. 1a,b** and **Supplementary Fig. 1** online). These exemplars were used in the within, identical and between conditions, respectively. Two such sets of three morphs were chosen from each continuum (closer to one or the other endpoint), although any one subject viewed only one of these sets, which was selected at random (for example, either the 'Maggie set' or the 'Marilyn set'; **Fig. 1b**). This meant that across subjects the direction of shift along the morph continuum was not confounded by the within or the between conditions. Note that the faces used in the identical and between conditions were roughly the same distance from an original face (at one or other endpoint of the continuum); in addition, across subjects, these face pairs could appear in either of the identical or between conditions in a counterbalanced manner (see above, **Fig. 1a** and **Supplementary Fig. 1** online).

Subjects first viewed all 86 original faces and rated their pre-experimental familiarity (as above). The experiment had a 3 (pair type) by 2 (ISI) factorial design. The three types of pair were identical (same image with same physical properties and identity), within (same side of identity boundary but 30% difference along the morph continuum), and between (different identity—that is, the other side of category boundary—and also 30% difference along the continuum)[20]. Note that the first faces used in each type of pair were held constant (**Fig. 1b**). Each pair was shown twice, once with a long (500-ms) ISI between first and second face, once with a short (75-ms) ISI. Each face was presented for 500 ms with a 2,500-ms stimuli onset asynchrony (SOA) between pairs. Subjects judged whether the two faces showed the same or different person, as were asked to respond as quickly and accurately as possible.

Data were analyzed by Matlab 6.0 and SPSS 11.0. Results for 'different' responses are presented (**Fig. 1c**) as a proportion of the total number of pairs per condition (43). All results were Greenhouse-Geisser corrected for nonsphericity. For the correlation between pre-experimental familiarity ratings and sensitivity to the identity boundary, we subtracted scores (the proportion of 'different' responses, collapsed across ISI) for the within condition from those for the between condition, because this comparison controls for the presence of physical change. The familiarity rating entered for each subject was their average for the whole face set.

Event-related fMRI experiment. Fourteen healthy right-handed volunteers (seven females; mean age, 28 years; range, 22–39 years) took part in the fMRI experiment. Stimuli were analogous to the same-different behavioral experiment above, except that only 39 morph continua were used owing to time constraints. Before scanning, subjects performed the familiarity rating task (see above).

During fMRI, a pair-repetition paradigm was used[21,32]. The factorial design and face-pair presentation durations were identical to the same-different experiment (see above). The order of trials was pseudorandomized to maximize the separation in time between trials with faces from the same continuum. Subjects maintained fixation, pressing a key only when a particular face identity appeared (randomly chosen from three continua that were excluded from the experimental design). The experiment had three phases (without a break in scanning) that started with presentation of a target identity (one of three possibilities) and ended with feedback on performance. There were 36 trials for each of the six conditions. A further 25% of the total trials were null events with only a fixation point and about 15% of the total trials were targets, which could appear as the first or the second face in a pair with either a long or a short ISI.

A 3T Allegra (Siemens) was used to acquired BOLD gradient echoplanar images (EPIs). Images were reconstructed using trajectory-based reconstruction to minimize ghosting and distortion effects in the images[43]. We acquired 44 oblique axial slices (2-mm thick with a 1-mm gap) with 64×64 pixels and an in-plane resolution of 3×3 mm^2, a 90° flip angle, a 30-ms echo time and a 2.860-s repeat time. To reduce susceptibility artifacts in temporal cortices, slices were tilted up 30° anterioraly[44]. A T1-weighted structural image ($1 \times 1 \times 1$ mm^3 resolution) was subsequently acquired.

Whole-brain voxel-based analyses were done with SPM2. The EPI volumes were spatially realigned[45] and unwarped[46]. The time series for each voxel was realigned temporally to acquisition of the middle slice. The EPI images were normalized to a standard MNI reference brain in Talairach space[47] and smoothed with an isotropic 9-mm full-width at half-maximum (FWHM) gaussian kernel. The time series in each voxel was high-pass-filtered to 1/128 Hz. Statistical analysis in SPM used a hierarchical random-effects model with two levels[48]. In the first, single-subject fMRI responses were modeled by a design matrix comprising the onset of the second event[32] for each pair for the six conditions, target events and feedback at end of each phase. These eight regressors were convolved with the canonical hemodynamic response function[49]. To assess and to control for pre-experimental familiarity with the faces, additional covariate regressors for each of the six conditions encoding the familiarity rating of each subject with each face pair were added. Linear contrasts pertaining to the main effects, interactions and simple effects were calculated for each subject.

To allow inferences at the population level, a second-level analysis considered the images resulting from contrasts for each subject in which subjects were treated as a random effect. The contrast images were first smoothed using a 3-mm FWHM gaussian kernel to account for residual intersubject differences[50] (the final estimated smoothness was about $11 \times 11 \times 11$ mm^3 FWHM). Simple effects were computed in SPSS by using parameter estimates (linearly transformed to a percentage signal change relative to the grand mean over all voxels and scans) extracted from the maxima (**Fig. 2a,b**). All reported results are Greenhouse-Geisser corrected to account for nonsphericity.

We tested for correlations with intersubject self-rated pre-experimental familiarity by using a simple correlation analysis across subjects between the relevant contrast images (for example, for identity change: between versus within) and the mean familiarity across all pairs for each subject.

Functional localizer and fROI analyses. Eight subjects (four females; mean age, 29 years; range, 22–39 years) who participated in the previous fMRI experiment (see above) had a functional localizer scan immediately after the main experiment. The stimuli comprised ten achromatic photos of neutral faces taken from Ekman's series cropped to remove hair, ten scrambled versions of these faces generated in a 15×10 grid (but with the face outline kept intact), and ten achromatic photos of houses cropped to produce a similar elliptical shape and size as the intact and scrambled faces.

Localizer stimuli were presented in a blocked design, with ten stimuli from the same category (faces, houses or scrambled faces) in each block. Each stimulus was presented for 250 ms with an SOA of 850 ms; the total duration of each block was 11 s. Blocks were separated with a 6.3-s fixation point presented on a gray background. The task was to select by pressing a button immediate repetitions, which occurred about 15% of the time across all conditions. Each block was repeated 4–5 times.

EPIs were acquired as in the main fMRI study. Identical preprocessing procedures were used as in the main fMRI experiment, except that there was no slice time realignment owing to the blocked design of the localizer. A design matrix with three regressors for the onset of each block type (face, house, scrambled face) convolved with the canonical HRF was used to model the predicted BOLD response. For each subject, we computed the contrast of 'face versus house plus scrambled face' and extracted the coordinates for the peak corresponding to right FFA and right OFA, as previously defined[3,24]. These regions were identified reliably for only seven of the eight subjects. Using these coordinates, we then

extracted the parameter estimates from the first level (that is, subject-specific) analysis of the event-related morphing experiment for each of the seven subjects. These parameter estimates were linearly transformed to a percentage signal change, and statistical analysis was done with SPSS. All reported results are Greenhouse-Geisser corrected for nonsphericity.

URLs. Morpher 3.1, http://www.asahi-net.or.jp; SPM2, http://www.fil.ion.ucl.ac.uk/spm.

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

1. Posamentier, M.T. & Abdi, H. Processing faces and facial expressions. *Neuropsychol. Rev.* **13**, 113–143 (2003).
2. Farah, M.J., Wilson, K.D., Drain, M. & Tanaka, J.N. What is 'special' about face perception? *Psychol. Rev.* **105**, 482–498 (1998).
3. 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).
4. Malach, R., Levy, I. & Hasson, U. The topographic of high-order human object areas. *Trends Cogn. Sci.* **6**, 176–184 (2002).
5. Haxby, J.V., Hoffman, E.A. & Gobbini, M.I. The distributed human neural system for face perception. *Trends Cogn. Sci.* **4**, 223–233 (2000).
6. Leonard, C.M., Rolls, E.T., Wilson, F.A. & Baylis, G.C. Neurons in the amygdala of the monkey with responses selective for faces. *Behav. Brain Res.* **15**, 159–176 (1985).
7. Ishai, A., Haxby, J.V. & Ungerleider, L.G. Visual imagery of famous faces: effects of memory and attention revealed by fMRI. *NeuroImage* **17**, 1729–1741 (2002).
8. Rajah, M.N., McIntosh, A.R. & Grady, C.L. Frontotemporal interactions in face encoding and recognition. *Brain Res. Cogn Brain Res.* **8**, 259–269 (1999).
9. Seeck, M. *et al.* Intracranial neurophysiological correlates related to the processing of faces. *Epilepsy Behav.* **2**, 545–557 (2001).
10. Perrett, D., Rolls, E.T. & Caan, W. Visual neurons responsive to faces in the monkey temporal cortex. *Exp. Brain Res.* **47**, 329–342 (1982).
11. McCarthy, G., Puce, A., Belger, A. & Allison, T. Electrophysiological studies of human face perception. II. Response properties of face-specific potentials generated in occipitotemporal cortex. *Cereb. Cortex* **9**, 431–444 (1999).
12. De Renzi, E. Prosopagnosia in two patients with CT scan evidence of damage confined to the right hemisphere. *Neuropsychologia* **24**, 385–389 (1986).
13. Tranel, D., Damasio, H. & Damasio, A.R. A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia* **35**, 1319–1327 (1997).
14. Dolan, R. *et al.* How the brain learns to see objects and faces in an impoverished context. *Nature* **389**, 596–569 (1997).
15. Henson, R.N. *et al.* Electrophysiological and heamodynamic correlates of face perception, recognition and priming. *Cereb. Cortex* **13**, 793–805 (2003).
16. Beale, J.M. & Keil, F.C. Categorical effects in perception of faces. *Cognition* **57**, 217–239 (1995).
17. Nosofsky, R.M. Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol. Gen.* **115**, 39–61 (1986).
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. Harnad, S. Psychophysical and cognitive aspects of categorical perception. in *Categorical Perception: The Ground of Cognition* (ed. Harnad, S.) 1–25 (Cambridge Univ. Press, New York, 1987).
20. Campanella, S. *et al.* Right N170 modulation in a face discrimination task: an account for categorical perception of familiar faces. *Psychophysiology* **37**, 796–806 (2000).
21. Kourtzi, Z. & Kanwisher, N. Representation of perceived object shape by the human lateral occipital complex. *Science* **293**, 1506–1509 (2000).
22. Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y. & Malach, R. Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron* **21**, 191–202 (1998).
23. Henson, R.N., Shallice, T. & Dolan, R. Neuroimaging evidence for dissociable forms of repetition priming. *Science* **287**, 1269–1272 (2000).
24. Henson, R.N. Neuroimaging studies of priming. *Prog. Neurobiol.* **70**, 53–81 (2004)
25. Grill-Spector, K., Kourtzi, Z. & Kanwisher, N. The fusiform face area subserves face perception not generic within-category identification. *Nat. Neurosci.* **7**, 555–562 (2004).
26. de Gelder, B. & Rouw, R. Beyond localization: a dynamical dual route account of face recognition. *Acta. Psychol.* **107**, 183–207 (2001).
27. Bruce, V. & Young, A. Understanding face recognition. *Br. J. Psychol.* **77**, 305–327 (1986).
28. Gauthier, I. *et al.* The fusiform 'face area' is part of a network that processes faces at the individual level. *J. Cogn. Neurosci.* **12**, 495–504 (2000).
29. George, N. *et al.* Contrast polarity and face recognition in the human fusiform gyrus. *Nat. Neurosci.* **2**, 574–580 (1999).
30. Hasselmo, M.E., Rolls, E.T. & Baylis, G.C. The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav. Brain Res.* **32**, 203–218 (1989).
31. Mundel, T. *et al.* Transient inability to distinguish between faces: electrophysiologic studies. *J. Clin. Exp. Neuropsychol.* **20**, 102–110 (2003).
32. Winston, J.S., Henson, R.N., Fine-Goulden, M.R. & Dolan, R.J. fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *J. Neurophysiol.* **92**, 1830–1839 (2004).
33. Sirovich, L. & Kirby, M. Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am.* **4**, 519–524 (1987).
34. Tanaka, J., Giles, M., Kremen, S. & Simon, V. Mapping attractor fields in face space: the atypicality bias in face recognition. *Cognition* **68**, 199–220 (1998).
35. Sigala, N. & Logothetis, N.K. Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* **415**, 318–320 (2002).
36. Rolls, E.T., Treves, A., Tovee, M.J. & Panzeri, S. Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. *J. Comput. Neurosci.* **4**, 309–333 (1997).
37. Gorno-Tempini, M.L. *et al.* The neural systems sustaining face and proper-name processing. *Brain* **121**, 2103–2118 (1998).
38. Leveroni, C.L. *et al.* Neural systems underlying the recognition of familiar and newly learned faces. *J. Neurosci.* **20**, 878–886 (2000).
39. Sugiura, M. *et al.* Activation reduction in anterior temporal cortices during repeated recognition of faces of personal acquaintances. *NeuroImage* **13**, 877–890 (2001).
40. Glosser, G., Salvucci, A.E. & Chiaravalloti, N.D. Naming and recognizing famous faces in temporal lobe epilepsy. *Neurology* **61**, 81–86 (2003).
41. Barton, J.J.S. & Cherkasova, M. face imagery and its relation to perception and covert recognition in prosopagnosia. *Neurology* **61**, 220–225 (2003).
42. Winston, J.S., O'Doherty, J. & Dolan, R.J. Common and distinct neural response during direct and incidental processing of multiple facial emotions. *NeuroImage* **20**, 84–97 (2003).
43. Josephs, O., Deichmann, R. & Turner, R. Trajectory measurment and generalised reconstruction in rectilinear EPI. *NeuroImage* **11**, S543 (2000).
44. Deichmann, R., Gottfried, J.A. & Turner, R. Optimized EPI for fMRI studies of the orbitofrontal cortex. *NeuroImage* **19**, 430–441 (2003).
45. Friston, K.J., Williams, S.R., Howard, R., Frackowiak, R.S.J. & Turner, R. Movement-related effects in fMRI time-series. *Magn. Reson. Med.* **35**, 346–355 (1996).
46. Andersson, J.L., Hutton, C., Ashburner, J., Turner, R. & Friston, K.J. Modeling geometric deformations in EPI time series. *NeuroImage* **13**, 903–919 (2001).
47. Ashburner, J. & Friston, K.J. Nonlinear spatial normalization using basis function. *Hum. Brain. Mapp.* **7**, 254–266 (1999).
48. Penny, W. & Friston, K.J. Hierarchical models. in *Human Brain Function II* (eds. Frackowiak, R.S., Friston, K.J., Dolan, R. & Ashburner, J.) 851–863 (Elsevier Academic, 2003).
49. Friston, K.J. *et al.* Event-related fMRI: characterizing differential responses. *NeuroImage* **7**, 30–40 (1998).
50. Friston, K.J., Holmes, A.P., Price, C.J., Buchel, C. & Worsley, K.J. Multisubject fMRI studies and conjunction analyses. *NeuroImage* **10**, 385–396.