# Individual faces elicit distinct response patterns in human anterior temporal cortex

**Nikolaus Kriegeskorte*†, Elia Formisano‡, Bettina Sorger‡, and Rainer Goebel‡**

*Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD 20892; and ‡Department of Cognitive Neuroscience, Faculty of Psychology, Universiteit Maastricht, 6200 MD Maastricht, The Netherlands

Edited by Leslie G. Ungerleider, National Institutes of Health, Bethesda, MD, and approved October 5, 2007 (received for review June 19, 2007)

**Visual face identification requires distinguishing between thousands of faces we know. This computational feat involves a network of brain regions including the fusiform face area (FFA) and anterior inferotemporal cortex (aIT), whose roles in the process are not well understood. Here, we provide the first demonstration that it is possible to discriminate cortical response patterns elicited by individual face images with high-resolution functional magnetic resonance imaging (fMRI). Response patterns elicited by the face images were distinct in aIT but not in the FFA. Individual-level face information is likely to be present in both regions, but our data suggest that it is more pronounced in aIT. One interpretation is that the FFA detects faces and engages aIT for identification.**

fMRI | information-based | population code

When we perceive a familiar face, we usually effortlessly recognize its identity. Identification requires distinguishing between thousands of faces we know. A puzzle to both brain and computer scientists, this computational feat involves a network of brain regions (1) including the fusiform face area (FFA) (2, 3) and anterior inferotemporal cortex (aIT) (4). There is a wealth of evidence for an involvement in face identification of both the FFA (1, 5–18) and aIT (4, 16, 19–26).

The FFA responds vigorously whenever a face is perceived (2, 3, 27). This implies that the FFA distinguishes faces from objects of other categories and suggests the function of face detection (27, 28). An additional role for the FFA in face identification has been suggested by three lines of evidence: (*i*) Lesions in the region of the FFA are frequently associated with deficits at recognizing individual faces (prosopagnosia) (6, 9, 10). (*ii*) The FFA response level covaries with behavioral performance at identification (11). (*iii*) The FFA responds more strongly to a sequence of different individuals than to the same face presented repeatedly (8, 12–17).

For aIT as well, human lesion and neuroimaging studies suggest a role in face identification. Neuroimaging studies (4, 22–24, 26) found anterior temporal activation during face recognition with the activity predictive of performance (22). Lesion studies (19, 20, 25) suggest that right anterior temporal cortex is involved in face identification. In monkey electrophysiology, in fact, face-identity effects appear stronger in anterior than in posterior inferotemporal cortex (29–31).

These lines of evidence suggest an involvement of both the FFA and aIT in face identification. A region representing faces at the individual level should distinguish individual faces by its activity pattern. However, it has never been directly demonstrated that either the FFA or aIT responds with distinct activity patterns to different individual faces.

We therefore investigated response patterns elicited by two face images by means of high-resolution functional magnetic resonance imaging (fMRI) at 3 Tesla (voxels: $2 \times 2 \times 2$ mm$^3$). We asked whether response patterns associated with the faces are statistically distinct. This would mean that the activity patterns allow the decoding from the fMRI signals (32–42) of the perceived individual.

The decision to use only two particular face images involves a trade-off. The disadvantage consists in the fact that any two face images necessarily differ along many dimensions. We are, thus, throwing a wide net: effects are expected in any brain region that represents at least one of the dimensions distinguishing the face images. Although these regions should include the putative individual-face representation, interpretation will be difficult if several regions are found to distinguish the faces.

The advantage of using only two face images consists in the fact that we do not need to average response patterns elicited by different images. For each image, averaging responses to its repeated presentation yields a sufficiently stable response pattern, which can be regarded as an estimate of the response on a single perceptual trial. In contrast, previous studies using large numbers of images needed to average response patterns elicited by different images, usually from the same category. Such average response patterns are hard to interpret, because it is unclear whether they actually arise on any single trial of perception.

The faces (one male, one female; Fig. 1) were presented in the same size, view, and lighting. Because of this matching of the two images and because faces in general are similar in their overall shape, the two face images are by many measures (e.g., spatial image correlation) much more similar than the images conventionally contrasted in object-vision neuroimaging. This raises the question whether fMRI will have sufficient resolution and sensitivity to detect any effect at all.

To be able to replicate previously described face-category effects (2), we included two house images (43) in the experiment as control stimuli. To minimize low-level confounds, we processed the four images to have identical histograms and, thus, identical light and spatial-signal energy. Subjects were presented with the images in a rapid event-related design, in which they performed an anomaly-detection task, requiring them to pay close attention to each repeated presentation of an image [Fig. 1; and see supporting information (SI) Figs. 4 and 5 for behavioral performance during fMRI].

## Results

**Conventional Activation-Based Analysis.** Conventional univariate mapping analysis of our data yielded the category effects expected on the basis of the literature. Face-category activation (faces–houses) was very strong in the right and left fusiform gyrus, revealing the FFA (SI Fig. 6*a*, single subject; SI Fig. 7, Talairach-space group map). Weaker face-category activation was found in right and left aIT in the group analysis (SI Fig. 7). Single-subject and group mapping analyses also revealed the parahippocampal place area (43).

Responses to single images from the same category have not previously been contrasted with fMRI. Contrasting the two faces in

Author contributions: N.K. designed research; N.K., E.F., B.S., and R.G. performed research; N.K. analyzed data; and N.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

†To whom correspondence should be addressed. E-mail: kriegeskorten@mail.nih.gov.

This article contains supporting information online at www.pnas.org/cgi/content/full/0705654104/DC1.

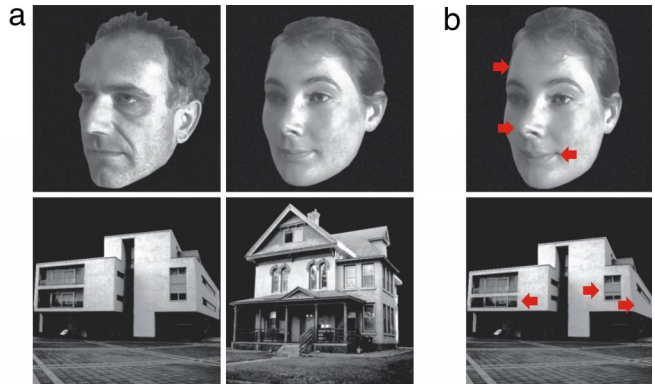


**Fig. 1.** Stimuli and anomaly-detection task. (*a*) The four particular images whose inferotemporal response patterns are investigated in this study. Each image was processed to have a precisely uniform histogram. The images were presented sequentially while subjects fixated a central cross (not shown). Subjects performed an anomaly-detection task: ≈12% of the images were subtle variations (*b*) of the four originals (*a*), in which the global shape of the object as well as details had been slightly distorted (red arrows). Anomalies were unpredictable because several anomalous versions were used for each original. The task required subjects to attend to each image presentation even after many repetitions and allowed us to monitor attentive viewing.

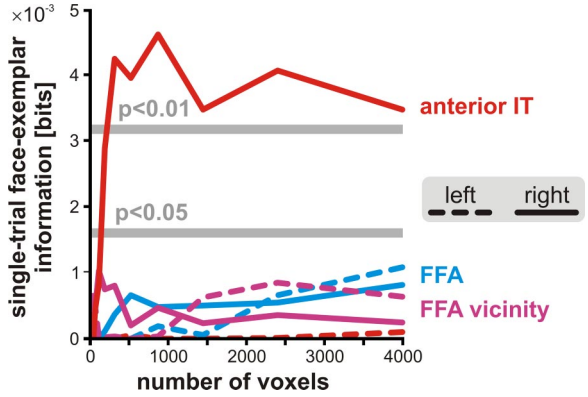


**Fig. 2.** Face-exemplar information as a function of region size. When we define the FFA by the category contrast (SI Fig. 6) and vary the threshold to select between 10 and 4,000 contiguous voxels, significant face-exemplar information is not found at any threshold (blue dashed line, left FFA; blue solid line, right FFA). When we define the ''FFA vicinity'' as the 4,000 cortical voxels in a sphere centered on the FFA in each subject and hemisphere (SI Fig. 6) and select the *n* voxels containing most face-exemplar information on independent data, significant face-exemplar information is not found for any threshold (magenta dashed line, left FFA vicinity; magenta solid line, right FFA vicinity). When we define aIT in each subject and hemisphere as the 4,000 most anterior voxels in temporal cortex and, again, select the *n* voxels containing most face-exemplar information on independent data, no significant face-exemplar information is found for the left hemisphere (red dashed line). However, robust face-exemplar information is found in right aIT (red solid line) The figure shows group results for regions of interest defined in each individual subject. Independent data were used for (*i*) defining the regions and voxel weights and (*ii*) testing the multivariate effects and estimating face-exemplar information.

a conventional univariate mapping analysis did not yield significant effects either in single subjects or at the group level for either unsmoothed or smoothed (6-mm full-width at half-maximum) data. Contrasting the two houses did reveal effects in early visual areas. This is plausible, given that the two house images have very different distributions of low-level features, whereas the two faces are very similar in terms of low-level image features. (For further activation-based control analyses, see *SI Text*, *Results of Control Analyses*.)

**Information-Based Activity-Pattern Analysis.** The signature of a distributed face-exemplar representation would be a subtle difference between the fine-grained regional response patterns elicited by the two face images—despite their visual similarity. Because univariate mapping is not sensitive to subtle response-pattern differences, we performed an information-based multivariate analysis, which is designed for this type of effect (44, 45). A significant multivariate difference between the response patterns elicited by the two faces in a given region of interest (ROI) would indicate the presence of face-exemplar information. The information estimate (in bits) reflects how accurately the face exemplar could be decoded from the ROI's multivariate response on a single trial (main results in Fig. 2, see *Methods*, SI Fig. 8 and *SI Text*). Independent data were used for (*i*) defining the regions and voxel weights and (*ii*) testing the multivariate effects and estimating face-exemplar information. All pattern-information analyses were performed on unsmoothed data.

**Is There Face-Exemplar Information in the FFA?** The FFA was defined by the category contrast (faces–houses) in each individual subject (false-discovery rate, $q < 0.05$). No significant face-exemplar information was found in the FFA in any subject ($P > 0.05$). To maximize statistical power, we combined the data from the individually defined FFAs in a fixed-effects group analysis (see *Methods*). The FFA face-exemplar information was insignificant ($P > 0.05$) in the group analysis as well.

Could the threshold used to define the FFA have excluded face-exemplar voxels at the fringe of the region? To include more voxels at the fringe of the FFA, we systematically varied the threshold of the category contrast to select a contiguous set of 10–4,000 voxels in each subject. Although the resulting extended "FFA" at 4,000 voxels is a huge region reaching far into occipital and anterior temporal cortices, this did not reveal any significant face-exemplar information in either hemisphere (Fig. 2, blue lines, multivariate fixed-effects group analysis, $P > 0.05$).

Could the conventional definition of the FFA by the category contrast have entailed a bias against inclusion of voxels carrying face-exemplar information? Face-exemplar voxels excluded by the conventional definition of the FFA might nevertheless belong to the same functional unit. Another possibility is that the use of a different reference category (e.g., objects instead of houses) would change the precise ROI for the FFA and reveal face-exemplar information. A third possibility is that the face exemplar is encoded in a more widely distributed fashion in the FFA and its vicinity.

To exclude all three possibilities, we asked whether there is face-exemplar information in the vicinity of the FFA (including the FFA itself). To find face-exemplar information, we searched for it in each subject separately using a multivariate searchlight (44): For each voxel, we selected the 3-mm-radius spherical neighborhood (comprising 19 voxels) and computed the Mahalanobis distance reflecting the difference between the activity patterns elicited by the two faces. The Mahalanobis distance for each voxel's spherical neighborhood was entered in a descriptive map called the "face-exemplar information map." This information-based map (as well as the activation-based map used to define the FFA) was based on half the data (data set A) of each subject. Statistical tests and information estimates were based on independent data (the other half: data set B).

The "FFA vicinity" was then defined, for each subject and hemisphere, as 4,000 cortical voxels within a sphere centered on (and including) the FFA (SI Fig. 6 *c* and *d*, magenta). For each subject and hemisphere, the FFA vicinity was tested for face-exemplar information by the following procedure for $n = 10$ to $n = 4{,}000$ voxels: (*i*) Select the *n* voxels (of the 4,000) with the greatest values in the face-exemplar information map. (*ii*) Perform a multivariate fixed-effects group analysis on this voxel set using independent data (see *Methods*). Fig. 2 shows the results. Face-exemplar

NEUROSCIENCE

information remains insignificant in the FFA vicinity of either hemisphere (magenta), independent of the number of voxels included.

**Is There Face-Exemplar Information in aIT?** Analogously to the FFA vicinity, we defined aIT in each subject and hemisphere as the 4,000 most anterior voxels in temporal cortex within our inferotemporal imaging slab (SI Fig. 6 *c* and *d*, red). We then tested aIT in exactly the same way as the FFA vicinity: we selected between 10 and 4,000 voxels with the greatest values in the face-exemplar information map in each subject and performed the same multivariate group test on independent data.

Fig. 2 shows the results. In left aIT (red dashed line), face-exemplar information remains insignificant, independent of the number of voxels included. In right aIT, by contrast, face-exemplar information becomes highly significant ($P < 0.01$) when more than approximately 200 voxels are included.

**Finding Face-Exemplar Information by Searchlight Mapping.** To address more broadly whether face-exemplar information is present in any region within our fMRI slab, we performed a group-statistical information-based brain mapping (44) with randomization inference. We first computed face-exemplar information maps with a 3-mm-radius searchlight in each subject separately, as above, but using all data. These maps were transformed into Talairach space and subjected to group-statistical inference (see *Methods*).

This method provides an alternative perspective, (*i*) because it is not restricted to predefined regions of interest and (*ii*) because informative regions need to correspond in Talairach space to be sensitively detected [although their intrinsic pattern representations can be unique to each individual subject (45)]. Despite these differences to the ROI-based analysis, results were consistent: Face-exemplar information was found only in right aIT (global maximum of the group-statistical map shown in Fig. 3 and SI Fig. 9*b*, Talairach coordinates of centroid: 38, 2, −38, $P < 0.0001$ at peak voxel).

**Summary of Results.** Information-based multivariate ROI and mapping analyses indicate that right aIT responds with a distinct activity pattern to each of the faces. We found no evidence of face-exemplar information in any other region within our temporal-occipital imaging slab in either the information-based searchlight mapping or the multivariate ROI analyses performed on the FFA and its vicinity.

## Discussion

**Individual-Face Information Is Unlikely to Be Completely Absent in the FFA.** At face value, our findings could be taken to suggest that the FFA is invariant to differences between faces. However, given the evidence from previous studies (1, 5–18, 51), we do not believe that the FFA is cleanly invariant to face identity.

We have made every effort to optimize sensitivity to fine-scale effects and go beyond previous work by combining high-resolution fMRI and information-based multivariate analysis of local response patterns. Nevertheless, the use of blood oxygen-level dependent (BOLD) fMRI with isotropic 2-mm-width voxels limits the neuronal activity pattern differences we are capable of detecting. Undetected information could reside in the fine-grained activity patterns beyond the limits imposed by voxel size and hemodynamics. Alternatively, face-exemplar information could be encoded in the temporal activity pattern, to which our statistical model here is insensitive. Our results therefore clearly do not imply the absence of face-exemplar information in the FFA or elsewhere in the brain.[§]

[§]The absence of an effect (however small) can never be statistically demonstrated. This is a general limitation in science, but particularly severe here, because each fMRI voxel reflects the activity of hundreds of thousands of neurons pooled across seconds.

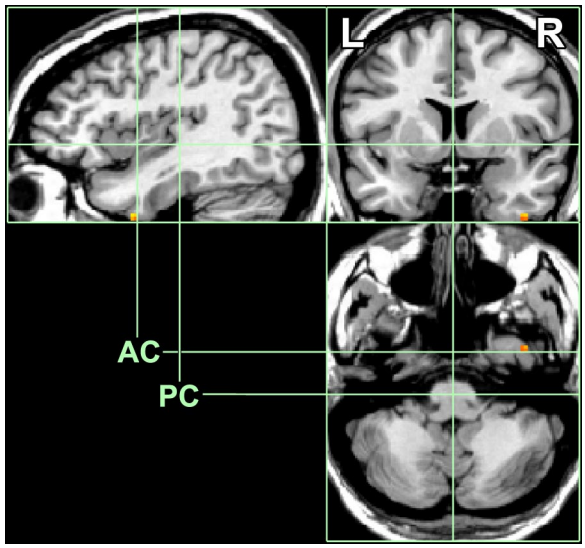


**Fig. 3.** Peak of distributed face-exemplar information in aIT. We used information-based functional brain mapping (44) to determine where locally distributed face-exemplar information was greatest within our occipitotemporal fMRI slab. The only region found was in right aIT (Talairach coordinates of centroid 38, 2, −38). The full information-based Talairach-space group map is shown in SI Fig. 9*b*. A single subject's map and the event-related spatial response patterns in the anterior face-exemplar region are shown in SI Fig. 11. The group map was thresholded to highlight voxels with $P < 0.001$, uncorrected (orange-yellow). The peak voxel had $P < 0.0001$ (yellow). Information-based mapping was performed in each subject by using a 3-mm-radius spherical information searchlight (44) (see *Methods*, SI Fig. 8*b*, and SI Text: *Information-Based Group Mapping in Talairach Space*); thus, each highlighted voxel indicates face-exemplar information distributed within a more extended local neighborhood (volume highlighted: 8 voxels = 64 mm$^3$, volume contributing information: 56 voxels = 448 mm$^3$). Single-subject information-based maps were transformed into Talairach space and averaged across subjects. Statistical inference was performed at each voxel by a randomization test involving random relabeling of the face trials. The background shows the MNI template brain transformed into Talairach space. Green boxes indicate the cuboid subvolumes of Talairach space. Anterior commissure (AC) and posterior commissure (PC) are indicated. The right hemisphere is on the right side in the coronal and axial slices.

Our findings do suggest that any face-identity effect is much weaker than the category effect in the FFA.

**Individual-Face Information Appears Most Pronounced in the aIT.** That fMRI could detect face-exemplar information in aIT, but not in the FFA or its vicinity, suggests that individual-level face information is, by at least one measure, more pronounced in aIT than in the FFA and its vicinity. The FFA and other regions are likely to contain face-exemplar information as well, at lower levels. In particular, the face images must have elicited subtly different activity patterns in early visual cortex with its retinotopic maps of low-level features. Because the faces were matched in size, view, lighting, and histogram, neither information-based mapping nor ROI analysis revealed a face-exemplar effect in early visual cortex, although all other pairs of stimuli could be distinguished (see SI Fig. 10 and *SI Text*, *Results of Control Analyses*). That our methods revealed the face-exemplar information in aIT, but not early visual cortex, suggests that the subtle differences in the early visual representation are magnified in ventral-stream processing to yield a much larger, and thus detectible, difference in aIT. The sensitivity of fMRI to the difference in aIT indicates that there are massive neuronal face-exemplar effects in that region. This suggests a functional role of this region in distinguishing individual faces (SI Fig. 11).

**Low-Level Confounds Cannot Account for the aIT Face-Exemplar Effect.** The absence of significant face-exemplar information in the early visual fMRI patterns (see SI Fig. 10 and *SI Text*, *Results of*

*Control Analyses*) suggests that the faces were appropriately matched for our purposes. This is plausible for a number of reasons. (*i*) Faces, in general, are similar in feature set and global configuration. One consequence of this is a similar spatial frequency spectrum. (*ii*) In addition, the faces were matched in size, view, and lighting, which yielded similar retinotopic images. (*iii*) Furthermore, we matched the image histograms. As a consequence, our stimuli also had identical light and spatial-signal energy. The two face images were, thus, much better matched for low-level confounds than can be achieved, for example, when a set of face images is contrasted against a set of other object images to localize face-sensitive regions including the FFA.

**Both the FFA and aIT May Be Necessary for Face Identification.** Our findings may appear at odds with the studies suggesting a role for the FFA in face identification (1, 5–18). However, the contrasting evidence can be reconciled: the FFA may detect faces (2, 3, 27, 28), engage aIT to identify them (4, 16, 19–26), and subsequently receive feedback from aIT. In this view, face identification requires both regions, and the activity of both should predict success and failure of the process. This would explain why (*i*) lesions in the region of the FFA are associated with deficits at recognizing individual faces (6, 9, 10) and why (*ii*) the FFA response level reflects behavioral performance at identification (11).

The face-processing stages of detection and identification have been associated with the successive components M100 and M170 in a magnetoencephalography study (46). Having detected a face, the FFA may not only trigger identification in aIT but, more generally, engage specialized nodes of the core and extended face network (1, 5) for detailed analysis, including analysis of facial expression in STS (7).

Our interpretation is also consistent with the third line of evidence for a role for the FFA in identification, namely that (*iii*) the FFA responds more strongly to a sequence of different individuals than to the same face image presented repeatedly (8, 12–17). The greater response to face-identity change than face-identity repetition is usually taken to indicate neuronal information about face identity. This interpretation is based on the idea of "fMRI release from adaptation" (47), which is expected to occur if each identity drives a different set of FFA neurons. If each set of neurons representing a face drove each of our voxels approximately equally, we might well have failed to detect the information, because our approach of direct measurement and analysis of the response patterns is limited by fMRI spatial resolution. The fMRI adaptation technique, by contrast, is not limited by the fMRI resolution. The presence of some amount of individual face information in the FFA appears likely and would be consistent with our interpretation here.

However, the interpretation of the fMRI adaptation results requires some caution (48, 49), because release-from-adaptation effects can carry over from a region A to another region B, even if the projection pools responses so that the selectivity causing the release in A is not present in B. For example, release from adaptation in a low-level region could carry over to the FFA, even if the projection pools responses so that the low-level selectivity is lost in the FFA. Similarly an aIT release from adaptation upon identity change as previously reported (16) could carry over to the FFA as an unspecific activation, even if identity could not be decoded from FFA neuronal responses. More generally, a change of perceived face identity is likely to trigger an attentional response entailing widespread activation. All affected regions (either within the face network or beyond it) would then exhibit a release-from-adaptation effect. Furthermore, if exact-image repetition defines the baseline (as in most cases, but see refs. 12, 15, and 16), it is unclear whether the release from adaptation is caused by face-identity change or low-level feature change.

From a computational perspective, face detection is a difficult task, particularly for cluttered scenes, and might well merit a dedicated functional region. There is no strong theoretical reason to believe that detection and identification must be colocalized. In fact, the representational basis functions optimal for face detection are very different from those optimal for distinguishing individuals. In a simple template-matching framework, detection would require something like an average-face template, whereas identification would require multiple templates sensitive to the subtle differences between faces.

**Our Findings Are Consistent with a Wide Range of Monkey and Human Studies.** ***Monkey electrophysiology, neuroimaging, and lesion studies.*** Our aIT finding is consistent with monkey cell recordings, where face-specific responses are found in many locations, but identity-specificity is strongest in the anterior temporal cortex (29–31). The monkey aIT representation has recently been described as a norm-based code for individual faces (50). Another recent study (51) investigated single-cell responses in the monkey middle face patch, which might be the homologue of the human FFA (52). These authors show that cell responses in the monkey middle face patch contain both face-category and face-identity information. However, category information is more pronounced, and identity information becomes available at a greater latency. An earlier monkey study (53) showed that bilateral ablation of the monkey STS (including the region of the middle face patch) does not entail face-identification deficits, suggesting that the middle face patch might not be the main locus of face identification.

All these results are consistent with the interpretation that the middle face patch detects faces, engages aIT to individuate them, and then receives feedback from aIT. It is unclear, however, how closely the human FFA resembles the monkey middle face patch at the level of single-cell responses.

***Human electrophysiology and neuroimaging.*** An early study using positron emission tomography (PET) describes bilateral anterior temporal activation associated with performance of a face-identity task (4). Bilateral anterior temporal lobe exhibits a reduced response to repeated presentations of familiar faces (24). Right temporal polar cortex, in particular, has been found to be active during face perception and recognition (with its activity predictive of performance) (22), during discrimination of familiar and unfamiliar faces (23), and during the naming of faces (54). The latter study suggests that the right temporal pole serves a face-specific function, whereas the left temporal pole is domain-generally involved in naming unique entities (54). These human imaging studies all used PET.

Using fMRI, right anterior temporal cortex has also been found active during face-from-name retrieval (26), but the region is superior to ours. Some fMRI studies may have missed aIT effects, because aIT is often affected by a large fMRI-signal dropout (55, 56) caused by heterogeneous magnetic susceptibility of the local anatomy. This would explain why studies employing PET (4, 22–24, 54) or higher-resolution fMRI (16) [which is less affected by the dropout (55)] more consistently report anterior temporal activity related to face recognition. For example, there is fMRI-adaptation evidence (16) for an involvement of both the FFA and bilateral aIT in face-identity representation. This study used 2-mm slices (1-mm gap) with $3 \times 3\text{-mm}^2$ in-plane voxel size.

Our findings are also consistent with three related studies describing intracranial electrophysiological recordings in human patients exposed to face stimuli (57–59). The authors describe a human "anterior face area" located in the right hemisphere and giving rise to the face-specific AP350 potential (57), which is shown to be reduced on repetitions of the same face (59). This response reduction on repetition is consistent with an identity representation in right aIT. A reduction on repetition was not found in the earlier face-specific N200 originating in more posterior ventral cortex (59).

***Acquired prosopagnosia.*** Prosopagnosia is the inability to recognize individual faces. This disorder can be acquired by brain damage. In particular, it can be caused by lesions in the general region of the FFA (6, 9, 10). Damage to the FFA in these cases may impair

engagement of aIT for identification. Prosopagnosia can also occur in the presence of a face-selectively responding FFA (17, 60), demonstrating that other face-specific regions besides the FFA are needed for identification. In the patient in question (PS) (17, 60), the FFA responses to sequences of faces of different identity appear altered (17), which could be caused by altered input or feedback to the FFA. Prosopagnosia can also be caused by anterior temporal lesions (19, 20, 25). Lesions in the right temporal polar cortex can impair face identification (19). Right anterior temporal atrophy is frequently associated with progressive prosopagnosia (20). Gainotti *et al.* (25) describe a patient with a right anterior temporal focal atrophy associated with impaired identification of familiar people from their faces or voices. The authors suggest that the impairment may be caused by damage to face-recognition units (61) in the right anterior temporal cortex.

***Congenital prosopagnosia.*** Prosopagnosia can also be congenital in the absence of any apparent brain damage (62). Like acquired prosopagnosics, congenital prosopagnosics can detect faces (63) and often exhibit intact FFA activity (64, 65). However, they cannot identify faces, and there is evidence of decreased cortical volume in the right anterior temporal cortex (62, 66).

**What Is the Nature of the Human aIT Face Representation?** The presence of face-exemplar information in aIT suggests that aIT contains a population code representing the subtle differences between individual faces. However, many questions remain.

***What face properties are represented in the aIT?*** First, individual faces in general and our stimuli in particular differ along many dimensions. There is an extended literature addressing face-space dimensions (67, 68) including gender, age, attractiveness, overall configuration, local features and skin texture. The relative importance of these dimensions in the aIT face representation needs to be elucidated.

***How does aIT face representation relate to memory?*** Anterior temporal activity has been found to be greater for familiar than unfamiliar faces and has therefore been associated with access to memory about people (1, 5, 24) (see also refs. 69 and 70). In our study, faces were perceptually familiar to the subjects from a task training immediately preceding the experiment, but subjects did not have any conceptual knowledge (e.g., names, biographical information) about the individuals. A perceptual representation of face identity in interaction with medial temporal memory regions might be expected to show greater activity for familiar faces. Feedback from memory could provide *a priori* information serving to stabilize the activity pattern representing the individual, thus reducing perceptual noise. As a consequence, familiar faces may elicit more distinct individual-face representations. This would be consistent with a report of an aIT face-identity-change effect correlated with face familiarity (16) (see also ref. 71). Clearly memory and perception depend on each other; in fact, it is difficult to draw a bold line between them (72). Another possibility, then, is that the aIT face representation itself contains long-term memory traces. For example, the basis patterns of the representation (or the attractors of its dynamics) may correspond to known faces. Haxby *et al.* (1, 5) suggest that anterior temporal cortex contains representations of person identity, name, and biographical information (see also refs. 73 and 74).

***Is the right aIT representation face-specific or domain-general?*** The anterior temporal cortex is thought to represent complex feature conjunctions suited for fine-grained discriminations (75–77), including the discrimination between individual faces. The right aIT representation did not distinguish the houses—despite their greater visual dissimilarity (SI Figs. 9*c* and 10). It thus does not appear to be completely domain-general. Previous studies suggest that right anterior temporal cortex processes face information (22, 23, 26, 54, 57, 59). However, the region could, for example, distinguish animate objects in general. Establishing face-specificity [as has been done for the FFA (11, 28, 78, 79) (but see refs. 80 and 81)] will require testing with exemplars from a range of different categories.

## Methods

In this section, we give only an abbreviated methods description. Details on subjects, stimuli, task, design, and analysis are in *SI Text*.

**Design and fMRI Measurements.** We used a rapid event-related design with a basic trial duration of 3 s (minimal stimulus–onset asynchrony), corresponding to two functional volumes (volume acquired every 1,500 ms). Each image was presented for 400 ms. We measured 15 transversal functional slices (including early visual regions as well as the entire ventral visual stream) with a Siemens Magnetom Trio scanner (3 Tesla). Voxels were isotropic: $(2\ mm)^3$.

**Statistical Analysis.** ***Significance testing of activity-pattern effects.*** We used a standard univariate *t* test to determine whether two images elicit distinct response patterns in an ROI (Fig. 2). The *t* test is performed after projecting the data onto a multivariate discriminant dimension determined with independent data (SI Fig. 8*a*). This univariate *t* test on the multivariate discriminant constitutes a multivariate test of response-pattern difference. Compared with classical multivariate tests, this test has the advantage of not requiring the assumption of multivariate normal errors; univariate normality (as is commonly assumed in univariate fMRI analysis) suffices.

For each subject, two data sets (A and B) of the same experiment are used in the analysis. Set A is used to form a subject-specific hypothesis regarding (*i*) the precise ROI discriminating the images and (*ii*) the multivariate dimension discriminating the images. Set B is then used to test this hypothesis.

For example, to test for a face-exemplar effect in a given ROI, we first determine the response patterns elicited by the faces in data set A by using standard linear modeling. We then determine the weighting of the voxels that best discriminates the two faces in set A (face-exemplar discriminant). This weighting is closely related to the *t* map for the contrast between the two faces (but normalized by error variance instead of standard error, equivalent to a Fisher discriminant with diagonal covariance). If the contrast pattern represents an actual difference between the response patterns elicited by the two face images in the ROI, it should replicate in data set B. We therefore compute a weighted sum of the ROI time courses in data set B using the weights determined from set A. This yields a single time course (the discriminant time course), which can be subjected to a *t* test as commonly used in fMRI analysis. We use this approach to perform a fixed-effects group analysis, using prewhitening to account for temporally autocorrelated errors.

***Information estimates.*** As a measure of the amount of information a region contains about which of two images is being perceived, we estimate the mutual information between the stimulus and the multivariate response it elicits on a single trial. For the two faces, one bit of single-trial face-exemplar information (Fig. 2) would imply that a single trial's fMRI response pattern (20 s of fMRI data acquired after a single 400-ms presentation of a face image) always suffices to determine, with perfect certainty, which of the two faces was shown. We apply this measure to all pairs of the four image conditions and refer to it as the single-trial pair-wise condition information. Because fMRI measurement is noisy and limited in resolution and because our estimate depends on assumptions (see *SI Text*), the pair-wise condition information is an estimate of a lower bound on the actual information present in the region.

***Information-based mapping.*** Face-exemplar information was mapped by information-based functional brain mapping (44, 45). This method scans the imaged volume with a spherical searchlight to find regions whose response pattern distinguishes the faces (SI Fig. 8*b*). More precisely, we used a spherical searchlight of 3-mm radius to highlight 19 voxels [size: $(2\ mm)^3$] at a time. This searchlight was centered at each imaged voxel in turn, highlighting overlapping spherical sets of voxels. Using the linear model, we estimate the response patterns associated with the two faces within the searchlight and compute the Mahalanobis distance as a multivariate contrast statistic. The Mahalanobis distance is recorded in a statistical map at the voxel at the center of the searchlight. This method yields a continuous map indicating the evidence for face-exemplar information in the local neighborhood of each voxel.

For the ROI-based analysis of face-exemplar information in the FFA and aIT (SI Fig. 6 and Fig. 2), we performed a descriptive information-based mapping for each subject separately using only data set A. The resulting face-exemplar information map served to define the ROI at a given number of voxels (see SI Fig. 6 and *SI Text*). Independent data (data set B) was then used to perform statistical inference for the ROI by means of the pattern-discriminant *t* test described above.

In addition, we also performed an information-based group mapping in Talairach space (Fig. 3 and SI Fig. 9*b*) using all data. Here, we used a randomization scheme involving permutation of the condition labels for statistical inference

(details in *SI Text*). All information-based mapping analyses were performed with custom software developed in Matlab.

**ACKNOWLEDGMENTS.** We thank Alex Martin, Bill Vinje, David Leopold, Douglas Ruff, Kyle Simmons, Marieke Mur, Peter Bandettini, and Ziad Saad for helpful comments and discussions. Special thanks to Judith Peters and Harry Smit for lending their faces (stimuli). This research was funded by the Universiteit Maastricht, the Donders Centre for Cognitive Neuroimaging (Nijmegen, The Netherlands), and intramural program of the National Institute of Mental Health, and NWO (MaGW-VIDI Grant 452-04-330) (to E.F.).

1. Haxby JV, Hoffman EA, Gobbini MI (2000) *Trends Cognit Sci* 4:223–233.
2. Kanwisher N, McDermott J, Chun MM (1997) *J Neurosci* 17:4302–4311.
3. Puce A, Allison T, Gore JC, McCarthy G (1995) *J Neurophysiol* 74:1192–1199.
4. Sergent J, Ohta S, MacDonald B (1992) *Brain* 115:15–36.
5. Haxby JV, Hoffman EA, Gobbini MI (2002) *Biol Psychiatry* 51:59–67.
6. Damasio AR, Damasio H, Van Hoesen GW (1982) *Neurology* 32:331–341.
7. Hoffman EA, Haxby JV (2000) *Nat Neurosci* 3:80–84.
8. Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW (2000) *J Cognit Neurosci* 12:495–504.
9. Barton JJ, Press DZ, Keenan JP, O'Connor M (2002) *Neurology* 58:71–78.
10. Hadjikhani N, De Gelder B (2002) *Hum Brain Mapp* 16:176–182.
11. Grill-Spector K, Knouf N, Kanwisher N (2004) *Nat Neurosci* 7:555–562.
12. Andrews TJ, Ewbank MP (2004) *NeuroImage* 23:905–913.
13. Winston JS, Henson RNA, Fine-Goulden MR, Dolan RJ (2004) *J Neurophysiol* 92:1830–1839.
14. Loffler G, Yourganov G, Wilkinson F, Wilson HR (2005) *Nat Neurosci* 8:1386–1390.
15. Pourtois G, Schwartz S, Seghier ML, Lazeyras F, Vuilleumier P (2005) *J Cognit Neurosci* 17:1043–1057.
16. Rotshtein P, Henson RN, Treves A, Driver J, Dolan RJ (2005) *Nat Neurosci* 8:107–113.
17. Schiltz C, Sorger B, Caldara R, Ahmed F, Mayer E, Goebel R, Rossion B (2006) *Cereb Cortex* 16:574–586.
18. Rhodes G, Byatt G, Michie PT, Puce A (2004) *J Cognit Neurosci* 16:189–203.
19. Evans JJ, Heggs AJ, Antoun N, Hodges JR (1995) *Brain* 118 (Pt 1):1–13.
20. Tranel D, Damasio H, Damasio AR (1997) *Neuropsychologia* 35:1319–1327.
21. Moscovitch M, Winocur G, Behrmann M (1997) *J Cognit Neurosci* 9:555–604.
22. Kuskowski MA, Pardo JV (1999) *NeuroImage* 9:599–610.
23. Nakamura K, Kawashima R, Sato N, Nakamura A, Sugiura M, Kato T, Hatano K, Ito K, Fukuda H, Schormann T, Zilles K (2000) *Brain* 123 (Pt 9):1903–1912.
24. Sugiura M, Kawashima R, Nakamura K, Sato N, Nakamura A, Kato T, Hatano K, Schormann T, Zilles K, Sato K, *et al.* (2001) *NeuroImage* 13:877–890.
25. Gainotti G, Barbier A, Marra C (2003) *Brain* 126:792–803.
26. Tsukiura T, Mochizuki-Kawai H, Fujii T (2006) *NeuroImage* 30:617–626.
27. Tong F, Nakayama K, Moscovitch M, Weinrib O, Kanwisher N (2000) *Cognit Neuropsychol* 17:257–279.
28. Kanwisher N (2000) *Nat Neurosci* 3:759–763.
29. Hasselmo ME, Rolls TE, Baylis GC (1989) *Behav Brain Res* 32:203–218.
30. Perrett DI, Hietanen JK, Oram MW, Benson PJ (1992) *Philos Trans R Soc London Ser B* 335:23–30.
31. Rolls ET, Treves A, Tovee MJ, Panzeri S (1997) *J Comput Neurosci* 4:309–333.
32. Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) *Science* 293:2425–2430.
33. Cox DD, Savoy RL (2003) *NeuroImage* 19:261–270.
34. Carlson TA, Schrater P, He S (2003) *J Cognit Neurosci* 15:704–717.
35. Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X (2004) *Machine Learn* 57:145–175.
36. Kamitani Y, Tong F (2005) *Nat Neurosci* 8:679–685.
37. Haynes JD, Rees G (2005) *Nat Neurosci* 8:686–691.
38. LaConte S, Strother S, Cherkassky V, Anderson J, Hu X (2005) *NeuroImage* 26:317–329.
39. Polyn SM, Natu VS, Cohen JD, Norman KA (2005) *Science* 310:1963–1966.
40. Mourao-Miranda J, Bokde AL, Born C, Hampel H, Stetter M (2005) *NeuroImage* 28:980–995.
41. O'Toole A, Jiang F, Abdi H, Haxby JV (2005) *J Cognit Neurosci* 17:580–590.
42. Haynes JD, Rees G (2006) *Nat Rev Neurosci* 7:523–534.
43. Epstein R, Kanwisher N (1998) *Nature* 392:598–601.
44. Kriegeskorte N, Goebel R, Bandettini P (2006) *Proc Natl Acad Sci USA* 103:3863–3868.
45. Kriegeskorte N, Bandettini P (2007) *NeuroImage* 38:649–662.
46. Liu J, Harris A, Kanwisher N (2002) *Nat Neurosci* 5:910–916.
47. Grill-Spector K, Malach R (2001) *Acta Psychol (Amsterdam)* 107:293–321.
48. Tolias AS, Keliris GA, Smirnakis SM, Logothetis NK (2005) *Nat Neurosci* 8:591–593.
49. Sawamura H, Orban GA, Vogels R (2006) *Neuron* 49:307–318.
50. Leopold DA, Bondar IV, Giese MA (2006) *Nature* 442:572–575.
51. Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) *Science* 311:670–674.
52. Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RBH (2003) *Nat Neurosci* 6:989–995.
53. Heywood CA, Cowey A (1992) *Philos Trans R Soc London Ser B* 335:31–37.
54. Grabowski TJ, Damasio H, Tranel D, Ponto LL, Hichwa RD, Damasio AR (2001) *Hum Brain Mapp* 13:199–212.
55. Devlin JT, Russell RP, Davis MH, Price CJ, Wilson J, Moss HE, Matthews PM, Tyler LK (2000) *NeuroImage* 11:589–600.
56. Bellgowan PS, Bandettini PA, van Gelderen P, Martin A, Bodurka J (2006) *NeuroImage* 29:1244–1251.
57. Allison T, Puce A, Spencer DD, McCarthy G (1999) *Cereb Cortex* 9:415–430.
58. McCarthy G, Puce A, Belger A, Allison T (1999) *Cereb Cortex* 9:431–444.
59. Puce A, Allison T, McCarthy G (1999) *Cereb Cortex* 9:445–458.
60. Rossion B, Caldara R, Seghier M, Schuller A-M, Lazeyras F, Mayer E (2003) *Brain* 126:1–15.
61. Bruce V, Young A (1986) *Br J Psychol* 77:305–327.
62. Behrmann M, Avidan G (2005) *Trends Cog Sci* 9:180–187.
63. Behrmann M, Avidan G, Marotta JJ, Kimchi R (2005) *J Cognit Neurosci* 17:1130–1149.
64. Hasson U, Avidan G, Deouell LY, Bentin S, Malach R (2003) *J Cognit Neurosci* 15:419–431.
65. Avidan G, Hasson U, Malach R, Behrmann M (2005) *J Cognit Neurosci* 17:1150–1167.
66. Behrmann M, Avidan G, Gao F, Black S (2007) *Cereb Cortex* 17:2354–2363.
67. Rhodes G, Carey S, Byatt G, Proffitt F (1998) *Vision Res* 38:2307–2321.
68. Leopold DA, O'Toole AJ, Vetter T, Blanz V (2001) *Nat Neurosci* 4:89–94.
69. Gobbini MI, Leibenluft E, Santiago N, Haxby JV (2004) *NeuroImage* 22:1628–1635.
70. Gobbini MI, Haxby JV (2007) *Neuropsychologia* 45:32–41.
71. Eger E, Schweinberger SR, Dolan RJ, Henson RN (2005) *NeuroImage* 26:1128–1139.
72. Bussey TJ, Saksida LM (2005) *Curr Opin Neurobiol* 15:730–737.
73. Gorno-Tempini ML, Price CJ, Josephs O, Vandenberghe R, Cappa SF, Kapur N, Frackowiak RS (1998) *Brain* 121:2103–2118.
74. Leveroni CL, Seidenberg M, Mayer AR, Mead LA, Binder JR, Rao SM (2000) *J Neurosci* 20:878–886.
75. Tyler LK, Stamatakis EA, Bright P, Acres K, Abdallah S, Rodd JM, Moss HE (2004) *J Cognit Neurosci* 16:351–362.
76. Bright P, Moss HE, Stamatakis EA, Tyler LK (2005) *Q J Exp Psychol B* 58:361–377.
77. Moss HE, Rodd JM, Stamatakis EA, Bright P, Tyler LK (2005) *Cereb Cortex* 15:616–627.
78. Yovel G, Kanwisher N (2004) *Neuron* 44:889–898.
79. Kanwisher N, Yovel G (2006) *Philos Trans R Soc London Ser B* 361:2109–2128.
80. Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC (1999) *Nat Neurosci* 2:568–573.
81. Gauthier I, Skudlarski P, Gore JC, Anderson AW (2000) *Nat Neurosci* 3:191–197.

Kriegeskorte *et al*. 10.1073/pnas.0705654104.

# Supporting Information

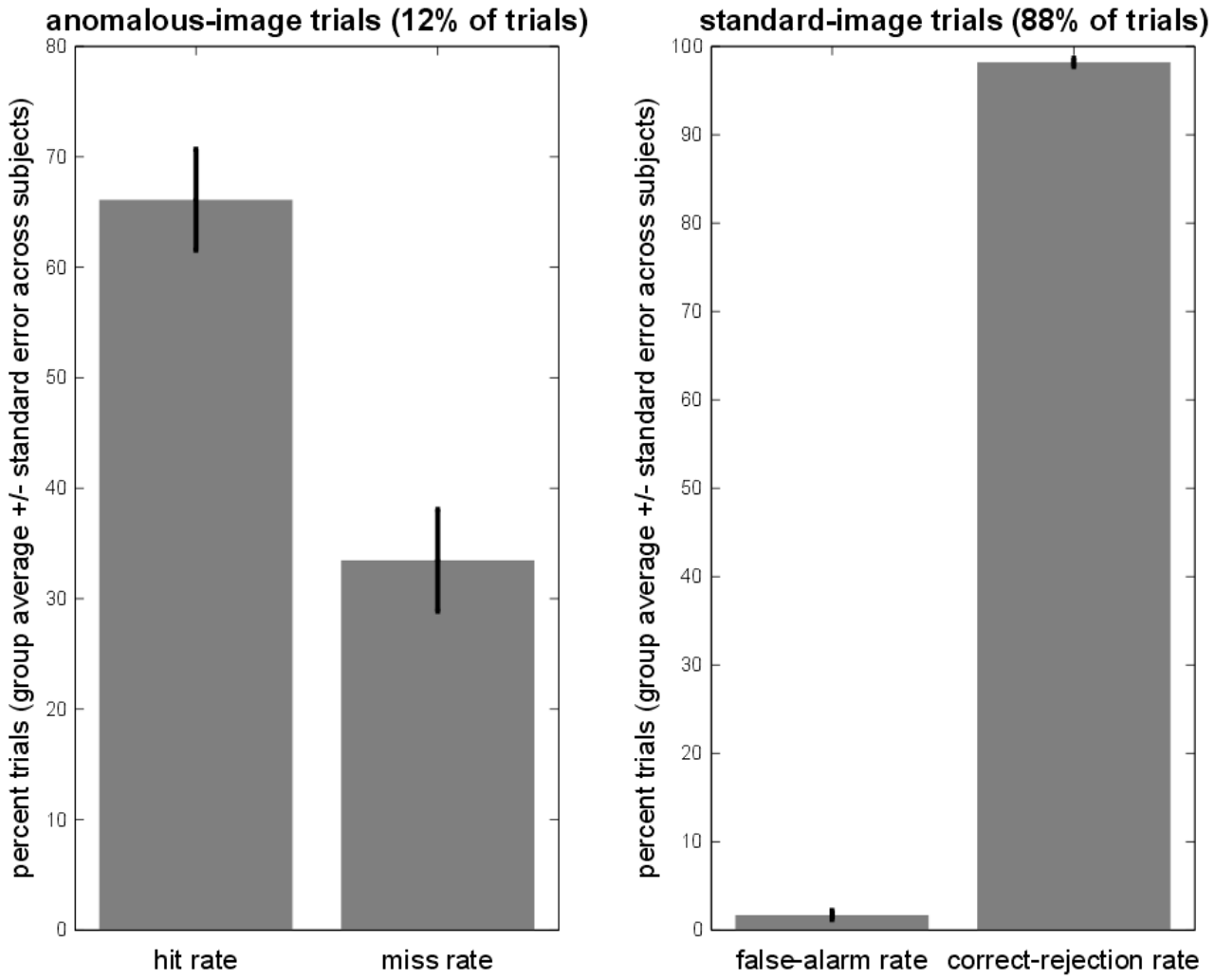


**Fig. 4. Subject error rates on anomaly-detection task during fMRI.** The anomaly-detection task (Fig. 1, *Methods*) was designed to require subjects to attend to every presentation, despite the fact that 88% of all trials consisted in repetitions of the four standard images. Subjects correctly detected approximately two-thirds (66%) of the anomalous versions of the images presented. Analysis of responses across time (data not shown) indicated that all subjects attentively viewed the stimuli throughout both runs. Bars show group-average percentages of trials with error bars indicating the standard error of the mean (computed from the standard deviation of the single-subject means).

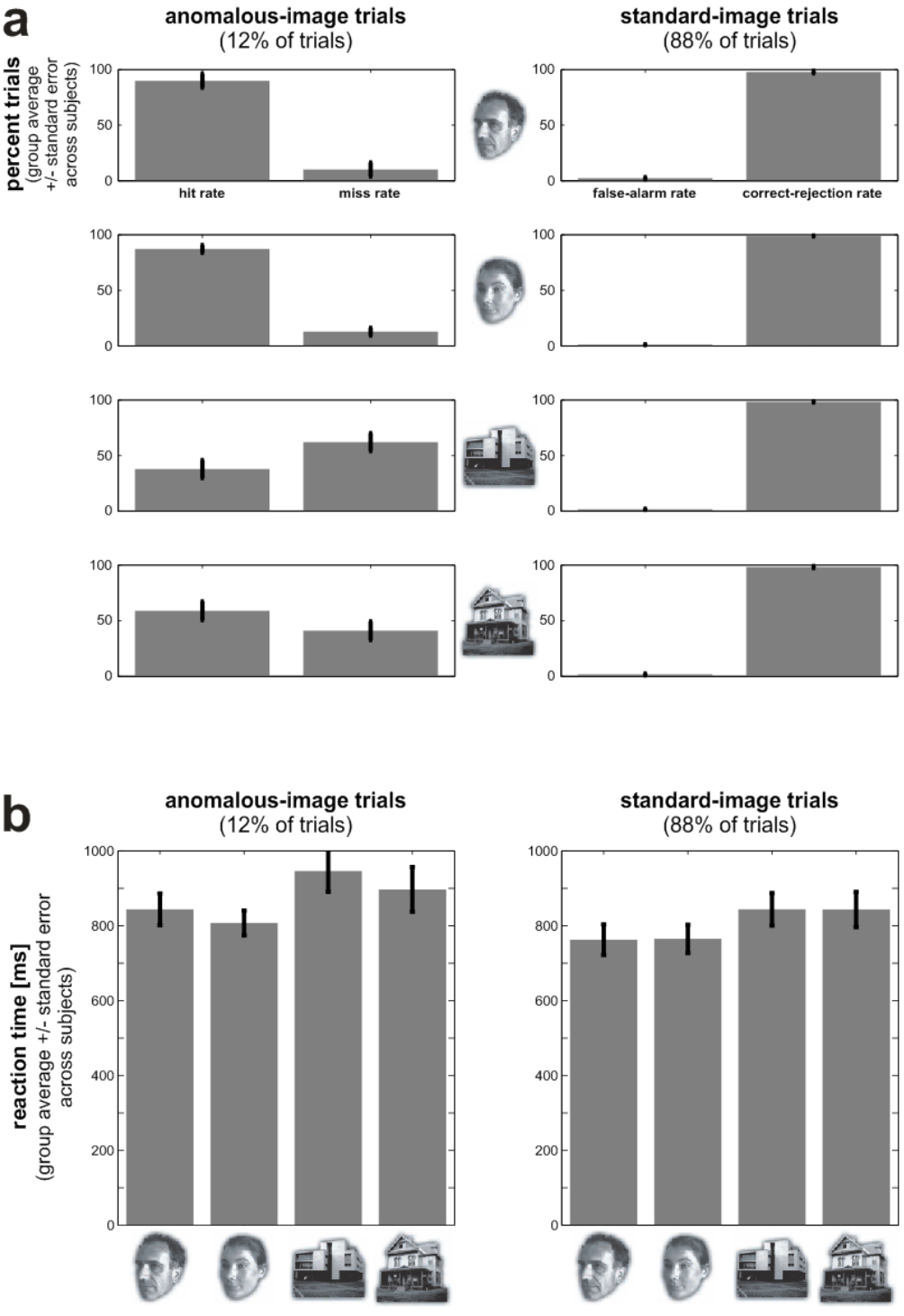


**Fig. 5. Detailed behavioral results for anomaly-detection task during fMRI (stimulus-specific reaction times and error rates).** (*a*) Anomaly-detection error analysis as in SI Fig. 10, but performed separately for each of the four standard images (right column) and its anomalous variants (left column). Subjects performed similarly on each of the four standard images. (Only standard-image trials were used for the fMRI analyses.) However, they missed more anomalies for the houses than for the faces. (*b*) Reaction-time analysis for each of the four standard images (right column) and its anomalous variants (left column). Both error rates and reaction times indicate that the anomaly detection task was slightly more challenging to subjects for the houses than for the faces. Bars show group-average percentages of trials (*a*) and reaction times (*b*) with error bars in both indicating the standard error of the mean (computed from the standard deviation of the single-subject means).

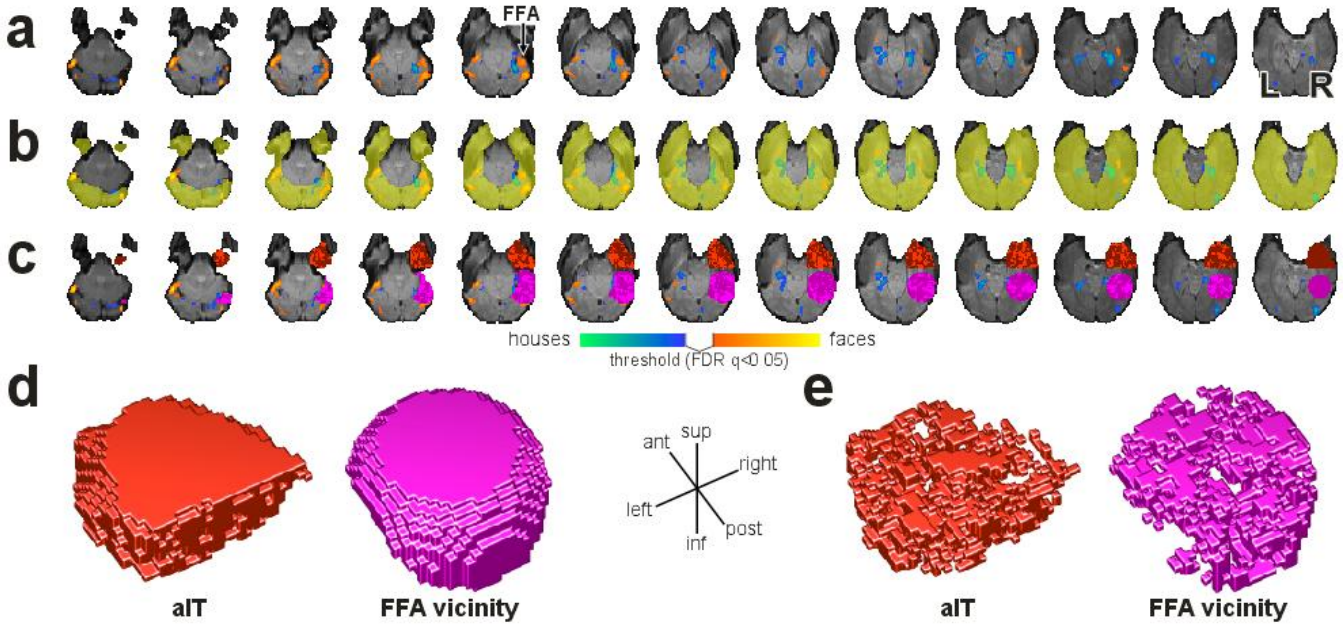


**Fig. 6. Definition of regions of interest in a single subject.** The three rows (*a*-*c*) show 13 axial brain slices as acquired (echoplanar fMRI slices averaged across time) from inferior to superior (top and bottom slices are omitted because they contain incomplete data after head-motion correction). (*a*) The univariate mapping for the contrast faces-houses reveals the fusiform face area (FFA) as indicated. The statistical map (see color bar in *c*) was thresholded to control the false-discovery rate, $q$ <0.05. For this activation analysis only, data were spatially smoothed by convolution with a Gaussian kernel of 6-mm full-width at half-maximum. (All pattern-information analyses were performed on unsmoothed data.) (*b*) A manually drawn cortex mask (transparent yellow) marks all cortical voxels in our imaging volume in each subject. (*c*) For each subject and hemisphere, the "FFA vicinity" is defined as 4,000 cortical voxels [voxel size: $(2\ \text{mm})^3$] within a sphere centered on (and including) FFA (ROI in light and dark magenta). Note that voxels within the sphere, but outside the cortex or imaging volume are not included and not counted. Analogously aIT is defined as the 4,000 most-anterior voxels in temporal cortex (ROI in light and dark red). Again only voxels within the cortex mask are included. (*d*) The 4,000-voxel ROIs from *c* for aIT (red, left) and the FFA vicinity (magenta, right). (*e*) These ROIs were analyzed for face-exemplar information (see Fig. 2) by using all 4,000 voxels and progressively reduced sets (1,400-voxel subsets shown) selected by thresholding the face-exemplar information map. The 1,400-voxel subsets are shown in light red and light magenta in *c*.

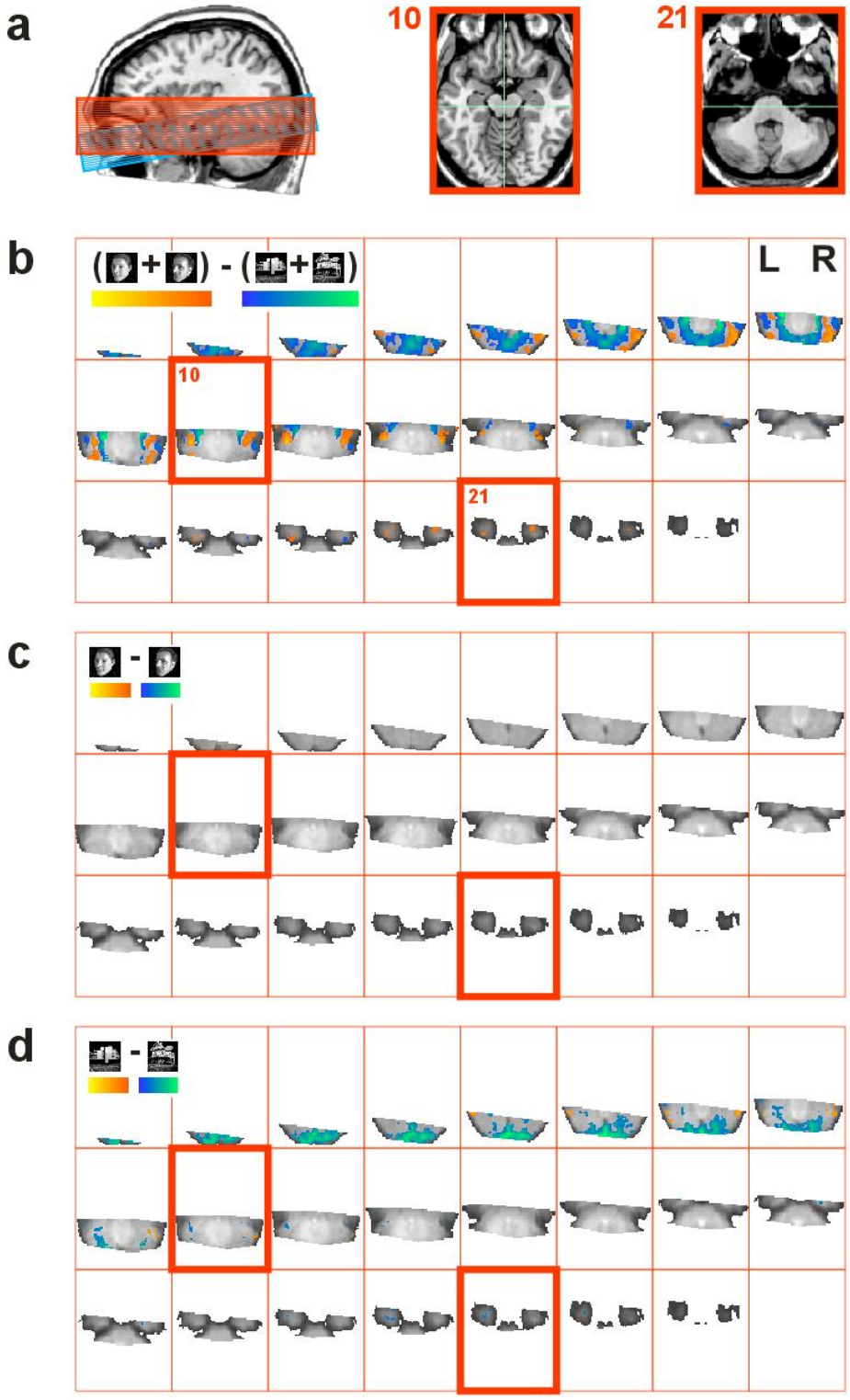


**Fig. 7. Activation-based group mapping in Talairach space.** (*a*) The occipito-temporal measurement slab (blue) and the Talairach-space slices shown in the other images (red) superimposed to a sagittal high-resolution anatomical MR image (MNI brain). Measured slices and Talairach slices are 2-mm thick with no gap. The anatomical locations of Talairach slices 10 and 21 (which contain FFA and aIT, respectively) are shown in the high-resolution anatomy (*a*, *Center* and *Right*). To indicate, where in Talairach space our measurements provided data for all subjects, we use a Talairach-space group average of the functional data as the background for the statistical maps in *b*-*d*. (*b*) Activation-based statistical map for the contrast faces versus houses. Right and left FFA appear in orange-yellow (face activation greater than house activation) in slice 10 and adjacent slices. The parahippocampal place area (PPA) appears bilaterally as well, medial to FFA (blue-green). (*c*) Activation-based statistical map for the contrast face 1 versus face 2. The two faces do not elicit different levels of overall activation in any region within our occipito-temporal imaging slab. This is plausible because the face images are physically very similar and because activation-based mapping as shown here involves smoothing out of fine-grained pattern information. (*d*) Activation-based statistical map for the contrast house 1 versus house 2. One of the house images elicits somewhat greater activity, particularly in early visual cortices. This is unsurprising because the two house images are very different physically, although they share the same category (house). (*b*-*d*) All maps in this figure show univariate fixed-effects group analyses performed on data smoothed with a kernel of 6-mm full-width at half-maximum (the kernel was a sphere of 3-mm radius). All maps are thresholded to control the false-discovery rate, $q < 0.05$. The right hemisphere is on the right side of each slice.

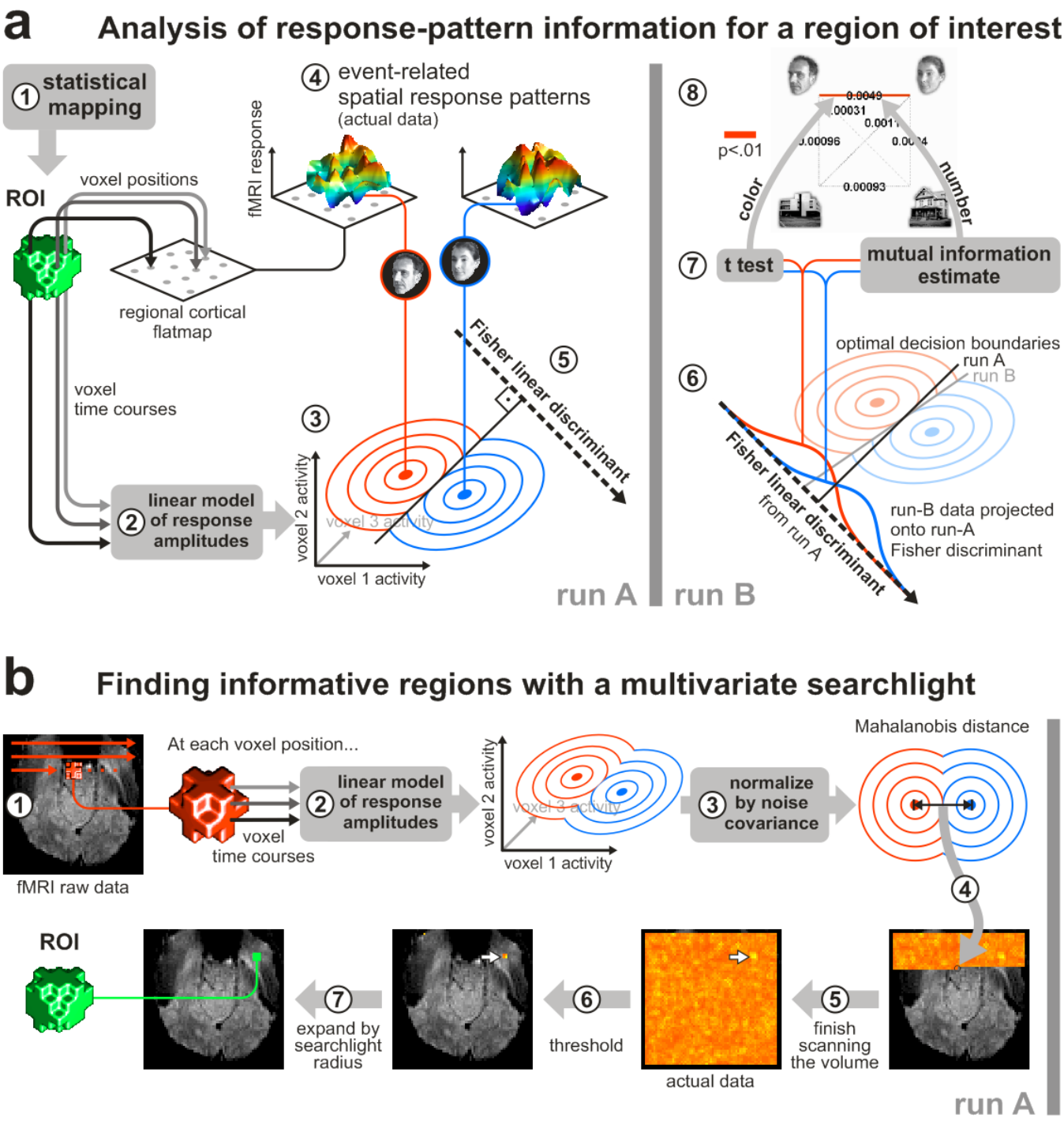


**Fig. 8. Information-based response-pattern analysis.** (*a*) For a region of interest (ROI, green voxel cluster) predefined by statistical mapping (1), a linear-model fit (2) provides an estimate of the response amplitude during each condition for each voxel of the region. For each condition, the pattern of responses across the voxels of the ROI can be thought of either as a point in the multidimensional space spanned by the voxel activities (3, red and blue central dots) or as an event-related spatial response pattern on the cortex (4). The analysis is applied to each pair of conditions in turn and is illustrated only for the face pair. For visualization of the ROI's response patterns, we compute an approximation to a regional cortical flatmap by a neighborhood-preserving self-organizing projection of the 3D voxel locations onto the unit square. The response patterns shown (4) are those found in the

anterior inferotemporal face-exemplar region of subject K. The residuals of the linear-model fit provide a multinormal model of the variability of the response-pattern estimates (3, red and blue iso-probability-density contours). Under multinormality and homoscedasticity, the optimal decision boundary for classification of response patterns is a hyperplane and the optimal discriminant dimension is the Fisher linear discriminant (5, dashed arrow), which is the dimension orthogonal to the optimal decision boundary (solid line separating red and blue distributions). All computations described thus far are performed on data from run A of each subject. We use independent data (run B) to estimate the pair-wise condition information (our effect measure, see *Methods* and *SI Text*, *Single-trial pair-wise condition information*) and to perform statistical inference (right side). The independent run-B data set is projected onto the Fisher discriminant and analyzed univariately. This projection amounts to a weighted sum across the voxels, where each voxel can have a positive or a negative weight depending on the sign of the difference in response between the two conditions. If multinormality holds and the two data sets are consistent, no information is lost by this projection. If multinormality does not hold, the analysis becomes conservative, i.e., the information estimate will be lowered and the sensitivity of statistical inference will suffer. Note that the specificity of the test, i.e., its validity, depends only on univariate normality after projection onto the Fisher discriminant. The results of analysis for all pairs of conditions are summarized in a pairwise-effects icon (see step 8 and SI Fig. 10). Group analysis (SI Fig. 10) is performed by averaging information effects and combining the t values (representing the individual response-pattern differences) across subjects. (*b*) To find regions whose response pattern distinguishes two conditions, we scan the measured volume with a 3-mm-radius spherical multivariate searchlight (1, red voxel cluster). Note that this aspect of the analysis is purely descriptive. Inference is later performed on independent data. The searchlight is centered on each voxel in turn (selecting overlapping voxel sets at adjacent positions). For each voxel position, the time courses of all voxels falling within the searchlight are subjected to joint multivariate analysis (2, 3). As a measure of response-pattern difference, we use the Mahalanobis distance. The Mahalanobis distance representing the response-pattern difference within the searchlight is recorded in a map at the central voxel position (4). The whole volume is scanned in this manner (5). Note that the resulting map represents local response-pattern information, not activation. The map shown is that of subject *K*. The white arrow marks the map maximum. The map is thresholded (6) to define the region distinguishing the two conditions. Optionally, the highlighted voxel cluster can be expanded by the searchlight radius (7), to obtain an ROI that includes all voxels that contributed to the local multivariate effects indicated by the superthreshold voxels. The location and shape of the region thus defined represents a subject-specific hypothesis, which is subsequently tested on independent data as described in *a*.

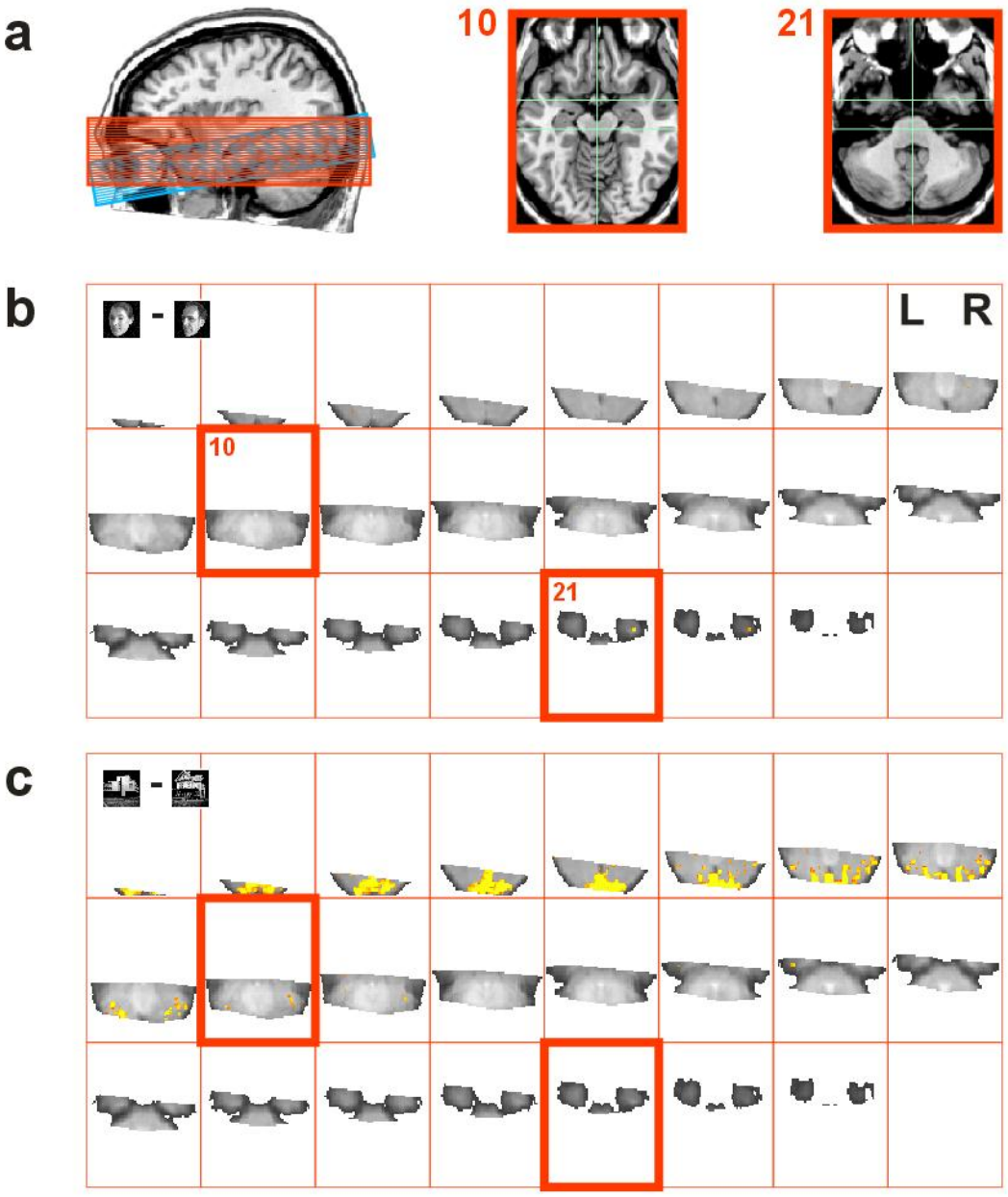


**Fig. 9. Information-based group mapping in Talairach space.** (*a*) Same as SI Fig. 7*a*, repeated for convenient reference. (*b*) Information-based group map showing regions whose local activity pattern distinguishes the two faces. This is the full information-based map shown selectively to display the aIT location in Fig. 3. Note that the only cluster is in right aIT (slices 21 and 22) and that there are a few isolated voxels in other slices as well. The highlighted aIT volume is 8 voxels = 64 $mm^3$; the volume contributing information is 56 voxels = 448 $mm^3$ (highlighted volume expanded by searchlight radius of 3 mm). Because of the effects of group-averaging and thresholding, these volumes should not be considered as estimates of the extent of the distributed code. The peak voxel had $P < 0.0001$. (*c*) Information-based group map showing regions whose local activity pattern distinguishes the two houses. Here, the information-based mapping reflects fine-grained pattern effects along with the activation-effects also seen in SI Fig. 7*d* for the house-exemplar contrast. (*b-c*) Information-based searchlight mapping is illustrated in SI Fig. 8*b*. Here, we used a randomization scheme for statistical inference (see *SI Text*: *Information-based group mapping in Talairach space*). Group maps were thresholded to highlight voxels with $P < 0.001$, uncorrected. All information-based analyses were performed on unsmoothed data. The right hemisphere is on the right side of each slice.

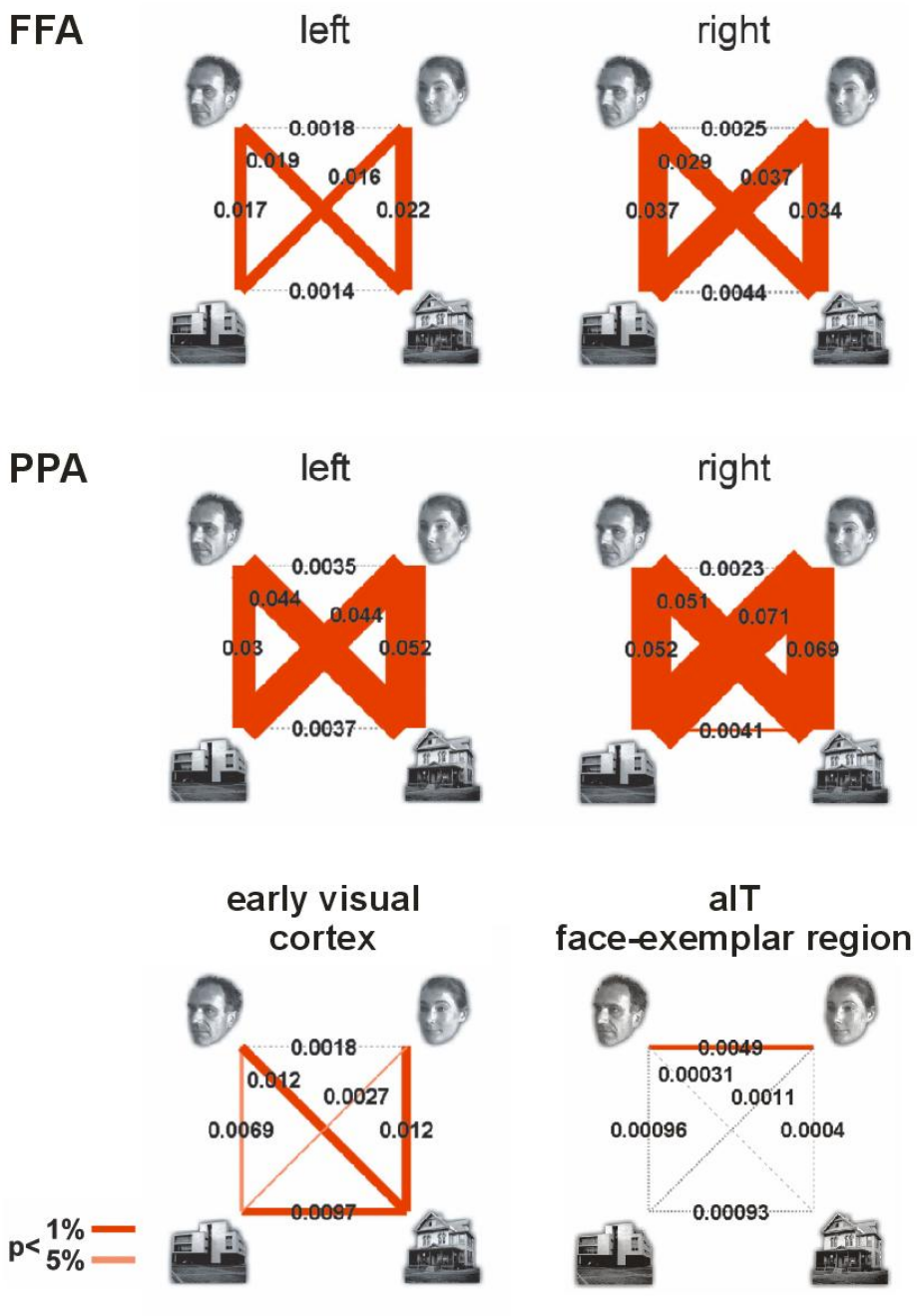


**Fig. 10. Response-pattern effects in key regions.** For each region of interest, a pair-wise-effects icon shows the multivariate effects for each pair of images. The color of each connection line indicates whether the response-pattern difference was significant for the group (red, $P < 0.01$; pink, $0.01 \leq P < 0.05$; dotted gray, $P \geq 0.05$, not significant). The thickness of each line reflects the multivariate effect size in terms of the pairwise condition information (see *SI Text*), which is also given explicitly in single-trial bits (numbers on lines). A pairwise condition information of 1 bit would indicate that the response pattern estimated from a single trial allows us to determine with perfect certainty, which of the two images has been presented. To focus the analysis on genuine combinatorial effects, the spatial-mean effect has been removed from the data before the multivariate analysis (significance testing and information estimation) by subtracting the spatial-mean time course of the region from each single-voxel time course. The analyses are fixed-effects group analyses (based on all 11 subjects) for regions of interest defined individually in each subject on the basis of mapping analyses (FFA, PPA, aIT face-exemplar region) or anatomical location (early visual cortex). Independent data were used for (1) defining the ROIs and (2) testing response-pattern effects and estimating pairwise condition information. Note the transformation of response-pattern similarity across regions: In retinotopic visual areas, all image pairs elicit distinct response patterns, except the two faces. This may reflect the greater physical similarity of the two face images. In FFA and PPA, the category distinction (faces versus houses) is emphasized, whereas within-category differences appear to be deemphasized. The IT face-exemplar region distinguishes the face images, but there are no significant effects for any other pair of images. This is consistent with the weaker overall response to houses in aIT (SI Fig. 7) and suggests that the house response patterns tend to lie in between the two distinct face response patterns in multivariate space, rendering them statistically indistinguishable from each of the face response patterns and from each other.

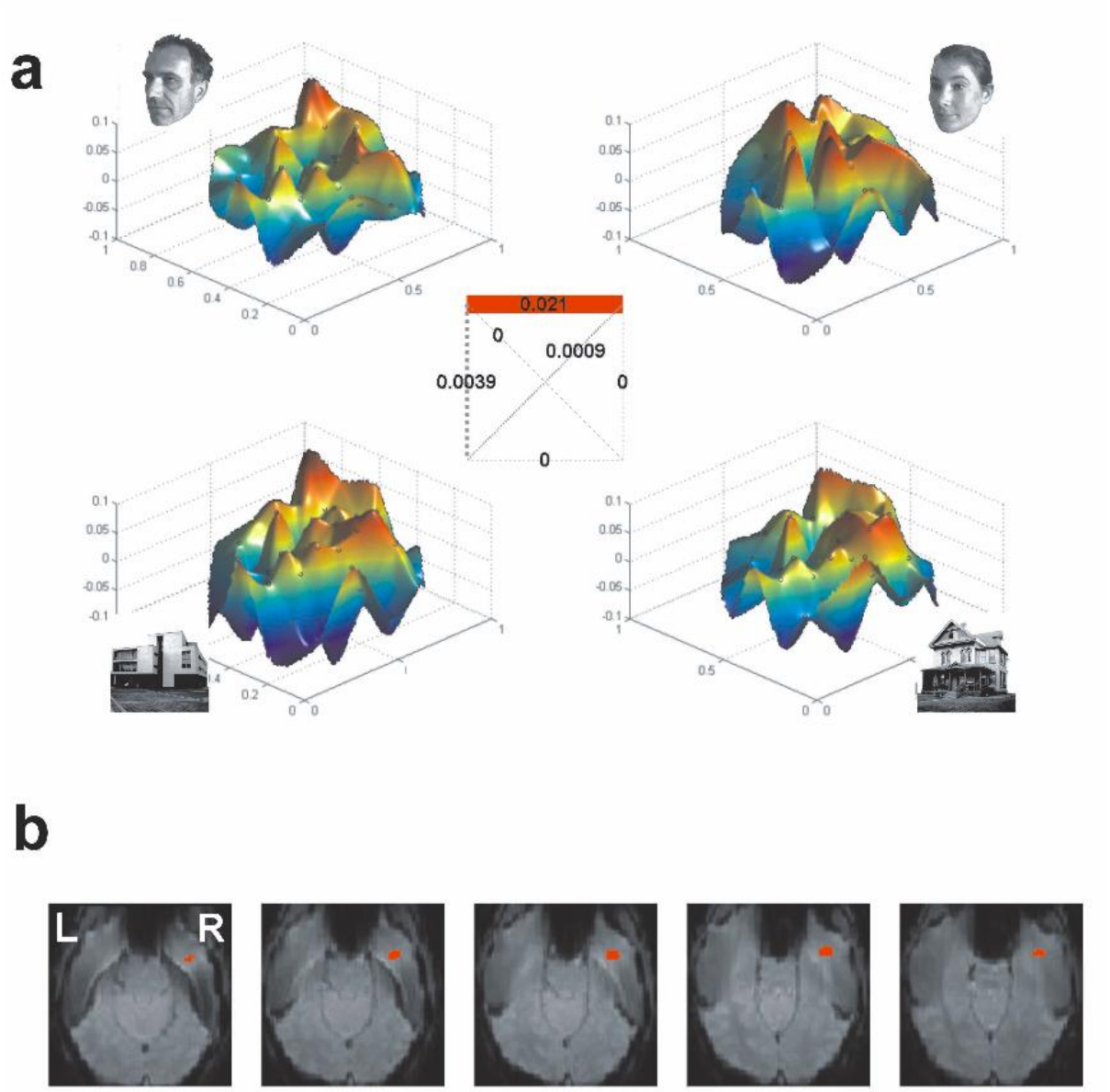


**Fig. 11. Anterior inferotemporal face-exemplar region (subject TS).** (*a*) Event-related spatial response patterns elicited by the four images in the anterior temporal face-exemplar region of subject TS. The face-exemplar effect, but none of the other pairwise effects, is significant in this subject. The two horizontal dimensions of each surface plot represent an approximate local cortical flatmap obtained by a neighborhood-preserving projection of the voxels onto the unit square. The vertical axes represent single-image beta estimates. Each voxel is represented by a little black circle and the pattern is interpolated to yield a smooth surface. For an explanation of the central pairwise-effects icon, see legend of SI Fig. 10. The region was defined based on data set A using information-based brain mapping (see *Methods*). Data set B was used (*i*) to compute the event-related spatial response patterns, (*ii*) to estimate single-trial pairwise condition information (numbers on central pairwise-effects icon; 0 for effects of inconsistent direction between data set A and data set B), and (*iii*) to test the effect (red connection indicates $P < 0.01$ in data-set-B *t* test on a linear discriminant defined by data set A). The spatial-mean effects have been removed by subtracting the spatial-mean time course of the region from each single-voxel time course before the analysis. (Omitting this step yields the same pattern of multivariate effects with negligible changes to the effect sizes.) (*b*) The anatomical location of the face-exemplar region in subject TS. The anatomical background slices were obtained by averaging the functional volumes across time. Slices progress from inferior to superior (left to right).

# SI Text

**Results of Control Analyses.** ***Information in early visual cortex.*** To investigate information in early visual cortex (EVC), we anatomically defined an ROI around the calcarine sulcus in each subject individually. Comparing the EVC response patterns for each pair of images (SI Fig. 10), we found significant response-pattern differences (multivariate fixed-effects group analysis, $P < 0.05$) for all pairs of images, except the two faces ($P > 0.05$). This reflects the physical similarity of the images (for example, spatial correlation of the face images is substantial, whereas all other pairs of images are essentially uncorrelated). Although the faces must have elicited subtly different response patterns in EVC, their retinotopic representations are too similar to be distinguished from our fMRI data. This suggests that our matching of view, lighting, and intensity histogram was successful at reducing low-level confounds to a negligible level.

***Activation effects in FFA and aIT.*** In addition to analyzing the information in the FFA and aIT response patterns (using unsmoothed single-subject data), we asked, more conventionally, what overall activation the images elicited. First, we performed an activation-based mapping (using data smoothed with a Gaussian kernel of 6-mm full-width at half-maximum) for the contrast faces-houses. This revealed face-category activation (*i*) in bilateral FFA (by definition), (*ii*) more posteriorly in bilateral regions including the lateral occipital complex and the occipital face area, and also (*iii*) more anteriorly in bilateral aIT (SI Fig. 7*b*, Talairach-space group maps). The face-category activation effects in aIT were weaker than in FFA and more posterior regions; and they were not detected in every subject (SI Fig. 6*a*, single-subject map).

Second, we performed an analysis of ROI-average activation. Independent data sets were used to (*i*) define the ROIs and (*ii*) analyze their activation effects. Our right-aIT face-exemplar region (defined by mapping for face-exemplar information) did not respond significantly more strongly to the faces than to the houses ($P > 0.05$), or vice versa ($P > 0.05$).

Note that absence of face-category activation in the right aIT face-exemplar region is not in contradiction to the distinctness of the two face response patterns: positive and negative single-voxel responses to a given face can yield an average across voxels that is close to the baseline, while the spatial response patterns are distinct.

FFA, as expected, did respond much more strongly to each of the faces than to each of the houses ($P < 0.01$). In both FFA and aIT, the two faces did not elicit significantly different overall activation ($P > 0.05$); the two houses did not elicit significantly different activation, either ($P > 0.05$).

***Adaptation effects caused by stimulus repetition.*** The design of this study is unconventional in that each stimulus image forms a separate condition. To be able to obtain stable estimates of the single-image response patterns, we present the same four images (Fig. 1*a*) many times in a pseudorandom sequence. A potential concern with such a design is that the repetitions could lead to reduced responses as a result of local neuronal adaptation or a more complex process of repetition suppression, which could include a gradual loss of attention directed at the stimuli. Although a response reduction cannot explain a positive finding (e.g., face-exemplar information in right aIT), it might explain a negative finding (e.g., the absence of a significant face-exemplar effect in FFA).

To assess whether the repeated presentation of the four images caused an overall decrease of the responses elicited, we divided each of the two fMRI runs performed with each subject into four equal temporal segments and analyzed the activation (ROI-average) elicited by each image in early visual cortex, FFA, and PPA. Results (data not shown) suggest that adaptation effects were small if they were present at all. A single-stimulus-per-condition design with only four stimuli can, thus, elicit stable responses throughout an event-related fMRI experiment. We think that our experimental task contributed to the stability of the responses across time. The anomaly-detection task (Fig. 1*b*) served to motivate subjects to attentively view of each presentation and allow us to monitor attentive viewing. Performance indicated that subjects viewed attentively throughout the experiment (SI Fig. 4).

**Details on Experimental Procedures.** ***Stimuli and task.*** The basic set of stimuli consisted of four photographs, depicting a woman's face, a man's face, a traditional house and a modern building (Fig. 1*a*). The images were in 8-bit grayscale and had a resolution of

512 × 512 pixels. Each image was processed to have a precisely uniform histogram. The images, thus, had identical light and spatial-signal energy.

Before the experiment, subjects were familiarized with the four images. They were instructed to continually fixate a central cross, which was always visible, and to perform an anomaly-detection task during the experiment (Fig. 1*b*). On 12% of the trials of each experimental run, subtle variations of the four images were presented. In each anomalous version, the global shape of the object as well as several details were slightly distorted. The particular changes were unpredictable to the subjects because several anomalous versions were used for each original. Subjects were asked to press a button placed underneath their right index finger on a regular trial and a button underneath their left index finger when they detected an anomalous image. The task served to motivate subjects to attend to each image presentation even after many repetitions and allowed us to monitor attentive viewing. Behavioral performance (SI Figs. 4 and 5) indicated that all subjects attentively viewed the stimuli throughout both runs.

***Experimental design.*** We used a rapid event-related design with a basic trial duration of 3 s (minimal stimulus-onset asynchrony) corresponding to two functional volumes of TR = 1500 ms. The stimulus sequence was optimized for estimation of the contrasts between the responses to the four original images by a method based on a genetic algorithm (1). Each image was presented for 400 ms. In each run, there were 63 presentations of each of the four original images, 33 presentations of anomalous versions of the images (see *Stimuli and task*, above), and 9 null trials, on which the image presentation was omitted and the fixation cross remained visible. The total number of 3-s time slots was, thus, $4 \times 63 + 33 + 9 = 294$, and the duration of the run including two empty time slots at the end was $(294 + 2) \times 3$ s = 14.8 min.

***Subjects.*** Eleven subjects between 18 and 30 years of age participated in the experiments (average age: 24.5 years). All had normal or corrected-to-normal vision. Five of them were female, six male. Ten of them were right-handed; one was left-handed. After receiving information about magnetic resonance imaging they gave their informed consent by signing a form. The experimental techniques used in this study and the consent form were approved by the ethical committee CWOM of the Academisch Ziekenhuis (university hospital) associated with the Katholieke Universiteit Nijmegen (The Netherlands).

***Magnetic resonance imaging.*** We measured 15 transversal functional slices with a Siemens Magnetom Trio scanner (3 Tesla) using a single-shot gradient-echo echo-planar-imaging (EPI) sequence and a standard birdcage headcoil. The imaged volume consisted in a 3-cm-thick temporal-occipital slab including early visual regions as well as the entire ventral visual stream. The pulse-sequence parameters were as follows: in-plane resolution: $2 \times 2$ mm$^2$, slice thickness: 2 mm, gap: 0 mm, slice acquisition order: interleaved, field of view (FoV): $256 \times 256$ mm$^2$, acquisition matrix: $128 \times 128$, time to repeat (TR): 1,500 ms, time to echo (TE): 32 ms, flip angle (FA): 75°. A functional run lasted 14.8 min. Each subject underwent a single imaging session including two functional runs and a high-resolution T1-weighted anatomical MPRAGE scan lasting 9.8 min (192 slices, slice thickness: 1 mm, TR: 2,300 ms, TE: 3.93, FA: 8°, FoV: $256 \times 256$ mm$^2$, matrix: $256 \times 256$). The experiments were performed at the Donders Centre for Cognitive Neuroimaging (Nijmegen, The Netherlands).

**Details on Statistical Analysis.** ***Preprocessing.*** The fMRI data sets were subjected to slice-scan-time adjustment and head-motion correction by using the BrainVoyager 2000 software package (version 4.8). (*i*) Slice-scan-time correction was performed by resampling the time courses with linear interpolation such that all voxels in a given volume represent the signal at the same point in time. (*ii*) Small head movements were automatically detected and corrected by using the anatomical contrast present in functional MR images. The Levenberg-Marquardt algorithm was used to determine translation and rotation parameters (six parameters) that minimize the sum of squares of the voxel-wise intensity differences between each volume and the first volume of the run. Each volume was then resampled in 3D space according to the optimal parameters by using trilinear interpolation.

***Design matrix and multiple linear regression.*** Single-subject analyses were performed by multiple linear regression of the response time course at each voxel. For each of the four original images, there was one predictor for the regular version and one predictor for

the anomalous versions presented. The predictor time courses were computed by using a linear model of the hemodynamic response (2) and assuming an immediate rectangular neural response during each condition of visual stimulation. For each 7.4-min subrun (see *Data splitting*, below), the design matrix included these cognitive predictors along with six head-motion-parameter time courses, a linear trend term, a six-predictor Fourier-basis for nonlinear trends (sines and cosines up to 3 cycles per subrun) and a confound-mean predictor. This design matrix was used for all univariate and multivariate analyses. Multiple linear regression was performed with custom software developed in Matlab.

***Data splitting.*** Each 14.8-min run was split into two 7.4-min subruns (first half, second half), yielding four subruns per subject. Odd subruns (when numbered chronologically as acquired) constituted data set A (used for mapping, ROI definition, and discriminant fitting). Even subruns constituted data set B (used for significance testing and information estimation). This splitting strategy is preferable to using each run as a separate data set here for three reasons: (*i*) The same stimulus sequence was used for run 1 and run 2. Set A and set B data correspond to independent stimulus subsequences. (*ii*) Each set contains an earlier and a later portion of the experimental session. (*iii*) Trend artefacts that have a similar time course across each run cannot introduce artefactual dependence between data set A and data set B.

***Definition of ROIs.*** ROIs were defined by thresholding statistical maps computed from data set A of each individual subject (SI Fig. 6). To avoid a dependence of our results on the threshold used, thresholds were varied in small steps to highlight between 10 and 4,000 voxels. All regions were restricted to a cortex mask manually defined in each subject (transparent yellow in SI Fig. 6*b*).

*Fusiform face area*. The FFA was defined in each subject and hemisphere by thresholding the *t* map for the contrast "faces minus houses." The *t* map was computed from data set A only, after spatial smoothing by convolution with a Gaussian kernel of 6-mm full-width at half-maximum. The region was defined as contiguous and seeded at the maximum of the face-house contrast map within the fusiform gyrus. (For small numbers of voxels, this definition matches what is called FFA in the literature. For large numbers of voxels, the region can extend far into anterior IT and posterior cortex. Nevertheless, it does not show evidence of face-exemplar information.)

*Anterior IT face-exemplar region.* The aIT face-exemplar region was defined in each subject and hemisphere by thresholding the face-exemplar information map obtained by using a 3-mm searchlight on data set A only (SI Fig. 8*b*). For a given number of voxels *n*, the region was defined as the discontiguous set of voxels (within the anterior 4,000 cortex-mask voxels in the hemisphere in question) with the highest entries in the data-set-A face-exemplar information map (computed by searchlight mapping on unsmoothed data).

*FFA-vicinity face-exemplar region.* To test for face-exemplar information in FFA and its vicinity in exactly the same way as in aIT, we defined the "FFA vicinity" as 4,000 cortex-mask voxels within a sphere around FFA in each subject and hemisphere. (First the center of FFA was defined as the peak of the face-house contrast map in the fusiform gyrus. Then a sphere was grown around this center. Voxels within both the sphere and the cortex mask were included in the ROI. The sphere was expanded until the ROI included exactly 4,000 voxels.) For a given number of voxels *n*, the region was then defined exactly the same way as the aIT face-exemplar region: as the discontiguous set of *n* voxels (within the 4,000-voxel FFA vicinity) that had the highest entries in the data-set-A face-exemplar information map.

***Significance testing of ROI response-pattern differences.*** *General approach and motivation.* To test in a single subject whether two stimuli elicit distinct response patterns in a given ROI, we first use data set A to formulate a subject-specific hypothesis as to the multivariate dimension and the direction of the effect. We then test this hypothesis by a single-sided *t* test performed on data set B. Forming a subject-specific hypothesis obviates the need for tests at multiple locations (multiple-comparisons problem) and allows us to apply a standard univariate *t* test, which requires fewer assumptions and affords more power than multivariate tests. As the statistical cost of these advantages, only half the data are available for the test.

*Discriminant estimation from data set A.* For a given contrast (e.g., face 1 vs. face 2) and ROI, we estimate the Fisher linear discriminant based on data set A using

the linear model described above (see *Design matrix and multiple linear regression*). The Fisher discriminant is a set of weights (one for each voxel in the ROI) defining the dimension (in the multivariate space spanned by voxel activities) that best separates two multinormal distributions of equal covariance (i.e., the dimension on which the ratio of between-class and within-class variance is maximal). The distributions here are distributions of spatial response patterns and each distribution corresponds to an experimental condition. The discriminant is defined by $w = (p_2 - p_1)\Sigma^{-1}$, where $p_1$ and $p_2$ are the two spatial response patterns, and $\Sigma$ is the error covariance matrix. We assume a diagonal covariance matrix for stability and to be able to test voxel sets larger than the number of time points (up to 4,000 voxels here). The resulting linear discriminant would be the optimally sensitive discriminant if the data were Gaussian with no dependence of errors between voxels. (The validity of the test, i.e., its specificity, is not affected by these assumptions.)

*Test on data set B.* We project the ROI time courses from data set B onto the data-set-A discriminant. This projection amounts to a weighted sum of the data-set-B time courses, yielding a single time course (discriminant time course) for a given ROI. We then perform a one-sided *t* test on the discriminant time course, with the direction of the test requiring consistency between data sets A and B. The *t* test assesses the same contrast that defines the discriminant and uses the same linear model (see above).

*Group analysis.* We perform a fixed-effects group analysis as defined in ref. 3 by concatenating the discriminant time courses of all subjects and fitting a composite design matrix with separate predictors for each subject.

*Temporal autocorrelation.* To ensure valid statistical inference in the presence of temporal autocorrelation of the errors, we apply the Cochrane-Orcutt prewhitening method (4) to the discriminant time courses and design matrix using an AR(1) model as described in ref. 5.

***Single-trial pairwise condition information.*** *General approach.* For each pair of conditions (corresponding to the four stimulus images here), we estimate a lower bound on the mutual information between the condition and the multivariate response in the ROI for a single trial. Although the experiment as a whole (comprising many trials) provides much more information, we use the mutual information for a single-trial response as a measure, because it is less dependent on accidental properties of the experiment such as the amount of data acquired and the efficiency of the design. Unlike classification accuracy, single-trial information estimates can, in principle, be compared between different experiments.

*Technical details.* The mutual information *I(S;R)* between stimulus and response is defined as follows:

$$I(S;R) = H(S) + H(R) - H(S,R) = \sum_{s,r} p(s,r) \cdot \log \frac{p(s,r)}{p(s) \cdot p(r)}$$

[**1**]

where $H(X) = -\sum_x p(x) \cdot \log p(x)$ is the entropy of a variable *X* with particular values *x*, *S* is the stimulus variable (dichotomous here; two stimuli considered at a time), *R* is the response variable (continuous here), *s* and *r* are particular values of *S* and *R*, respectively, *p(x)* is the discrete probability mass function of random variable *X*, and log is the base-2 logarithm.

We first estimate the probability distributions, then plug them into the above formula to estimate the mutual information. The stimulus variable is dichotomous (two images at a time) and uniform (each image has the same probability of occurrence). The response is continuous and multivariate with one dimension for each voxel in the region. For stability of the estimate and to be able to deal with large voxel sets (up to 4,000 voxels here), we assume the errors to be multinormal with diagonal covariance (i.e., Gaussian and independent between voxels) and equal across conditions. A multinormal response can be projected onto the Fisher linear discriminant (see above) without loss of information about the stimulus. (This is because the likelihood ratio is constant on hyperplanes orthogonal to the Fisher discriminant.) If the true population means and the true population covariance are known, the mutual information can thus equivalently be computed from the one-dimensional response distributions of the two stimuli on the Fisher discriminant. However, as the Fisher discriminant maximizes class separation, estimation of the Fisher dimension and the distributions from the same noisy data gives strongly positively biased information

estimates (overfitting). To avoid this bias, we use data set A to determine the Fisher discriminant and estimate the mutual information on the basis of the scatter and separation of the means of data set B on the data-set-A Fisher discriminant.

To estimate the distribution of the response estimates on the Fisher discriminant for single trials, we assume a design matrix $X$ with two nonoverlapping predictors, each of which describes a complete hemodynamic response (time window considered: 20 s). This provides the scaling factor that relates the standard deviation of the measurement error to the standard error of the response estimates on the Fisher discriminant for a single trial. The set-B response estimates and their standard errors on the data-set-A Fisher discriminant define two univariate normal distributions (one for each stimulus), which are plugged into Eq. **1** to obtain an estimate of the mutual information.

Because of instrumental measurement noise and limited measurement resolution in space and time and because of the assumptions involved (multinormality, independent voxel responses, no temporal pattern information), our estimate should be considered an estimate of a lower bound on the actual information carried by the region.

**Information-Based Group Mapping in Talairach Space.** The information-based group mapping in Talairach space (Fig. 3, SI Fig. 9) differed from the information-based mapping performed in single subjects for the ROI analyses in two respects: (*i*) Statistical inference was performed by using a randomization scheme (instead of the pattern-discriminant $t$ test based on splitting the data). This allowed us to use all data for the mapping. (*ii*) For computational efficiency the Mahalanobis distance was replaced by the mean squared t value (which is closely related to the Euclidean distance). The steps of the procedure are as follows:

*(i) Null-simulation design matrices from randomization of condition-labels.* For a given contrast (e.g., face 1 versus face 2) the condition labels (e.g., "face 1" and "face 2") were randomly reassigned within that set (i.e., labels were scrambled within the face and the house set, but a house trial would never receive a face label). This random relabeling was repeated 1,000 times. For each relabeling, a new design matrix with hemodynamic response predictors (2) was constructed. The resulting design matrices were spectrally similar to the actual design matrix used. (This was a concern because of the stimulus-sequence optimization, which should otherwise have been used to create the relabeling sequences as well.) Each of these design matrices was extended to include trend and head-motion components as described above (see *Design matrix and multiple linear regression*) and a confound mean predictor for each run.

*(ii) Randomization distribution of single-subject information-based maps.* Information-based mapping was performed in each subject separately for each of the 1,000 null-simulation design matrices using a 3-mm-radius spherical searchlight. To be able to perform these mappings efficiently the Mahalanobis distance (SI Fig. 8*b*) was replaced by the mean squared $t$ (MST) value within the searchlight for the contrast of interest. (The MST is closely related to the Euclidean distance: to obtain the Euclidean distance, the MST needs to be multiplied by the number of voxels entering the estimate and the square root taken.)

*(iii) Talairach-space group averaging of information-based maps.* For each subject, each of the 1,000 information-based maps was projected into Talairach space. We used BrainVoyager to define this transformation based on the T1-weighted anatomical volumes. For each of the 1,000 null-simulations, the resulting information-based maps were averaged across the 11 subjects in Talairach space, yielding 1,000 group-average null-simulation maps. The information-based maps obtained using the design matrix for the true labeling of the experimental trials were averaged in Talairach space in the same way.

*(iv) P values from voxel-specific randomization distributions.* For each voxel, we used the 1,000 values in null-simulation maps as a voxel-specific randomization distribution. The $P$ value of the voxel was estimated as the percent rank of the actual map's value at that voxel in the voxel-specific randomization distribution, divided by 100. The resulting map for face 1 versus face 2 highlights right aIT at $P < 0.001$. The voxels exceeding this threshold have actual values greater than all values in the randomization distribution. However, because there are only 1,000 values in the randomization distribution, this distribution does not allow us to estimate how much smaller than 0.001 the $P$ value is.

*(v) P values from pooling randomization distributions across voxels.* To obtain more precise *P* value estimates, we pooled randomization distributions across voxels. To account for inhomogeneities across voxels, the voxel-specific randomization distributions were first normalized: For each voxel, the mean and standard deviation of the voxel-specific randomization distribution was computed, then each value of that randomization distribution as well the actual value of the information-based map at that voxel were normalized by subtracting the mean and dividing by the standard deviation. The resulting normalized randomization distributions were combined across cortical voxels. The normalized values of the actual information-based map were converted to p values as described above but using the randomization distribution pooled across voxels.

The resulting *P* map for face 1 versus face 2 (Fig. 3, SI Fig. 9) was thresholded at $P < 0.001$. The peak voxel has $P < 0.0001$, thus surviving small-volume Bonferroni correction for 500 voxels (4,000 $mm^3$). Similar results were obtained by using a randomization distribution of null-simulation map maxima to correct for multiple comparisons.

1. Wager TD, Nichols TE (2003) *NeuroImage* 18:293-309.

2. Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) *J Neurosci* 16:4207-4221.

3. Lazar NA, Luna B, Sweeney JA, Eddy WF (2002) *NeuroImage* 16:538-550.

4. Cochrane D, Orcutt GH (1949) *J Am Stat Assoc* 44:32-61.

5. Bullmore E, Long C, Suckling J, Fadili J, Calvert G, Zelaya F, Carpenter TA, Brammer M (2001) *Hum Brain Mapp* 12:61-78.