Behavioral/Systems/Cognitive

Categorical, Yet Graded – Single-Image Activation Profiles of Human Category-Selective Cortical Regions

Marieke Mur,[1,3] Douglas A. Ruff,[1] Jerzy Bodurka,[2] Peter De Weerd,[3] Peter A. Bandettini,[1,2] and Nikolaus Kriegeskorte[1]

[1]Section on Functional Imaging Methods, Laboratory of Brain and Cognition and [2]Functional Magnetic Resonance Imaging Facility, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland, 20892 and [3]Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6200 MD Maastricht, The Netherlands

Human inferior temporal cortex contains category-selective visual regions, including the fusiform face area (FFA) and the parahippocampal place area (PPA). These regions are defined by their greater category-average activation to the preferred category (faces and places, respectively) relative to nonpreferred categories. The approach of investigating category-average activation has left unclear to what extent category selectivity holds for individual object images. Here we investigate single-image activation profiles to address (1) whether each image from the preferred category elicits greater activation than any image outside the preferred category (categorical ranking), (2) whether there are activation differences within and outside the preferred category (gradedness), and (3) whether the activation profile falls off continuously across the category boundary or exhibits a discontinuity at the boundary (category step). We used functional magnetic resonance imaging to measure the activation elicited in the FFA and PPA by each of 96 object images from a wide range of categories, including faces and places, but also humans and animals, and natural and manmade objects. Results suggest that responses in FFA and PPA exhibit almost perfect categorical ranking, are graded within and outside the preferred category, and exhibit a category step. The gradedness within the preferred category was more pronounced in FFA; the category step was more pronounced in PPA. These findings support the idea that these regions have category-specific functions, but are also consistent with a distributed object representation emphasizing categories while still distinguishing individual images.

Introduction

Human inferior temporal (hIT) cortex has been shown to contain category-selective regions that respond more strongly to object images of one specific category than to images belonging to other categories. The two most well known category-selective regions are the FFA, which responds selectively to faces (Puce et al., 1995; Kanwisher et al., 1997), and the PPA, which responds selectively to places (Epstein and Kanwisher, 1998). The category selectivity of these regions has been shown for a wide range of stimuli (Kanwisher et al., 1999; Downing et al., 2006). However, previous studies grouped stimuli into predefined natural categories and assessed only category-average activation. To investigate responses to individual stimuli, each stimulus needs to be treated as a separate condition (single-image design). Despite common use of single-image designs in monkey electrophysiology (Vogels, 1999; Földiák et al., 2004; Tsao et al., 2006; Kiani et al., 2007) and occasional use of item-specific designs in human studies in other domains (Bedny et al., 2007), single-image responses in human visual cortex have not been thoroughly investigated in object-vision functional magnetic resonance imaging (fMRI).

We measured single-image fMRI activity elicited by 96 stimuli from a wide range of object categories without assuming any predefined grouping in design or analysis. In Kriegeskorte et al. (2008), we analyzed these data for multivoxel pattern effects. We found that single-image activity patterns in hIT (including the lateral occipital complex [Malach et al., 1995], FFA and PPA) reflect natural categories: when activity patterns are grouped by their similarity, patterns elicited by images of the same category fall into the same cluster. Here, we focus on category-selective regions (rather than hIT as a whole) and on regional-average activation (rather than pattern information), thus relating the single-image approach to the earlier literature on category selectivity in human visual cortex. This enables us to investigate (1) whether each image from the preferred category elicits greater activation than any image outside the preferred category (categorical ranking), (2) whether there are activation differences within and outside the preferred category (gradedness), and (3) whether the activation profile (with stimuli ordered by the activation they elicit) falls off continuously across the category boundary or exhibits a discontinuity at the boundary (category step). We introduce a number of specialized analyses for addressing these three questions. Our analyses rely on dividing the 96-image data into two independent sets, estimating the activation profile from one dataset and then using the other dataset to test for (1) replicable inversions of rank, i.e., a member of a nonpre-

Received May 6, 2011; revised April 7, 2012; accepted May 1, 2012.

Author contributions: D.A.R., J.B., P.A.B., and N.K. designed research; M.M., D.A.R., J.B., and N.K. performed research; M.M., D.A.R., and N.K. analyzed data; M.M., P.D.W., P.A.B., and N.K. wrote the paper.

This work was supported by the Intramural Research Program of the U.S. National Institutes of Mental Health (Bethesda, Maryland) and Maastricht University (Maastricht, The Netherlands).

The authors declare no competing financial interest.

Correspondence should be addressed to either Marieke Mur or Nikolaus Kriegeskorte, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 7EF, UK. E-mail: marieke.mur@mrc-cbu.cam.ac.uk or nikolaus.kriegeskorte@mrc-cbu.cam.ac.uk.

D.A. Ruff's present address: Department of Neuroscience, University of Pittsburgh, Pittsburgh, PA 15213.

J. Bodurka's present address: Laureate Institute for Brain Research, 6655 South Yale Avenue, Tulsa, OK 74136-3326.

DOI:10.1523/JNEUROSCI.2334-11.2012

 0270-6474/12/328649-14$15.00/0

ferred category eliciting greater activation than a member of the preferred category (indicating a violation of categorical ranking); (2) replicable rankings (indicating graded responses); and (3) the necessity of a category step in modeling the falloff of activation from strongly to weakly activating stimuli.

Materials and Methods

Experiments

The fMRI experiment has been described in detail in Kriegeskorte et al. (2008). We therefore only describe the essential features here.

Subjects

Four healthy human volunteers participated in the fMRI experiment (mean age = 35 years; two females). Subjects were right-handed and had normal or corrected-to-normal vision. Before scanning, the subjects received information about the procedure of the experiment and gave their written informed consent for participating. The experiment was conducted in accordance with the Institutional Review Board of the National Institutes of Mental Health (Bethesda, Maryland).

Experimental stimuli, designs, and tasks

Ranking experiment. We used 96 colored photos of isolated objects spanning a wide range of categories, including faces and places (subset of stimuli from Kiani et al., 2007). Stimuli were presented using a rapid event-related design (stimulus duration: 300 ms, interstimulus interval: 3700 ms) while subjects performed a fixation-cross-color detection task. Stimuli were displayed at fixation on a uniform gray background at a width of 2.9° visual angle. Each of the 96 object images was presented once per run in random order. Each run included 40 randomly interleaved baseline trials where no stimulus was shown. Subjects participated in two sessions of six 9-min runs each. The sessions were acquired on separate days.

Localizer experiment. Subjects participated in an independent block-design experiment that was designed to localize regions of interest (ROIs) for the ranking analysis. The block-localizer experiment used the same fMRI sequence as the ranking experiment and a separate set of stimuli. Stimuli were grayscale photos of faces, objects, and places, displayed at a width of 5.7° of visual angle, centered with respect to a fixation cross. The photos were presented in 30 s category blocks (stimulus duration: 700 ms, interstimulus interval: 300 ms), intermixed with 20 s fixation blocks, for a total run time of ~8 min. Subjects performed a one-back repetition-detection task on the images.

fMRI

Blood oxygen level-dependent (BOLD) fMRI measurements were performed at high spatial resolution (voxel volume: $1.95 \times 1.95 \times 2\ mm^3$), using a 3 T General Electric HDx MRI scanner, and a custom-made 16-channel head coil (Nova Medical). Single-shot gradient-recalled echo-planar imaging with sensitivity encoding (matrix size: 128×96, TR: 2 s, TE: 30 ms, 272 volumes per run) was used to acquire 25 axial slices that covered IT and early visual cortex (EVC) bilaterally.

Analyses

fMRI data preprocessing

fMRI data preprocessing was performed using BrainVoyager QX 1.8 (Brain Innovation). The first three data volumes of each run were discarded to allow the fMRI signal to reach a steady state. All functional runs were subjected to slice-scan-time correction and 3D motion correction. In addition, the localizer runs were high-pass filtered in the temporal domain with a filter of two cycles per run (corresponding to a cutoff frequency of 0.004 Hz) and spatially smoothed by convolution of a Gaussian kernel of 4 mm full-width at half-maximum. Data were converted to percentage signal change. Analyses were performed in native subject space (i.e., no Talairach transformation).

Definition of ROIs

All ROIs were defined based on the independent block-localizer experiment and restricted to a cortex mask manually drawn on each subject's fMRI slices. The FFA was defined in each hemisphere as a cluster of contiguous face-selective voxels in IT cortex. These clusters were defined at five sizes, ranging from 10 to 300 voxels in each hemisphere. Clusters were obtained by selecting the peak face-selective voxel in the fusiform gyrus, and then growing the region from this seed by an iterative process. During this iterative process, the region is grown one voxel at a time, until an a priori specified number of voxels is selected. The region is grown by repeatedly adding the most face-selective voxel from the voxels that are directly adjacent to the current ROI in 3D space, i.e., from those voxels that are on the "fringe" of the current ROI (the current ROI is equivalent to the seed voxel during the first iteration). This region-growing procedure implies that each ROI is a cluster of spatially contiguous voxels, and that the larger ROIs subsume all voxels included in the smaller ROIs (same seed voxel for all ROI sizes). Face-selectivity was assessed by the contrast faces minus places and objects. The PPA was defined in an identical way but then using the contrast places minus faces and objects, growing the region from the peak place-selective voxel in the parahippocampal cortex in each hemisphere. Analyses were performed in native subject space, but for comparison to previous studies we computed the subject-average Talairach coordinates of the peak category-selective voxels used for seeding our ROIs. The coordinates (x, y, z) and their SD across subjects (in millimeters) are as follows. Left FFA: −37, −48, −13 (4, 5, 6); right FFA: 40, −50, −15 (5, 7, 2); left PPA: −15, −38, −5 (6, 5, 3); and right PPA: 23, −38, −7 (3, 5, 4). These coordinates are in the range expected based on previous literature (Kanwisher et al., 1997; Epstein and Kanwisher, 1998; Grill-Spector et al., 2004).

For control analyses, we defined the following two regions. hIT was defined by selecting the most visually responsive voxels within the IT portion of the bilateral cortex mask. It was defined at five sizes as well, ranging from 20 to 600 voxels. Visual responsiveness was assessed by the contrast visual stimulation (face, object, place) minus baseline. To ensure that hIT results would not be driven by face-selective or place-selective voxels, FFA and PPA were excluded from selection. For this purpose, FFA and PPA were defined at 150 and 200 voxels in each hemisphere, respectively. To define EVC, we selected the most visually responsive voxels, as for hIT, but within a manually defined anatomical region around the calcarine sulcus within the bilateral cortex mask. EVC was defined at the same five sizes as hIT.

Estimation of single-image activation

Single-image BOLD fMRI activation was estimated by univariate linear modeling. We concatenated the runs within a session along the temporal dimension. For each ROI, data were extracted and averaged across space. We then performed a single univariate linear model fit for each ROI to obtain a response-amplitude estimate for each of the 96 stimuli. The model included a hemodynamic-response predictor for each of the 96 stimuli. Since each stimulus occurred once in each run, each of the 96 predictors had one hemodynamic response per run and extended across all within-session runs. The predictor time courses were computed using a linear model of the hemodynamic response (Boynton et al., 1996) and assuming an instant-onset rectangular neuronal response during each condition of visual stimulation. For each run, the design matrix included these stimulus-response predictors along with six head-motion-parameter time courses, a linear-trend predictor, a six-predictor Fourier basis for nonlinear trends (sines and cosines of up to three cycles per run), and a confound-mean predictor. The resulting response-amplitude (β) estimates, one for each of the 96 stimuli, were used for the ranking analyses.

Novel analyses of single-image activation profiles

Receiver-operating characteristic. To investigate the category selectivity of single-image responses, the 96 object images were ranked by their β estimates, i.e., by the activation they elicited in each ROI. To quantify how well activation discriminated faces from nonfaces and places from nonplaces, we computed receiver operating characteristic (ROC) curves and associated areas under the curves (AUCs) for each ROI. The AUC represents the probability that a randomly chosen face (or place) is ranked before a randomly chosen nonface (or nonplace) based on the activation elicited by these two images. In other words, the AUC is a threshold-independent measure of discriminability. Taking faces as an

example, an AUC of 0.5 indicates chance performance at discriminating faces from nonfaces. An AUC of 1 indicates perfect discriminability, i.e., each face is ranked before each nonface. An AUC of 0 indicates perfect discriminability as well, but based on the opposite response pattern, i.e., each nonface is ranked before each face. To determine whether discrimination performance was significantly different from chance, we used a two-sided label-randomization test on the AUC (10,000 randomizations). *p* values were corrected for multiple comparisons using Bonferroni correction based on the number of ROI sizes tested per region. For group analysis, we averaged the activation profiles across sessions and subjects, and performed the ranking and AUC test on the subject-average activation profile (see Figs. 1, 2).

Proportion of replicated inverted pairs. We expect category-selective regions to discriminate preferred images (i.e., images from the preferred category) from nonpreferred images (i.e., images from other, nonpreferred, categories) significantly above chance. However, taking FFA as an example, even if each face elicits greater regional-average activation than any nonface, we still expect the AUC to be smaller than 1 because of the noise in the data. We therefore need a separate test for violation of category-consistent ranking. If there are indeed nonfaces that consistently activate FFA more strongly than faces, these inverted pairs (i.e., nonpreferred image ranked before preferred image) should replicate. We used the proportion of replicated inverted pairs (PRIP) from one session to the next as our test statistic. We computed the PRIP for each subject by dividing the number of inverted pairs that replicated from session 1 to session 2 by the total number of inverted pairs in session 1. A PRIP of 1 indicates that all inverted pairs replicated from one session to the next (perfect replicability). A PRIP of 0 indicates that none of the inverted pairs replicated from one session to the next (zero replicability). In other words, all inverted pairs reverted to category-preferential order (i.e., preferred image ranked before nonpreferred image). A PRIP of 0.5 indicates that half of the inverted pairs replicated from one session to the next. This is the level that we expect under the null hypothesis that the apparently inverted pairs actually have equal activation (the probability of inversion due to noise is ~0.5 for these image pairs). We used a two-sided label-randomization test (10,000 randomizations) to determine whether the PRIP differed significantly from 0.5. A PRIP significantly larger than 0.5 indicates that most inverted pairs replicate, suggesting the presence of true inversions and therefore a violation of category-consistent ranking. A PRIP significantly smaller than 0.5 indicates that most inverted pairs revert to category-preferential order, suggesting that most of the inversions observed in a single session were due to noise. In other words, the evidence points in the direction of category-consistent ranking. *p* values were corrected for multiple comparisons using Bonferroni correction based on the number of ROI sizes tested per region. For group analysis, we used the subject-average PRIP as our test statistic (see Fig. 3). We performed statistical inference using a simulated null distribution of subject-average PRIPs obtained by randomization of the condition labels. Note that this procedure allows the particular image pairs inverted to differ across subjects.

Replicability of largest-gap inverted pairs. The test of the proportion of replicated inverted pairs has the power to demonstrate that most inversions either replicate or revert to category-preferential order. However, this test is not appropriate for detecting a small number of true inverted pairs among many apparent inversions caused by noise. For example, 10 highly replicable inversions would almost certainly go undetected if they were hidden among a hundred pairs inverted by noise in one session's data. Given the gradedness of responses within and outside the preferred category (see Figs. 1, 5, 6), it is plausible that many stimuli near the category boundary might be inverted by noise. We therefore devised an alternative test for preference inversions, which focuses on the most egregious inversions, i.e., those associated with the largest activation gap between the stimuli from the nonpreferred and the preferred category. We can use the activation estimates of session 1 to find the largest-gap inverted pair. In this pair of stimuli, the stimulus from the nonpreferred category exhibits the largest dominance over the stimulus from the preferred category. If noise equally affects all stimuli (a reasonable assumption here, because all stimuli were repeated an equal number of times and fMRI time series are widely assumed to be homoscedastic), then this inverted pair is least likely to be spurious. This motivates us to test whether the inversion replicates in session 2. However, since this is a single pair of stimuli, we have very limited power for demonstrating the replicated inversion. To test for a small proportion of true inverted pairs, it is more promising to combine the evidence across multiple pairs. However, if we include too many pairs, we might lose power by swamping the truly inverted pairs in spurious inversions caused by noise. We therefore consider, first, the largest-gap inverted pair, then the two largest-gap inverted pairs and so on, up to the inclusion of all inverted pairs. Each of these replication tests subsumes the inverted pairs of all previous tests, thus the tests are highly statistically dependent. The loss of power due to the necessary adjustment for multiple testing might therefore not be severe if the dependency is appropriately modeled.

For $k = 1 \ldots n$, where n is the number of session 1 inverted pairs, we find the k largest-gap inverted pairs in the session 1 activation profile, estimate the activation gaps for these pairs from the session 2 activation profile, and average the gaps. This provides the average replicated gap as a function of k (ARG(k)). We also compute the SE of the estimate of the ARG from the SEs of the activation estimates of session 2 and take the repeated use of the same stimuli in multiple pairs into account in combining the SEs of the estimates. To stabilize the estimates, we compute the ARG statistic and its SE also with reverse assignment of the two sessions (session 2 for finding and ranking the inverted pairs and session 1 for estimating the ARG). For each k, the two ARG statistics and their SEs are averaged. Note that the two directions are not statistically independent and that averaging the SEs does not assume such independence, yielding a somewhat conservative estimate of the SE. Note also that one of the sessions will typically exhibit a larger number of inverted pairs. The number of inverted pairs considered in the average across the two directions is therefore the lower one of the two sessions' numbers of inverted pairs.

If ARG(k) is significantly positive for any value of k (accounting for the multiple tests), then we have evidence for replicated inversions. To test for a positive peak of ARG(k), we perform a Monte Carlo simulation. The null hypothesis is that there are no true inversions. Our null simulation needs to consider the worst-case null scenario, i.e., the one most easily confused with the presence of true inverted pairs. The worst-case null scenario most likely to yield high ARGs is the case where the inverted pairs all result by chance from responses that are actually equal. (If inverted pairs result from responses that are actually category-preferential with a substantial activation difference, these are less likely to replicate.) We estimate the set of inverted pairs using session 1 data. We then simulate the worst-case null scenario that the stimuli involved all actually elicit equal responses. For each stimulus, we then use the SE estimates from the session 2 data to set the width of a 0-mean normal distribution for the activation elicited by that stimulus. We then draw a simulated activation profile and compute the ARG(k). We repeat this simulation using sessions 1 and 2 in reversed roles and average the ARG(k) across the two directions as explained above. We then determine the peak of the simulated average ARG(k) function. This Monte Carlo simulation of the ARG(k) is based on reasonable assumptions, namely normality and independence of single-stimulus activation estimates. It accounts for all dependencies arising from the repeated appearance of the same stimuli in multiple pairs and from the averaging of partially redundant sets of pairs for different values of k. For each ROI, this Monte Carlo simulation was run 1000 times, so as to obtain a null distribution of peaks of ARG(k). Top percentiles 1 and 5 of the null distribution of the ARG(k) peaks provide significance thresholds for $p < 0.01$ and $p < 0.05$, respectively. We performed two variants of this analysis that differed in the way the data were combined across subjects. In the first variant (see Fig. 4), we performed our ARG analysis on the group-average activation profile. This variant is most sensitive to preference inversions that are consistent across subjects. In the second variant, we computed ARG(k) and its SE independently in each subject. We then averaged the ARG across subjects for each k, and computed the SE of the subject-average ARG for each k. The number of inverted pairs considered in the average across subjects was the lowest one of the four subjects' numbers of inverted pairs. Inference on the subject-average ARG(k) peak was performed using Monte Carlo simulation as described above, but now averaging across subjects was performed at the level of ARG(k) instead of at the level of the activa-

tion profiles. This second variant is sensitive to subject-unique preference inversions.

Replicability of within-category activation profiles. Do images of a region's preferred category all activate the region equally strongly or do some of them activate the region more strongly than others? To address this question, we tested whether within-category ranking order replicated across sessions. If all images of one specific category would activate a region equally strongly (i.e., flat within-category activation profile), we would expect their ranking order to be random and therefore not replicable across sessions. If, however, some images of a specific category would consistently activate the region more strongly than other images of the same category (i.e., graded within-category activation profile), we would expect the ranking order of these images to replicate across sessions. We assessed replicability of within-category activation profiles by computing Spearman's rank correlation coefficient (Spearman's r) between activation estimates for one specific category of images in session 1, and activation estimates for the same subset of images in session 2. We performed a one-sided test to determine whether Spearman's r was significantly larger than zero, i.e., whether replicability of within-category activation profiles was significantly higher than expected by chance. p values were corrected for multiple comparisons using Bonferroni correction based on the number of ROI sizes tested per region. For group analysis, we combined single-subject data separately for each session, and then performed the across-session replicability test on the combined data (see Fig. 5). We used two approaches for combining the single-subject data. The first approach consisted in concatenating the session-specific within-category activation profiles across subjects, the second in averaging them across subjects. The concatenation approach is sensitive to replicable within-category ranking across sessions even if ranking order would differ across subjects. The averaging approach is sensitive to replicable within-category ranking that is consistent across subjects.

Joint falloff model for category step and within-category gradedness. If the activation profile is graded within a region's preferred category and also outside of that category, the question arises whether the category boundary has a special status at all. Alternatively, the falloff could be continuously graded across the boundary without a step. A simple test of higher category-average activation for the preferred category cannot rule out a graded falloff without a step. To test for a step-like drop in activation across the category boundary requires a joint falloff model for gradedness and category step.

To fit such a falloff model, we first need to have a ranking of the stimuli within and outside the preferred category. We therefore order the stimuli by category (preferred before nonpreferred) and by activation within preferred and within nonpreferred. Note that inspecting the noisy activation profile after ranking according to the same profile (see Figs. 1, 2) cannot address either the question of gradedness or the question of a category step. Gradedness cannot be inferred because the profile will monotonically decrease by definition: the inevitable noise would create the appearance of gradedness even if the true activations were equal for all stimuli. Similarly, a category step might be obscured in a ranked activation profile, because ranking the noisy activation estimates will artifactually smooth the transition. After obtaining a ranking hypothesis from a given dataset, we therefore need independent data to test for gradedness and for a category step. We use session 1 to obtain the ranking hypothesis. We then apply the order (preferred before nonpreferred, and ordered according to session 1 within preferred and within nonpreferred) to the activation profile estimated from session 2 and fit the falloff model to the session 2 activations.

We use a simple linear falloff model with four predictors (see Fig. 6*A*). The predictors are (1) a linear-ramp predictor for the preferred category, which ranges from 1 (at the most activating preferred stimulus) to 0 (at the category boundary) and is constant at 0 within the nonpreferred category; (2) a linear-ramp predictor for the nonpreferred category, which is constant at 0 within the preferred category and ranges from 0 (at the category boundary) to −1 (at the least activating nonpreferred stimulus); (3) a confound-mean predictor spanning all stimuli (1 for all stimuli); and (4) a category-step predictor (1 for the preferred and −1 for the nonpreferred category). The estimated parameters of this linear model reflect the gradedness within (predictor 1) and outside (predictor 2) the preferred category, the activation average between the two categories (predictor 3), and the size of the category step (predictor 4), i.e., the drop-off at the category boundary that is not explained by the piecewise linear gradation within and outside the preferred category.

To improve the estimates, we perform the same model fitting in reverse (using session 2 to obtain the ranking hypothesis and session 1 to fit the model) and average the estimated parameters across both directions. Note that the two directions do not provide fully independent estimates; we do not assume such independence for inference. Statistical inference is performed by bootstrap resampling of the stimulus set (10,000 resamplings). The motivation for bootstrap resampling the stimuli is to simulate the variability of the estimates across samples of stimuli that could have been used. Our conclusions should be robust to the particular choice of exemplars from each category. We therefore view our stimuli as a random sample from a hypothetical population of stimuli that might equally well have been used. Repeating the analysis (ranking with each session's data, fitting the model to the other session's data, and averaging the gradation- and step-parameter estimates across the two directions) for each bootstrap resampling, provides a distribution of fits (shown transparently overplotted in gray in Fig. 6*B*) and parameter estimates, from which we compute confidence intervals and p values (one-sided test).

We performed two variants of this analysis that differed in the way the data were combined across subjects. In the first variant (see Fig. 6*B*), we averaged the activation profiles across subjects to obtain a group-average activation profile for ranking (based on one session) and for fitting the falloff model (based on the other session). This analysis is most sensitive to activation profiles that are consistent across subjects. In the second variant (results not shown, but described below), we fitted the falloff model independently for each subject and averaged the parameter estimates across subjects. This analysis is sensitive to subject-unique activation profiles (where different particular images may evoke higher activation in each subject).

Results

Good discriminability of object category at the single-image level

To visualize the degree of category selectivity for single images, we ranked the 96 object images by the activation they elicited in each ROI (Figs. 1, 2). Visual inspection of the ranking results indicates that category-selective regions FFA and PPA show a clear preference for images of their preferred category: activation of PPA ranks (almost) all places before all nonplaces and activation of FFA ranks most faces before most nonfaces (Fig. 1). Control regions hIT and EVC do not show a clear category preference at first inspection (Fig. 2). To quantify these results, we computed ROCs and AUCs for each ROI. Consistent with visual inspection, single-image activation of FFA showed very good discrimination of faces from nonfaces, with right FFA (AUC = 0.94) showing better performance than left FFA (AUC = 0.82). Single-image activation of PPA showed (near) perfect discrimination of places from nonplaces (AUC = 1). A two-sided condition-label randomization test on the AUCs indicated that discrimination performance of FFA and PPA for their preferred category was significantly above chance (Fig. 1; $p < 0.001$ for each region). In addition, discrimination performance of FFA and PPA for the "opposite", nonpreferred, category (i.e., places for FFA and faces for PPA) was significantly below chance (Fig. 1; $p < 0.05$ for FFA, $p < 0.001$ for PPA). In other words, activation of FFA ranked most nonplaces before most places and activation of PPA ranked most nonfaces before most faces. Could this finding simply be due to FFA's strong activation to faces (which were among the nonplaces) and PPA's strong activation to places (which were among the nonfaces)? If so, removing the faces from the nonplaces (FFA) and the places from the nonfaces (PPA) should abolish the effect. This was indeed the case for FFA, but not for PPA, indicating that PPA responds more weakly to faces than to

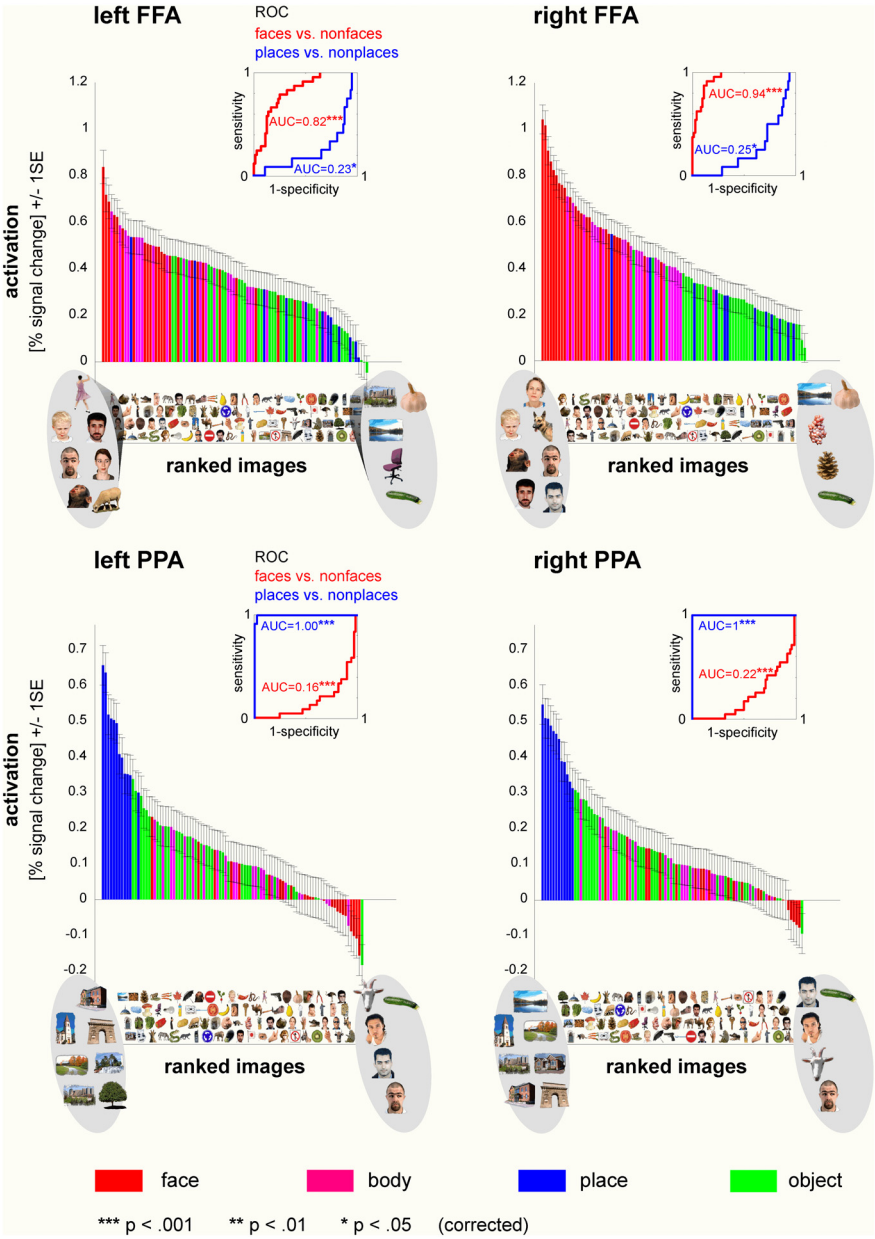


Figure 1. Single-image activation of FFA and PPA discriminates preferred from nonpreferred images. The graphs show the 96 object images ranked by the activation they elicited in each ROI. Each bar represents activation to one of the 96 object images in percent signal change averaged across four subjects. Each image is placed exactly below the bar that reflects its activation, so that the images are ordered from left to right (i.e., only the *x*-coordinate is meaningful). The leftmost image activated the region most strongly, the rightmost image activated the region most weakly. The highest- and lowest-ranked images are enlarged to give a first impression of the region's response preference. The bars are color-coded for category to give an overall impression of category selectivity without having to inspect all single images. Insets show ROC curves and associated AUCs, indicating performance for discriminating faces from nonfaces (red) and places from nonplaces (blue). We used a two-sided label-randomization test to determine whether discrimination performance was significantly different from chance (H0: AUC = 0.5). Since we tested discrimination performance at five different ROI sizes for each region, we corrected *p* values for multiple (five) comparisons using Bonferroni correction. Error bars indicate SE of the activation estimates, averaged across four subjects. FFA and PPA were each defined at 128 voxels in each hemisphere, based on an independent block-localizer experiment. Note that the smooth falloff is a necessary consequence of the rank ordering of the activation profile. Therefore further analyses are required to test for preference inversions (Figs. 3, 4), gradedness (Figs. 5, 6), and a categorical step (Fig. 6).

nonfaces even if there are no places among the nonfaces. Discrimination performance of hIT was not significantly different from chance for either category. EVC showed above-chance performance for places (Fig. 2; AUC = 0.74, $p < 0.05$) but not for faces. This suggests that place images differ to some extent from nonplace images in terms of their lower level visual properties (Rajimehr et al., 2011).

No evidence for preference inversions in PPA and right FFA

Figure 1 indicates that, despite the clear preference of FFA and PPA for images of their preferred category, some nonpreferred images appear before some preferred images in this descriptive analysis. This can be seen most clearly for FFA: some nonface images activated FFA more strongly than some face images. To test whether high-ranked nonpreferred images consistently activated the category-selective regions more strongly than lower-ranked preferred images, we computed the PRIP (Fig. 3; see Materials and Methods). The PRIP gives an indication of the rate at which inverted pairs (i.e., nonpreferred image ranked before preferred image) replicate from one session to the next. Statistical inference was performed using a two-sided label-randomization test on the PRIP. We would expect to find a PRIP of ~0.5 under the null hypothesis that the apparently inverted pairs actually have equal activation. Results show that the PRIP for both FFA and PPA was significantly <0.5 for almost all ROI sizes (Fig. 3*B*), indicating that inverted pairs had a significant tendency to revert to category-preferential order from one session to the next. In other words, the evidence points in the direction of category-consistent ranking.

The fact that most inverted pairs did not replicate does not eliminate the possibility that there is a small number of true inverted pairs. We therefore performed another analysis focused on the replicability of those inverted pairs that showed the largest activation difference between the two images (the "largest-gap inverted pairs"). Assuming that noise equally affects all images, the largest-gap inverted pairs are the most likely candidates for true inversions. Our test of replicability of largest-gap inverted pairs (Fig. 4; see Materials and Methods), performed on the subject-average activation profile, showed no evidence for replicated inverted pairs in either FFA or PPA at any ROI size (Fig. 4*B*; smallest two ROI sizes not shown). Statistical inference was performed using a Monte Carlo simulation of the null hypothesis of no true inverted pairs. We additionally performed a modified version of our largest-gap inverted-pairs analysis, which is sensitive to subject-unique preference inversions. Results for this analysis did not differ from the results shown in Figure 4*B*, except that left FFA now showed replicated inverted pairs at one of the five ROI sizes (23 voxels,

$p < 0.01$). This suggests some evidence for the presence of truly inverted pairs in left FFA at the individual subject level. We therefore subsequently performed the largest-gap inverted-pairs analysis on the subject-average activation profile again, but this time with optimal linear weighting of the activation estimates for the four subjects. Each single-image activation in each subject was given the weight $1/SE^2$, where SE is the standard error of the estimate. This weighting yields the minimum-variance weighted average for the group. For PPA and right FFA, no inversions were detected consistent with the analysis shown in Figure 4*B*, where activation profiles were averaged across subjects with equal weights. However, for left FFA (defined at 55 or 128 voxels), we found evidence for replicated inverted pairs. In sum, our findings are consistent with the idea that right FFA will prefer any face over any nonface (in terms of its regional-average activation), and that left and right PPA will similarly prefer any place over any nonplace. Only for left FFA was there some evidence for preference inversions for particular images.

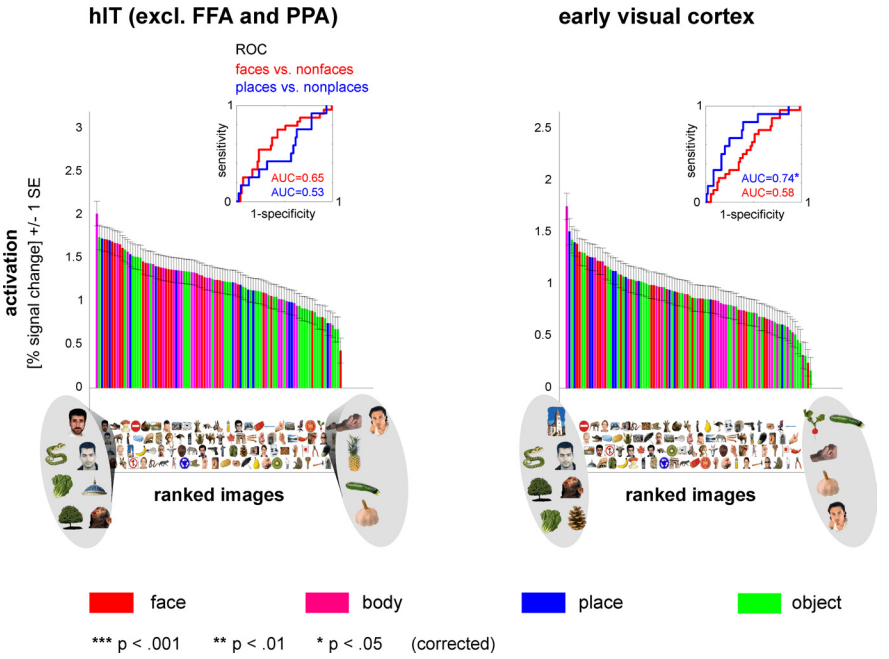


Figure 2. Single-image activation of hIT and EVC does not show a strong category preference. As in Figure 1, images are ranked by the activation they elicited in each ROI. Insets show discrimination performance. Statistical tests as described in Figure 1. hIT and EVC were defined bilaterally at 256 voxels each, based on visual responsiveness during an independent block-localizer experiment.

Control regions hIT and EVC do not have a strong category preference. However, for completeness, we performed the same analyses for these regions and found that their PRIP values were not significantly different from chance (Fig. 3*B*). Our more sensitive largest-gap inverted-pairs test showed evidence for a small number of replicated inverted pairs in both hIT and EVC at all (EVC) or most (hIT) ROI sizes (Fig. 4*B*; smallest two ROI sizes and hIT not shown). The evidence for face–nonface inversions remained present in the subject-unique group analysis, but the evidence for place–nonplace inversions largely disappeared (it only remained present in EVC at 46 voxels). Overall, these results are consistent with our expectation that particular images drive these regions to slightly different degrees, but the preferences do not conform to the category definitions.

Category discriminability and preference inversions across ROI sizes

We performed our analyses of category discriminability and preference inversions for five different ROI sizes, ranging from 10 to 300 voxels for unilateral FFA and PPA and from 20 to 600 voxels for bilateral hIT and EVC. Testing across multiple ROI sizes enables assessment of the robustness of our effects against changes in ROI size. Figure 1 shows discrimination performance (AUC) for an intermediate ROI size (128 voxels) chosen to approximately match the previously reported volume of right FFA (Kanwisher et al., 1997). Discrimination performance for the other ROI sizes can be found in Table 1.

The top panel of Table 1 shows very good discrimination of faces from nonfaces based on single-image activation of both left and right FFA at all ROI sizes. Discrimination performance was best for the smallest ROI size (10 most face-selective voxels) and decreased with increasing ROI size. The effect of ROI size was more pronounced for left than right FFA, resulting in a considerable difference in performance between left and right FFA for the two largest ROI sizes. The bottom panel of Table 1 shows near-perfect discrimination of places from nonplaces based on single-image activation of both left and right PPA at all ROI sizes. Discrimination performance of right PPA was not influenced by ROI size; performance of left PPA was a bit lower for the two smallest ROI sizes. It should be noted that hIT showed above-chance performance for discriminating faces from nonfaces at small ROI sizes ($0.71 < AUC < 0.72$, $p < 0.01$), which can be attributed to the inclusion of some weakly face-selective voxels in a subset of the subjects. Furthermore, the above-chance performance of EVC for discriminating places from nonplaces reported in Figure 1, where EVC was defined at 256 voxels, was only marginally significant for the other four ROI sizes ($0.71 < AUC < 0.73$, $p < 0.10$). With respect to preference inversions, Figure 3*B* indicates that right FFA and PPA showed PRIP effects for their preferred category at almost all ROI sizes. Left FFA and PPA showed PRIP effects at most ROI sizes, with stronger effects at smaller ROI sizes for FFA and larger ROI sizes for PPA. ROI size did not have a noticeable effect on largest-gap inverted-pairs results for either FFA or PPA (Fig. 4*B*).

In sum, these findings indicate that the strong single-image preference for faces over nonfaces in FFA and places over nonplaces in PPA can be found at all ROI sizes. Nevertheless, ROI size does affect measured category selectivity. Strongest category selectivity is found at smaller ROI sizes for FFA and at larger ROI sizes for left PPA. The clear decrease in discrimination performance for left FFA with increasing ROI size might simply reflect the previously reported finding that left FFA contains fewer strongly face-selective voxels than right FFA (Kanwisher et al., 1997).

Within-category activation profiles are graded

Figure 1 suggests that, within the preferred category, some images activated category-selective regions more strongly than others. We tested this hypothesis by examining the replicability of

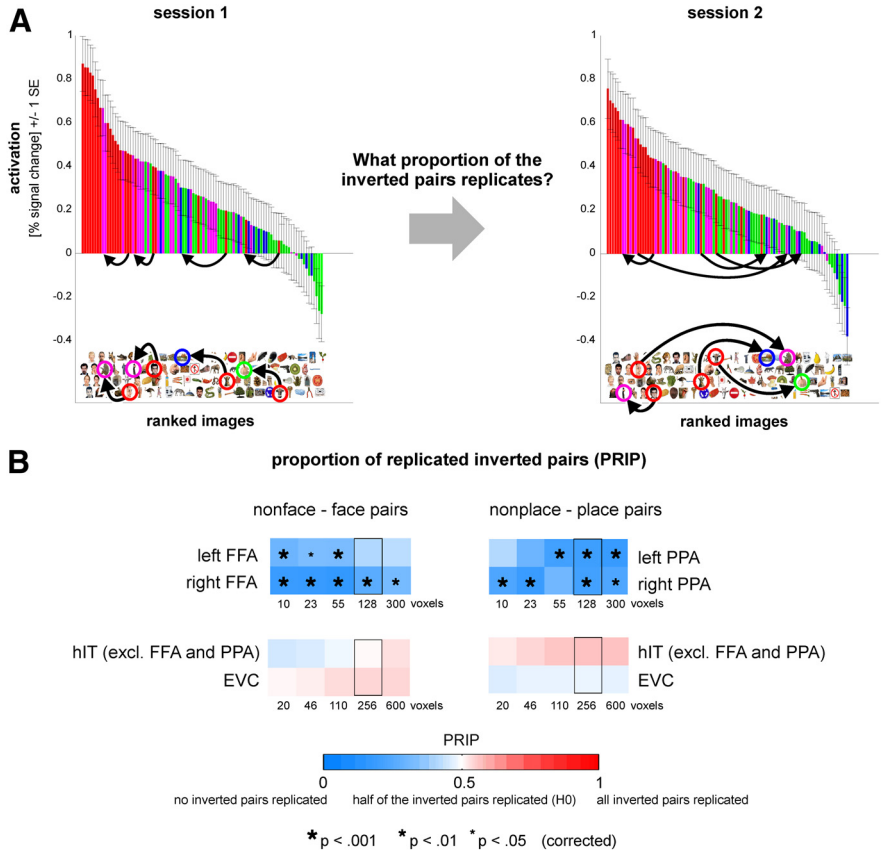


Figure 3. Most inverted pairs do not replicate. How prevalent are true inversions of category preference in FFA and PPA? We investigated this by computing the PRIP. ***A***, Computation of the PRIP. The bar graphs display single-image activation of right FFA (defined at 128 voxels in one subject) for each session. Colored circles connected by black arrows highlight four inverted pairs (i.e., nonpreferred image ranked before preferred image) in session 1. Only 1 of the 4 inverted pairs replicates in session 2, the other 3 revert to category-preferential order (i.e., preferred image ranked before nonpreferred image). If these 4 example pairs were the only inverted pairs in session 1, the PRIP would be 0.25. Color coding is the same as in Figure 1. ***B***, Results of statistical group analysis of the PRIP for category-selective regions FFA and PPA and control regions hIT and EVC. The PRIP was averaged across subjects, allowing for different particular image pairs to be inverted in each subject. Since inverted pairs are defined based on the notion of category preference, the analysis was based on nonface–face pairs for FFA and nonplace–place pairs for PPA. hIT and EVC do not have a strong category preference and were tested for both types of pairs, serving as a control for category-selective regions. We used a two-sided label-randomization test to determine whether the PRIP differed significantly from 0.5; the level we expect under the null hypothesis that the apparently inverted pairs actually have equal activation. A PRIP significantly >0.5 indicates that most inverted pairs replicate. A PRIP significantly <0.5 indicates that most inverted pairs revert to category-preferential order. Results show that most inverted pairs revert to category-preferential order for FFA and PPA for most ROI sizes. The p values were corrected for multiple comparisons as described in Figure 1. Black boxes highlight the ROI sizes used in Figures 1 (FFA and PPA) and 2 (hIT and EVC).

within-category ranking (Fig. 5), which we estimated by rank correlating within-category activation profiles across sessions (Fig. 5*A*). If the ranking for a category of images (e.g., faces) was replicable across sessions, this would indicate that some of these images consistently activated the region more strongly than others. Group results are shown in Figure 5*B*.

Left and right FFA showed replicable ranking for faces, especially at smaller ROI sizes (Fig. 5*B*, top). This indicates that some faces consistently activated FFA more strongly than others. The significant results for group-average activation profiles further suggest that within-face activation profiles were similar across the four subjects. This conclusion was supported by visual inspection of single-subject within-face activation profiles and by intersubject correlation analyses. Effects were somewhat stronger in left than right FFA. Control regions hIT and EVC showed replicable within-face ranking as well, but only for concatenated single-subject activation profiles. This suggests that activation profiles were not consistent across subjects. In addition, effects in EVC were small and were significant at one ROI size only.

Right, but not left, PPA showed replicable ranking for places at most ROI sizes (Fig. 5*B*, bottom). This indicates that some places consistently activated right PPA more strongly than others. The significant results for the group-average activation profiles suggest that within-place activation profiles were similar across the four subjects. This conclusion was supported by visual inspection of single-subject within-place activation profiles and by intersubject correlation analyses. Right FFA and control regions hIT and EVC showed replicable within-place ranking as well. Right FFA showed effects at smaller ROI sizes and hIT at larger ROI sizes. Effects in hIT and EVC were present for the group-average activation profiles, and effects in right FFA were only present for concatenated single-subject activation profiles. These findings suggest that within-place activation profiles were similar across the four subjects for hIT and EVC, but not for right FFA. This conclusion was supported by visual inspection of single-subject within-place activation profiles and by intersubject correlation analyses.

These findings confirm that category-selective regions are activated more strongly by some images of their preferred category than by others, i.e., they show a graded activation profile for images of their preferred category. These effects are not confined to category-selective regions: hIT and EVC show graded within-category activation profiles as well, especially for places.

The presence of within-category activation differences naturally leads us to ask how these differences can be explained. Previous studies have suggested that human faces might activate FFA more strongly than animal faces (Kanwisher et al., 1999). This raises the possibility that graded within-category profiles reflect the existence of subcategories that elicit different levels of activation. For faces, we investigated this possibility by performing a one-sided t test on the within-face activation estimates averaged across subjects and sessions, comparing activation to human versus animal faces in each ROI. This analysis showed that left FFA at the smallest two ROI sizes was indeed activated more strongly by human than animal faces ($t_{(22)} = 3.6$, $p < 0.01$ for 10 voxels; $t_{(22)} = 2.8$, $p < 0.05$ for 23 voxels; p values were corrected for multiple [five] comparisons using Bonferroni correction). Right FFA showed a similar tendency, but results did not reach significance. Other regions did not show differential activation to human versus animal faces. For places, an intuitive subdivision would be natural versus man-made places. A two-sided t test investigating this distinction did not yield significant results in any of our ROIs. Consistent with this, a recent pattern-information study reported that the natural/manmade distinc-

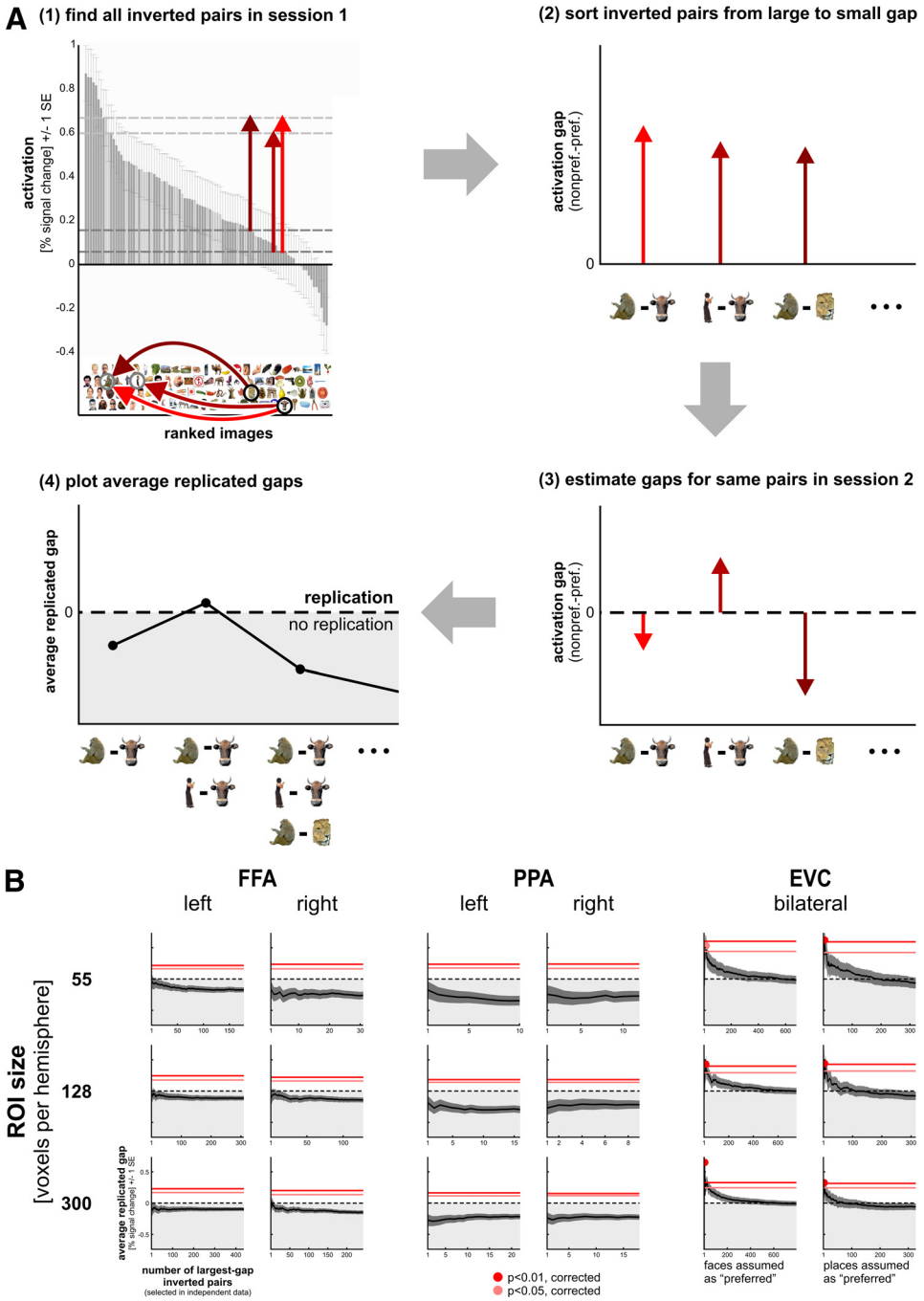


Figure 4. No evidence for any inversions of category preference for particular image pairs in PPA or FFA. Figure 3 showed that most inverted pairs do not replicate. This leaves open the possibility that some inverted pairs do replicate. Here we test the apparently inverted pairs with the largest activation difference (which are least likely to be inverted by noise), using independent data. ***A***, Computation of the average replicated gap. We first find all inverted pairs using session 1 data (step 1). We then sort these pairs according to the size of the activation gap, from the inverted pairs with the largest gaps to the inverted pairs with the smallest gaps (step 2). We estimate the activation gaps for all session 1 inverted pairs using the independent data from session 2 (step 3). Note that these gap estimates are negative for session 1 inverted pairs that do not replicate. For each number $k = 1...n$ of session-1 largest-gap inverted pairs, we average the session 2 gap estimates. We then plot these session 2 average replicated gaps versus k (step 4). We also perform this analysis with sessions 1 and 2 in reverse order and average results from the two directions (without assuming independence of the directions in statistical inference). If there are very few true inverted pairs, the leftmost part of the average replicated gap function has the greatest power to reveal these. If there are more true inverted pairs, averaging replicated gaps for more pairs has greater power for revealing the presence of true inverted pairs. ***B***, The average replicated gap (black solid lines) is plotted as a function of the number of largest-gap inverted pairs for FFA, PPA, and EVC. The dark gray error regions indicate $+/-1$ SE of the estimate. If the average replicated gap function does not emerge into the positive range (above the dashed black line), then even the most promising inverted pairs tend to revert to category-preferential order. If the average replicated gap exceeds the pink or red lines, then there is evidence for truly inverted pairs at $p < 0.05$ (pink line) or $p < 0.01$ (red line) and the peak is marked by a circle whose color indicates the level of significance in the same way. The significance thresholds for the peak were computed by Monte Carlo simulation, accounting for multiple use of the same single-image activation estimate in multiple pairs and for the multiple comparisons along the horizontal axis. Results provide evidence for nonface > face and nonplace > place pairs in EVC, but no evidence for nonface > face pairs in FFA and no evidence for nonplace > place pairs in PPA. Activation profiles were first averaged across subjects; a modified version of this analysis that is sensitive to subject-unique activation profiles revealed some evidence for preference inversions in left FFA.

Table 1. Discriminability (AUC) for faces and places

ROI size (voxels)	10 (20)	23 (46)	55 (110)	128 (256)	300 (600)
Faces versus nonfaces					
Left FFA	0.96***	0.96***	0.94***	0.82***	0.75**
Right FFA	0.98***	0.99***	0.99***	0.94***	0.91***
Left PPA	0.25**	0.22***	0.18***	0.16***	0.16***
Right PPA	0.27**	0.28**	0.22***	0.22***	0.21***
hIT	0.71**	0.72**	0.72**	0.65	0.58
EVC	0.62	0.63	0.62	0.58	0.56
Places versus nonplaces					
Left FFA	0.27*	0.20**	0.18***	0.23*	0.32
Right FFA	0.22**	0.24*	0.25*	0.25*	0.22*
left PPA	0.97***	0.99***	1.00***	1.00***	1.00***
Right PPA	1.00***	1***	1.00***	1***	1***
hIT	0.56	0.5	0.5	0.53	0.54
EVC	0.71	0.71	0.72	0.74*	0.73

p values were computed using a two-sided label-randomization test and were corrected for multiple (five) comparisons using Bonferroni correction. Voxel numbers in between parentheses describe ROI sizes for bilateral hIT and EVC. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$ (corrected).

tion did not explain the overall organization of response patterns in PPA (Kravitz et al., 2011). The main organizational principle of PPA seemed to be spatial expanse (open vs closed places) (Kravitz et al., 2011). Since our stimulus set contained open places only, it is unlikely that the open/closed distinction can explain our within-place activation differences.

Evidence for category step and within-category gradedness in PPA and FFA

The gradedness of within-category activation profiles raises the question of whether the category boundary has a special status at all: Does activation drop in a step-like fashion at the boundary or does it continuously fall off across the boundary? Note that inspecting the noisy activation profile after ranking according to the same profile (Figs. 1, 2) cannot address either the question of gradedness or the question of a category step (see Materials and Methods). Testing for a drop-off of activation at the category boundary requires joint modeling of the category step and the gradedness within and outside the preferred category. This test was implemented by our category-step-and-gradedness analysis (Fig. 6; see Materials and Methods), which uses one session to derive a ranking hypothesis from the data and the other to test a piecewise linear falloff model including predictors for category step and gradedness. Statistical inference was performed by bootstrap resampling of the image set.

Figure 6*B* displays the results of our category-step-and-gradedness analysis (smallest two ROI sizes not shown). Right FFA and right and left PPA showed a significant drop-off of activation at the category boundary at all ROI sizes ($p < 0.0025$ for PPA; $p < 0.05$, with $p < 0.0025$ in several cases, for right FFA). Left FFA did not show a significant drop-off of activation at the category boundary, except at 55 voxels ($p < 0.05$). hIT and EVC both did not show a significant drop-off of activation at the category boundary for either faces or places at any ROI size.

We additionally performed a modified version of our analysis, which is sensitive to subject-unique activation profiles. This analysis again showed a significant category step for right FFA and right and left PPA at all ROI sizes. Left FFA now showed a significant category step at three of five ROI sizes ($p < 0.05$, with $p < 0.0025$ for 55 voxels). There was no evidence for a category step in hIT and EVC.

Results for gradedness within the preferred category (Fig. 6*B*) were consistent with the results on the replicability of within-category ranking reported in the previous section (Fig. 5*B*). Both left and right FFA showed graded within-face activation profiles. Left FFA showed gradedness at the smallest two ROI sizes ($p < 0.0025$) and right FFA at all ROI sizes ($p < 0.0025$) except the largest one. Right but not left PPA showed a graded within-place activation profile at three of five ROI sizes ($p < 0.0025$ for 23 voxels, $p < 0.05$ for 55 and 128 voxels). hIT and EVC showed graded within-place but not within-face activation profiles at most or all ROI sizes ($p < 0.05$, with $p < 0.0025$ in several cases). The subject-unique analysis showed similar results for FFA and PPA. For hIT and EVC, gradedness within places disappeared, while gradedness within faces remained absent. The lack of within-place and within-face gradedness in hIT for the subject-unique analysis forms the only inconsistency with the replicability-of-within-category-ranking results (Fig. 5*B*, left column), and suggests that the subject-unique activation profiles for faces and places in hIT do not fall off linearly. (The category-step-and-gradedness analysis modeled the falloff of activation as linear within preferred and within nonpreferred categories, whereas the replicability-of-within-category-ranking analysis is sensitive to nonlinear graded activation profiles.)

Category-selective regions FFA and PPA also showed graded activation profiles for nonpreferred images at most ROI sizes (Fig. 6*B*). This effect likely reflects both between- and within-category activation differences among the nonpreferred images. In any case, this finding indicates that the activation profile of category-selective regions is graded for images outside the preferred category. hIT and EVC did not show graded activation profiles for nonplaces or nonfaces (except for nonfaces in EVC at the smallest ROI size, $p < 0.05$, data not shown). The subject-unique group analysis showed similar results for FFA. For PPA, gradedness within nonplaces disappeared at most ROI sizes. Results for hIT did not change, while EVC now exhibited gradedness within nonfaces and nonplaces at a small number of ROI sizes ($p < 0.05$).

In sum, our findings indicate that the category boundary has a special status in category-selective regions, especially in right FFA and right and left PPA. The presence of a drop-off of activation at the category boundary in the absence of gradedness would suggest a binary response profile. However, category-selective regions showed gradedness of activation within (except left PPA) and outside the preferred category in addition to the category step at the boundary. This suggests that a binary response function is not sufficient to explain the activation profiles of category-selective regions.

Correlation of activation profiles across regions

Our results suggest functional similarities between certain regions, which we explored further by rank-correlating activation profiles between ROIs (Fig. 7). This exploratory analysis can provide further information on functional similarities between regions, and, more specifically, on the extent to which activation profiles of category-selective regions are inherited from EVC. As for the replicability of within-category ranking (Fig. 5), we combined data across subjects either by concatenating or averaging the activation profiles across subjects. The concatenation approach is sensitive to inter-region correlations of activation profiles even if the particular activation profiles differ across subjects. The averaging approach is sensitive to inter-region correlations of activation profiles that are consistent across subjects. We investigated the inter-region correlations for (1) the full activation profile, (2) the within-face activation profile, and (3) the within-place activation profile. Statistical inference was performed by a standard one-sided test on Spearman's *r*. *p* values were corrected

for multiple testing using Bonferroni correction based on the total number of tests performed.

Figure 7 shows the inter-region correlation results. The main pattern that emerges is that activation profiles are correlated between hemispheres for corresponding regions, and between hIT and EVC (red blocks on diagonals). We subsequently inspected the within-face correlation between FFA and EVC, and the within-place correlation between PPA and EVC. The within-face activation profile was correlated between left but not right FFA and EVC, and the within-place activation profile was not significantly correlated between PPA and EVC. Results were similar across ROI sizes. These results suggest that EVC is not a major contributor to the within-category activation profiles of PPA and right FFA. We then inspected the correlation between category-selective regions (FFA/PPA) and EVC for the full activation profile (top row). The full activation profile was correlated between EVC and both category-selective regions, especially PPA. One interpretation of this finding would be that some degree of category selectivity is already present at the level of EVC, implying that low-level feature differences contribute to some extent to category-selective responses. For places, this seems a plausible interpretation, consistent with our finding that single-image activation of EVC can discriminate places from nonplaces at an above-chance level (Fig. 2). For faces, this interpretation seems less likely: the correlation between EVC and FFA is not significant for the subject-average activation profile, suggesting that the correlation is driven by subject-specific effects (e.g., idiosyncratic arousal effects) and not by face-selectivity of responses (shared across subjects in FFA).

Categorical, yet graded

Figure 8 summarizes our results. Single-image activation profiles of category-selective regions (1) show near-perfect discrimination of preferred from nonpreferred images and no preference inversions for particular object images, (2) show a step-like drop-off at the category boundary, and (3) are graded within and outside the preferred category. It can further be noted that single-image category selectivity is stronger in right than left FFA. In addition, gradedness seems to be more pronounced in FFA; the category step seems to be more pronounced in PPA. In sum, our findings indicate that the activation profiles of category-selec-

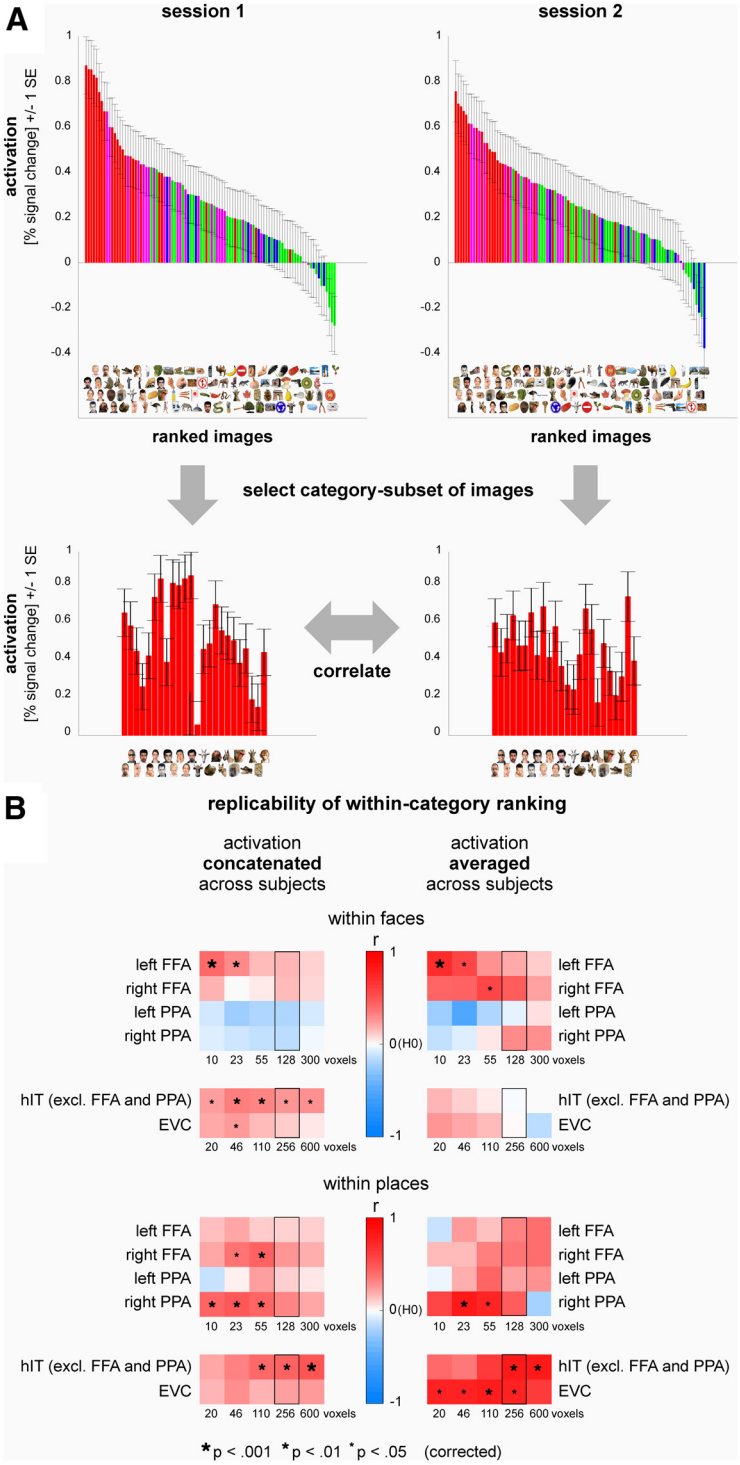
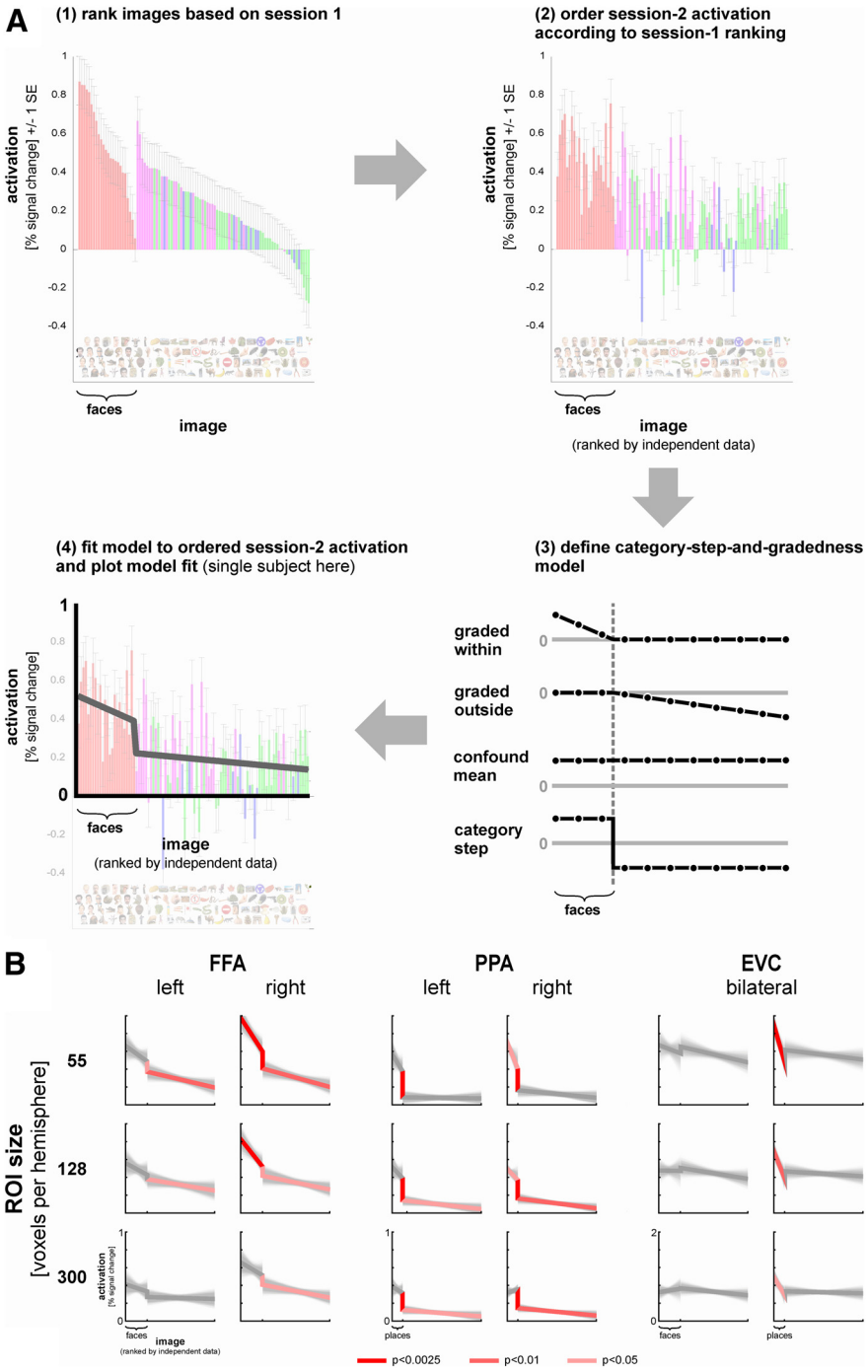


Figure 5. Category-selective regions show graded activation profiles for images of their preferred category. ***A***, If some images consistently activated a region more strongly than other images of the same category (i.e., graded within-category activation profile), the within-category ranking order should replicate across sessions. We computed the replicability of within-category ranking by selecting the same category subset of images in both sessions and correlating their activation estimates using Spearman's *r*. This procedure is illustrated for the within-face activation profile in right FFA defined at 128 voxels in one specific subject. Color coding is the same as in Figure 1. ***B***, Group analysis of replicability of within-category activation profiles for category-selective regions FFA and PPA and for control regions hIT and EVC. Analysis was performed for the image subsets of faces (top) and places (bottom), either using the concatenation approach (left) or the averaging approach (right) for combining single-subject data. Analysis of concatenated single-subject activation profiles is sensitive to replicable ranking regardless of differences in particular ranking order among subjects, while analysis of subject-average activation profiles is sensitive to replicable ranking that is consistent among subjects. We performed a standard one-sided test on Spearman's *r* to determine whether replicability of within-category activation profiles was significantly higher than expected by chance (H0: $r = 0$). *p* values were corrected for multiple comparisons as described in Figure 1. Black boxes highlight the ROI sizes that results were displayed at in Figures 1 (FFA and PPA) and 2 (hIT and EVC).



tive regions in IT cortex are categorical, yet graded.

Discussion

FFA and PPA might respond more strongly to every single member of their preferred category than to any nonmember

We measured single-image activation of human category-selective regions to 96 object images from a wide range of categories, and investigated whether category selectivity holds in general or is violated by particular single images. We found good discrimination of preferred from nonpreferred stimuli based on single-image activation of category-selective regions FFA and PPA across a wide range of ROI sizes. Furthermore, we did not find evidence for violations of category-consistent ranking by particular single images, except in left FFA. Together, these findings suggest the possibility that right FFA and left and right PPA respond more strongly to every single member of their preferred category than to any nonmember.

This conclusion is consistent with several single-image studies in monkeys that showed strong face-selectivity in the macaque middle and anterior superior temporal sulcus (STS) (Földiák et al., 2004; Tsao et al., 2006). These studies reported cells that responded almost exclusively to faces. It should be noted that many of the recorded cells in the middle macaque face patch, a suggested homolog of FFA located in the STS (Tsao et al., 2003, 2006), also responded significantly to several nonface images (Tsao et al., 2006). These nonface images shared lower-level visual properties with face images (e.g., round shape). However, at the population level (i.e., when responses were averaged across the population of visually responsive cells in the middle face patch), the responses elicited by these nonface images were

←

Figure 6. Category steps and graded activation profiles in PPA and FFA. Figures 1 and 5 suggest that activation profiles might be graded in FFA and PPA. This raises the question whether the category-average activation difference (by which these regions are functionally defined) can be accounted for by a continuously graded falloff without a step at the category boundary. Inspecting the noisy activation profile after ranking according to the same profile (Figs. 1, 2) cannot address this question (see Materials and Methods). Testing for a step-like drop in activation across the boundary requires joint modeling of category step and gradedness. ***A***, Implementation of the category-step-and-gradedness analysis. We first rank the images within and outside the preferred category according to session 1 activation (step 1). We then order the session 2 activation profile according to the session 1 ranking (step 2). We define a linear falloff model consisting of four predictors: a positive ramp predictor for the preferred category (0 elsewhere), a negative ramp predictor for the nonpreferred category (0 elsewhere), a confound mean predictor (1 everywhere), and a category-step predictor (1 for preferred, −1 for nonpreferred) (step 3). The ramps were defined such that setting the category step to 0 would yield a piecewise linear falloff with a kink, but no step (no discontinuity), at the category boundary (gray dashed line). We then fit the model by ordinary least-squares to the activation profile estimated from session 2 and plot the model fit (step 4). The procedure is illustrated for the activation profile of right FFA defined at 128 voxels in one specific subject. The procedure was repeated using the sessions in reverse order, and the resulting β estimates averaged. ***B***, Model fits for FFA, PPA, and EVC. To assess the dependency of the estimates on the particular sample of stimuli, we bootstrap-resampled the stimulus set 10,000 times and performed the model-fitting procedure on each bootstrap sample in both directions. We computed a p value for the β estimate of each predictor of the falloff model as the percentile of 0 within the bootstrap distribution of the β estimates (one-sided tests). The panels show the fitted falloff model predictions with the color of each line section coding for the significance of the corresponding model component (gray, not significant; light pink, $p < 0.05$; bright pink, $p < 0.01$; red, $p < 0.0025$). In the background, the 10,000 bootstrap model predictions are transparently overplotted in gray. Results show a large, significant category step in PPA (left and right); a small significant category step in right FFA; evidence for graded preferred activation profiles in FFA and right PPA; and evidence for graded nonpreferred activation profiles in right and left FFA and PPA. Activation profiles were first averaged across subjects; a modified version of this analysis that is sensitive to subject-unique activation profiles gave similar results.

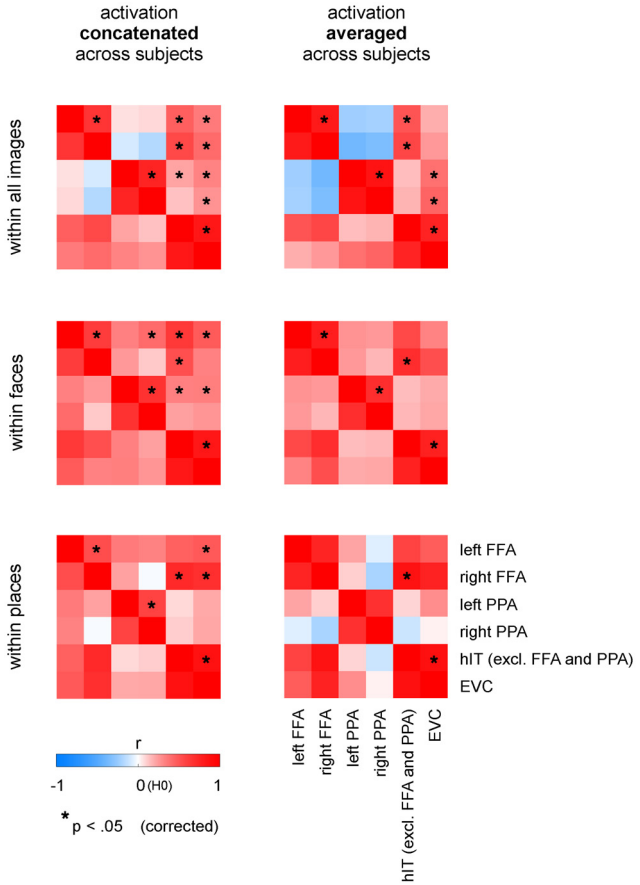


Figure 7. Activation profiles are correlated between early visual and IT cortex, and between hemispheres for corresponding regions. We rank-correlated activation profiles across regions to explore which regions showed similar activation profiles. We performed our correlation analysis for all images, for faces only, and for places only. Results are shown both for the concatenation approach (left) and the averaging approach (right) for combining single-subject data. Each correlation matrix is mirror-symmetric about a diagonal of ones. We performed a standard one-sided test on Spearman's r to determine whether inter-region correlations of activation profiles were significantly higher than expected by chance (H0: $r = 0$). p values were corrected for multiple comparisons using Bonferroni correction for the whole figure (15 inter-region comparisons × 6 matrices = 90 comparisons). Results are shown for the ROI sizes that results were displayed at in Figures 1 (FFA and PPA) and 2 (hIT and EVC). Results show correlated activation profiles between EVC and IT, and between hemispheres for FFA and PPA. In addition, the activation profile of EVC is correlated with that of category-selective regions when considering the full image set, but not within places for PPA, and within faces only for left (not right) FFA. This suggests that EVC is not a major contributor to the within-category activation profiles of PPA and right FFA.

clearly weaker than those elicited by any of the face images (Tsao et al., 2006). Kiani et al. (2007) reported a similar finding: they measured responses of face-selective cells in macaque IT cortex, and reported imperfect face selectivity at the single-cell level but close-to-perfect face selectivity when responses were averaged over a small population of face-selective IT cells. These findings are consistent with the idea that category membership of natural objects is encoded at the population level (Vogels, 1999; Kiani et al., 2007; Kriegeskorte et al., 2008).

In sum, our results suggest that category selectivity of FFA and PPA, conventionally investigated using category-average activation, might hold for single images. FFA single-image activation profiles appear similar to those described for the macaque middle face patch, consistent with a homology or functional analogy. If the recently reported place-selective region in the macaque (Bell et al., 2009) is the homolog or functional analog of the human PPA, then the monkey region should similarly exhibit essentially perfect categorical ranking and a pronounced category step.

Activation profiles are graded with step, not binary

Previous category-average studies left open whether category-selective regions simply act as a binary classifier or whether they show graded responses to individual exemplars of a category. This suggests three different possible scenarios. In the first scenario, the activation profile of the region follows a binary response function, i.e., there are only two possible levels of activation: high for exemplars of the preferred category and low for exemplars of nonpreferred categories. In the second scenario, the activation profile of the region still shows a category step, but is graded within and/or outside the preferred category, i.e., some category members activate the region more strongly than others. In the third scenario, the activation profile of the region falls off continuously, i.e., there is no step at the category boundary. Our results support the second scenario: FFA and PPA showed a category step, but also a graded activation profile for exemplars within and outside their preferred category.

The presence of gradedness is consistent with a recent monkey fMRI study that reported activation differences in face- and place-selective regions in IT between visually dissimilar exemplars of the preferred category (Bell et al., 2009). It is also in line with an earlier monkey electrophysiology study that reported a population of tree-selective cells in IT whose mean response differed across tree exemplars (Vogels, 1999). Other reports on gradedness of activation focused on differences between nonpreferred categories (Downing et al., 2006; Kiani et al., 2007) and did not investigate differences between exemplars.

There are several possible interpretations of the within-category activation differences reported here. First, it could be that activation differences between exemplars reflect differences in low-level visual features. Consistent with this idea, we found within-category activation differences in EVC, especially for places. However, the lack of correlation between within-place activation profiles of PPA and EVC suggests that the place exemplar differences in PPA do not reflect low-level visual differences represented at the level of overall activation of EVC. Second, within-category activation differences could be driven by subcategories that elicit different levels of activation. Consistent with this explanation, we found stronger activation to human than animal faces in FFA. Third, our within-category activation differences could be interpreted as attentional effects. Attention enhances responses to stimuli in object-selective cortex (Wojciulik et al., 1998; O'Craven et al., 1999) and early visual regions (Liu et al., 2005). Stimuli might differ in the extent to which they trigger attention. For example, high-valence stimuli (e.g., angry face) might trigger more attention than low-valence stimuli (e.g., neutral face), resulting in activation differences among stimuli (Breiter et al., 1996; Lane et al., 1999; Palermo and Rhodes, 2007). Fourth, activation differences between exemplars might reflect differences between the underlying distributed patterns of activity that are thought to represent them (Young and Yamane, 1992; Edelman et al., 1998; Tsao et al., 2006; Kiani et al., 2007; Eger et al., 2008; Kriegeskorte et al., 2008). Exemplar information carried by distributed activity patterns might get lost by pooling (Kriegeskorte et al., 2006, 2007; Eger et al., 2008), but could also to some extent be reflected in regional-average activation. Further single-image studies are needed to address these possibilities and test specific hypotheses as to the causes of the within-category activation differences.

		(A) Category discrimination: Are the categories discriminated? If so, how well? (Figs. 1, 2)	(B) Inversions: Is there evidence for preference inversions for particular object images? (Figs. 3, 4)	(C) Step: Is there a step-like activation drop at the category boundary? (Fig. 6)	(D) Graded for preferred: Are activations graded within the preferred category? (Figs. 5, 6)	(E) Graded for nonpreferred: Are activations graded within the nonpreferred categories? (Fig. 6)
FFA	**left**	**Yes,** quite good.	No, for average activation profiles. Yes, for subject-specific weighting and subject-unique profiles.	Yes, for 55 voxels. No, for other ROI sizes.	**Yes**, but not for large ROIs.	**Yes**, but not for 300-voxel ROI.
	right	**Yes,** very good.	**No evidence**.	**Yes**, small step.	**Yes**, but not for 300-voxel ROI.	**Yes**.
PPA	**left**	**Yes,** near perfect.	**No evidence**.	**Yes**, large step.	**No evidence.**	**Yes**, but not for small ROIs.
	right	**Yes,** perfect.	**No evidence**.	**Yes**, large step.	**Yes**, but not for 10- and 300-voxel ROIs.	**Yes**, but not for small ROIs.
control region: EVC (bilateral)		Weak place preference (Fig. 2).	Yes, evidence for inversions irrespective of face and place categories (Fig. 4).	**No evidence**.	Graded activation weak within faces, strong within places (Figs. 5, 6).	

yes
no evidence
mixed results

Figure 8. Summary of results.

Single-image designs for studying regional-average activation and pattern information

The classical fMRI category-block-design studies (e.g., Kanwisher et al., 1997; Epstein and Kanwisher, 1998) averaged across stimuli within predefined categories and across response channels (i.e., voxels within contiguous regions). Haxby et al. (2001) studied pattern information, but still averaged patterns within predefined categories. Kriegeskorte et al. (2008) studied pattern information of single-image response patterns, enabling data-driven discovery of category structure (Edelman et al., 1998). The present study constitutes a missing link in the sense that it considers single-image responses, but in terms of regional-average activation levels.

Building on previous single-image fMRI approaches (Edelman et al., 1998; Aguirre, 2007; Kriegeskorte et al., 2007, 2008; Eger et al., 2008; Haushofer et al., 2008; Kravitz et al., 2011), this study further demonstrates the feasibility of single-image fMRI experiments. Single-image designs reduce experimenter bias because they do not assume any grouping of the stimuli in design or analysis. They enable exemplar-based analyses and empirical discovery of categorical and continuous response characteristics in high-level visual cortex. The novel single-image analyses introduced in this paper for fMRI data might also be useful to cell-recording studies. Homologies or functional analogies between monkey and human category-selective regions are not established, and could be probed using single-image designs. However, it should be kept in mind that the fMRI-based regional-average activation analyses we pursue here operate at a different scale than pattern-information fMRI and cell recordings.

In what sense is the representation categorical? And in what sense is it not categorical?

The object representation in IT does not seem to be categorical in the sense of a binary response function. This has now been demonstrated both at the level of single-cell responses in the monkey (Vogels, 1999; Tsao et al., 2006; Kiani et al., 2007) and at the level of regional-average activation in the human (current study). Within-category response variation in IT has also been shown in the form of pattern-information differences between exemplars of the same category (Tsao et al., 2006; Kriegeskorte et al., 2007; Eger et al., 2008). Lateral prefrontal cortex, which receives input from IT, seems a more likely candidate for binary neuronal category representations (Freedman et al., 2001). However, the object representation in IT *is* categorical in the sense of potentially perfect rank-ordering by category (current study), the presence of a category step (current study), and categorical clustering of activity patterns (Kiani et al., 2007; Kriegeskorte et al., 2008).

One overall interpretation of these findings is that the object representation in IT strikes a balance between maximizing the between- and the within-category information. The optimal solution would enable representation of both object category (largest component of variance) and object identity. Such a solution might be implemented by feature selectivity at the columnar level (Tanaka, 1996) which is tuned to those object features that are most informative for discriminating categories as well as exemplars (Sigala and Logothetis, 2002; Ullman et al., 2002; Lerner et al., 2008), while untangling category and exemplar distinctions from accidental properties in multivariate space (DiCarlo and Cox, 2007).

Notes

Supplemental material for this article is available at http://www.mrc-cbu.cam.ac.uk/research/visualobjectslab/supplementary/MurEtAl-CategoricalYetGraded-Supplement.pdf. The supplemental material consists of results of several analyses that were reported in the results section of the main paper but that were not shown in the main figures. The supplemental material includes (1) results for all five ROI sizes for the largest-gap-inverted-pairs test, the category-step-and-gradedness test, and the inter-region-activation-

profile correlation test, (2) subject-unique group results for the largest-gap-inverted-pairs test and category-step-and-gradedness test, and (3) optimally weighted subject-average group results for the largest-gap-inverted-pairs test. This material has not been peer reviewed.

References

Aguirre GK (2007) Continuous carry-over designs for fMRI. Neuroimage 35:1480–1494.

Bedny M, Aguirre GK, Thompson-Schill SL (2007) Item analysis in functional magnetic resonance imaging. Neuroimage 35:1093–1102.

Bell AH, Hadj-Bouziane F, Frihauf JB, Tootell RB, Ungerleider LG (2009) Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging. J Neurophysiol 101:688–700.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207–4221.

Breiter HC, Etcoff NL, Whalen PJ, Kennedy WA, Rauch SL, Buckner RL, Strauss MM, Hyman SE, Rosen BR (1996) Response and habituation of the human amygdala during visual processing of facial expression. Neuron 17:875–887.

DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. Trends Cogn Sci 11:333–341.

Downing PE, Chan AW, Peelen MV, Dodds CM, Kanwisher N (2006) Domain specificity in visual cortex. Cereb Cortex 16:1453–1461.

Edelman S, Grill-Spector K, Kushnir T, Malach R (1998) Toward direct visualization of the internal shape representation space by fMRI. Psychobiology 26:309–321.

Eger E, Ashburner J, Haynes JD, Dolan RJ, Rees G (2008) fMRI activity patterns in human LOC carry information about object exemplars within category. J Cogn Neurosci 20:356–370.

Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature 392:598–601.

Földiák P, Xiao D, Keysers C, Edwards R, Perrett DI (2004) Rapid serial visual representation for the determination of neural selectivity in area STSa. Prog Brain Res 144:107–116.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. Science 291:312–316.

Grill-Spector K, Knouf N, Kanwisher N (2004) The fusiform face area subserves face perception, not generic within-category identification. Nat Neurosci 7:555–562.

Haushofer J, Livingstone M, Kanwisher N (2008) Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. PloS Biol 6:e187.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and object in ventral temporal cortex. Science 293:2425–2430.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: A module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.

Kanwisher N, Stanley D, Harris A (1999) The fusiform face area is selective for faces not animals. Neuroreport 10:183–187.

Kiani R, Esteky H, Mirpour K, Tanaka K (2007) Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J Neurophysiol 97:4296–4309.

Kravitz DJ, Peng CS, Baker CI (2011) Real-world scene representations in high-level visual cortex: It's the spaces more than the places. J Neurosci 31:7322–7333.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868.

Kriegeskorte N, Formisano E, Sorger B, Goebel R (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc Natl Acad Sci U S A 104:20600–20605.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60:1126–1141.

Lane RD, Chua PM, Dolan RJ (1999) Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. Neuropsychologia 37:989–997.

Lerner Y, Epshtein B, Ullman S, Malach R (2008) Class information predicts activation by object fragments in human object areas. J Cogn Neurosci 20:1189–1206.

Liu T, Pestilli F, Carrasco M (2005) Transient attention enhances perceptual performance and fMRI response in human visual cortex. Neuron 45:469–477.

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci U S A 92:8135–8139.

O'Craven KM, Downing PE, Kanwisher N (1999) fMRI evidence for objects as the units of attentional selection. Nature 401:584–587.

Palermo R, Rhodes G (2007) Are you always on my mind? A review of how face perception and attention interact. Neuropsychologia 45:75–92.

Puce A, Allison T, Gore JC, McCarthy G (1995) Face-sensitive regions in human extrastriate cortex studied by functional MRI. J Neurophysiol 74:1192–1199.

Rajimehr R, Devaney KJ, Bilenko NY, Young JC, Tootell RB (2011) The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. PLoS Biol 9(4):e1000608.

Sigala N, Logothetis NK (2002) Visual categorization shapes feature selectivity in the primate temporal cortex. Nature 415:318–320.

Tanaka K (1996) Inferotemporal cortex and object vision. Annu Rev Neurosci 19:109–139.

Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque cerebral cortex. Nat Neurosci 6:989–995.

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. Science 311:670–674.

Ullman S, Vidal-Naquet M, Sali E (2002) Visual features of intermediate complexity and their use in classification. Nat Neurosci 5:682–687.

Vogels R (1999) Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. Eur J Neurosci 11:1239–1255.

Wojciulik E, Kanwisher N, Driver J (1998) Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. J Neurophysiol 79:1574–1578.

Young MP, Yamane S (1992) Sparse population coding of faces in inferotemporal cortex. Science 256:1327–1331.