Behavioral/Cognitive

The Functional Organization of High-Level Visual Cortex Determines the Representation of Complex Visual Stimuli

Libi Kliger[1] and Galit Yovel[1,2]
[1]The School of Psychological Sciences, Tel Aviv University, Tel Aviv 6997801, Israel, and [2]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv 6997801, Israel

A hallmark of high-level visual cortex is its functional organization of neighboring areas that are selective for single categories, such as faces, bodies, and objects. However, visual scenes are typically composed of multiple categories. How does a category-selective cortex represent such complex stimuli? Previous studies have shown that the representation of multiple stimuli can be explained by a normalization mechanism. Here we propose that a normalization mechanism that operates in a cortical region composed of neighboring category-selective areas would generate a representation of multi-category stimuli that varies continuously across a category-selective cortex as a function of the magnitude of category selectivity for its components. By using fMRI, we can examine this correspondence between category selectivity and the representation of multi-category stimuli along a large, continuous region of cortex. To test these predictions, we used a linear model to fit the fMRI response of human participants (both sexes) to a multi-category stimulus (e.g., a whole person) based on the response to its component stimuli presented in isolation (e.g., a face or a body). Consistent with our predictions, the response of cortical areas in high-level visual cortex to multi-category stimuli varies in a continuous manner along a weighted mean line, as a function of the magnitude of its category selectivity. This was the case for both related (face + body) and unrelated (face+wardrobe) multi-category pairs. We conclude that the functional organization of neighboring category-selective areas may enable a dynamic and flexible representation of complex visual scenes that can be modulated by higher-level cognitive systems according to task demands.

Key words: **category-selective visual cortex; face; high-level vision; neuroimaging; normalization model**

Significance Statement

It is well established that the high-level visual cortex is composed of category-selective areas that reside in nearby locations. Here we predicted that this functional organization together with a normalization mechanism would generate a representation for multi-category stimuli that varies as a function of the category selectivity for its components. Consistent with this prediction, in an fMRI study we found that the representation of multi-category stimuli varies along high-level visual cortex, in a continuous manner, along a weighted mean line, in accordance with the category selectivity for a given area. These findings suggest that the functional organization of high-level visual cortex enables a flexible representation of complex scenes that can be modulated by high-level cognitive systems according to task demands.

Introduction

A fundamental feature of the high-level visual cortex of primates is its division into category-selective areas, such as face, body, or object-selective regions that reside in nearby locations (Malach et al., 1995; Kanwisher et al., 1997; Downing et al., 2001; Kanwisher and Yovel, 2006; Grill-Spector and Weiner, 2014). This division into category-selective areas has led to numerous studies that have examined the profile of response of these areas to isolated stimuli of these categories. Nevertheless, visual scenes are typically composed of multiple objects, and it is therefore essential to understand the nature of their representation in high-level visual cortex.

To study the representation of multi-category stimuli, previous single-neuron and fMRI studies have examined the relative contributions of the isolated stimuli to the response of multi-category stimuli. These studies found different patterns of response in different areas of high-level visual cortex. Whereas the response in object-general areas, such as inferior temporal cortex in monkeys (Zoccolan et al., 2005) or lateral occipital cortex in

Received Feb. 25, 2020; revised July 28, 2020; accepted Aug. 4, 2020.

Author contributions: L.K. and G.Y. designed research; L.K. performed research; L.K. analyzed data; L.K. and G.Y. wrote the paper.

The authors declare no competing financial interests.

This work is supported by a grant from the Israeli Science Foundation (ISF 446/16). We thank Tom Schonberg, Roy Mukamel, Jonathan Rosenblatt, Matan Mazor, and Nathaniel Daw for helpful input on this work; and Talia Brandman and Michal Bernstein for comments on this manuscript.

Correspondence should be addressed to Libi Kliger at libikl@mail.tau.ac.il.

https://doi.org/10.1523/JNEUROSCI.0446-20.2020

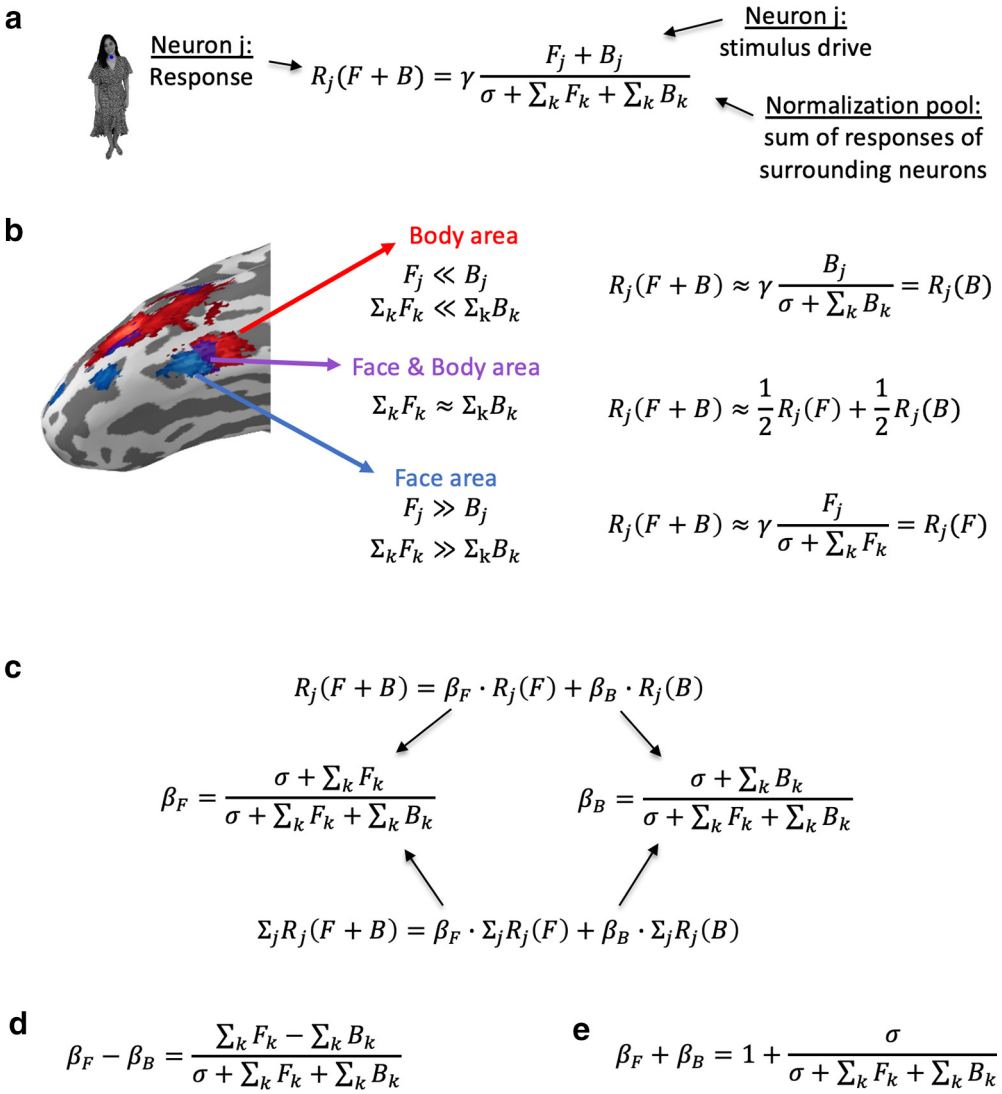

Figure 1. ***a***, The normalization equation (Reynolds and Heeger, 2009). The response of a neuron is divided (normalized) by the sum of the responses of the surrounding neurons. Here we show the response to a face (F) and a body (B) presented together. ***b***, A surface map of face- and body-selective areas with the predicted response based on the normalization equation: a face-selective area (blue) and a body-selective area (red) contain homogeneous surrounding neurons that are selective for the same category, therefore resulting in a maximum-like response. An area in the border between the face- and body-selective areas (purple) contains a heterogeneous surrounding of face-selective neurons and body-selective neurons. If half of the neurons are face selective and half are body selective, then the response to a face and a body should be the mean of the responses to the isolated stimuli. ***c***, Using mathematical derivations of the normalization equation (***a***), the response to a pair of stimuli can be described as a weighted mean of the responses to the isolated stimuli. The weights (β_F and β_B) are the contributions of the face and the body to the face + body response and are determined by the proportions of face- and body-selective neurons within the normalization pool. The fMRI BOLD signal reflects the response of a sum of neurons with similar normalization pools, and therefore the same linear relationship between the pair and the isolated stimuli also applies for the fMRI response, with the same weights as for the single-neuron equation. ***d***, ***e***, The normalization equation further predicts that the difference between the weights corresponds to the difference in the proportions of face and body selective neurons (***d***), and that the sum of weights is slightly >1 (i.e., 1 plus a small positive term; ***e***). Formal derivations can be found at https://github.com/gylab-TAU/multiple_objects_fMRI_analysis.

humans (MacEvoy and Epstein, 2009; Baeck et al., 2013) was a mean or a weighted mean response of the isolated stimuli, the response in category-selective areas, such as face or scene areas (Reddy et al., 2009; Bao and Tsao, 2018), was similar to the response to the preferred category (i.e., a maximum response). A normalization model was proposed to account for these findings. According to the normalization model, the response of a neuron to a stimulus is divided by the response of its surrounding neurons (Carandini and Heeger, 2011; Fig. 1*a*), and therefore it reduces the response to multi-category stimuli relative to the response to the preferred stimulus when presented alone. Nonetheless, the differences between specific implementations of the normalization model (i.e., responses diverging from mean to maximum) that were found in different category-selective areas were not addressed. To account for these differences, Bao and Tsao (2018) suggested that the response to multiple-category stimuli may vary as a function of the homogeneity of the normalization pool. If the surrounding neurons are selective for the same category as the recorded neuron (i.e., a face neuron in a face-selective area), the normalization pool is unresponsive to the nonpreferred stimulus and therefore does not reduce the response of the recorded neuron to its preferred stimulus, yielding a maximum response.

Here we provide a general framework for the relationship between category selectivity and the representation of multi-category stimuli, as detailed below (Fig. 1), by showing this correspondence with fMRI across a large continuous area of cortex. Category selectivity, as measured with fMRI, can provide an

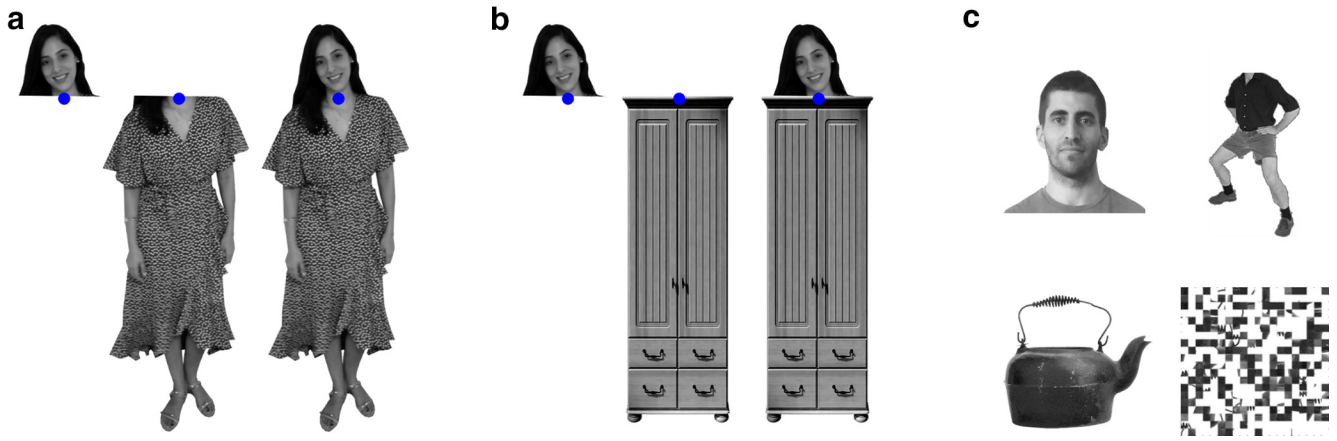

Figure 2. ***a***, A face–body stimulus set: face, body, and face+body stimuli, taken from the same images. The fMRI response to these stimuli was used to estimate the contribution of the face and the body to the face+body representation. Participants were asked to fixate on the blue dot and perform a 1-back task (see Materials and Methods). ***b***, A face–object stimulus set: face, object, and face+object stimuli, all taken from the same images. Participants were asked to fixate on the blue dot and perform a 1-back task. We used wardrobes as the objects, which were matched to the body stimuli in terms of low-level visual properties. The fMRI response to these stimuli was used to estimate the contribution of the face and the object to the face+object representation. ***c***, Functional localizer stimulus set: faces, bodies, objects, and scrambled objects. Functional localizer data were used to define category-selective regions of interest and to measure the voxelwise selectivity for specific categories, independently from the data that were used to estimate the contribution of each part to the multi-category representation.

estimate of the proportion of neurons that are selective for each of the measured categories and therefore with a measure of the homogeneity of the normalization pool. A voxel that shows high selectivity for a given category has a larger proportion of neurons selective for this category and therefore a homogeneous normalization pool. A voxel that shows a similar response to different categories reflects a mixture of category-selective neurons and therefore a heterogeneous normalization pool. We therefore predict that the response to multi-category stimuli will vary from a maximum response in category-selective areas to a mean response in areas that show similar responses to multiple categories, such as in the borders between two category-selective areas (Fig. 1*b*). More generally, we predict that the response to multi-category stimuli will be a weighted mean of the response to each of its components, and that the magnitude of category selectivity for each of the stimuli will determine its weights (Fig. 1*c–e*). Support for this prediction will offer a general framework for the various findings reported in previous studies that looked at the representation of multi-category stimuli in different category-selective regions.

Materials and Methods

To test the correspondence between the magnitude of category selectivity and the representation of multi-category stimuli in high-level visual cortex, we ran two fMRI studies. In the first study the multi-category stimulus was a whole person (face + body; Fig. 2*a*), and we estimated the response to the multi-category stimulus based on the response to the isolated components, a face and a body, by fitting a linear model to the data (Reddy et al., 2009). In a second experiment, we replicated these findings and generalized them to a face + object stimulus (Fig. 2*b*).

Participants

Thirty-two healthy volunteers with normal or corrected-to-normal vision participated in both experiments. Fifteen volunteers (6 women; age range, 19–37 years; 13 right handed) participated in experiment 1, and 17 healthy volunteers (11 women; age range, 20–30 years; 14 right handed) who did not participate in experiment 1 participated in experiment 2. Two participants were excluded from analysis of experiment 2 because of technical difficulties. Participants were paid $15/h. All participants provided written informed consent to participate in the study, which was approved by the ethics committees of the Sheba Medical Center and Tel Aviv University, and were performed in accordance with relevant guidelines and regulations. The sample size for each experiment ($N = 15$) chosen for this study was similar to the sample sizes of other fMRI studies that examined the representation of multiple objects in high-level visual cortex (10–15 subjects/experiment; see MacEvoy and Epstein, 2009, 2011; Reddy et al., 2009; Baeck et al., 2013; Song et al., 2013; Kaiser et al., 2014; Baldassano et al., 2016; Kaiser and Peelen, 2018).

Stimuli

Face + body stimuli. The face + body stimuli set was used in both experiment 1 and experiment 2. Stimuli consisted of 40 grayscale images of a whole person standing in a straight frontal posture with their background removed, which were downloaded from the Internet (20 men and 20 women identities). Each image of a person was cut into two parts approximately in the neck area, resulting in a face stimulus and a headless body stimulus for each identity (Fig. 2*a*). The isolated face and body stimuli were presented in the same location they occupied in the whole-person stimulus. A blue fixation dot was presented at a constant location around the neck on the screen across all conditions (Fig. 2*a*, middle and top part of the display). The size of the whole-person image was ~3.5° × 12.2° of visual angle.

Face + object stimuli. The face + object stimuli set was used in experiment 2 in addition to the face + body stimuli set. Stimuli consisted of pictures of faces, wardrobes, and faces-above-wardrobes (Fig. 2*b*). The face stimuli were the same 40 images of faces used in the face + body stimuli. For the object stimuli, we used 40 images of grayscale wardrobes with their backgrounds removed that were taken from the Internet. We digitally manipulated the images of the wardrobes so that the object location, size (number of pixels on the screen), contrast, and luminance will be matched to the 40 pictures of headless bodies from the face + body stimuli. The face + object stimuli were created by placing the wardrobe images right below the face in the same location of the body (i.e., a face above a wardrobe with no gap between them). A blue fixation dot was presented at a constant location on the screen across all conditions right over the neck in the same location as in experiment 1 (Fig. 2*b*). The size of the face + object pair was similar to that of the face + body and was ~3.5° × 12.2° of visual angle.

Functional localizer stimuli. Functional localizer stimuli of experiment 1 were grayscale images of faces, headless bodies, and nonleaving objects (Fig. 2*c*), and images of the whole person that were not included in analyses of this study. Functional localizer stimuli of experiment 2 were grayscale pictures of faces, headless bodies, nonleaving objects, and scrambled objects (Fig. 2*c*). The size of the stimuli was ~5.5° × 5.5° of visual angle.

Apparatus and procedure

fMRI acquisition parameters. fMRI data were acquired using a 3T Siemens MAGNETOM Prisma MRI scanner at Tel Aviv University,

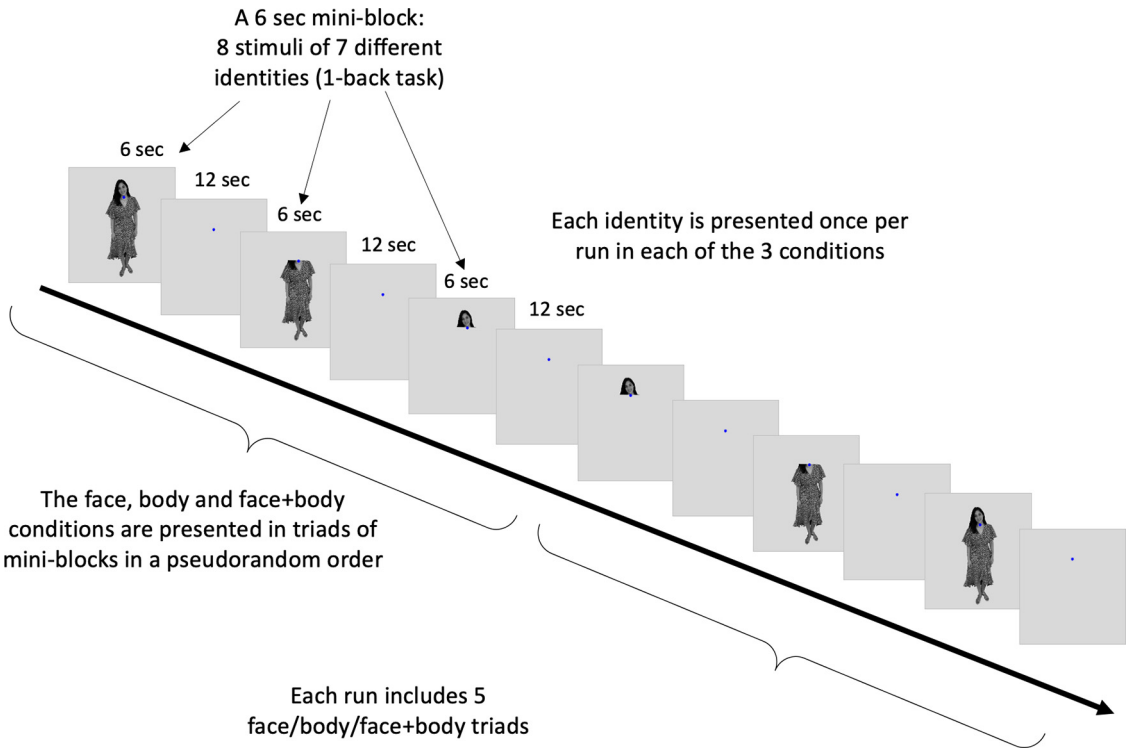

Figure 3. Experimental procedure. Each run had 15 mini-blocks of three conditions (5 blocks each). See Materials and Methods for a full description of the procedure.

using a 64-channel head coil. Echoplanar volumes were acquired with the following parameters: repetition time (TR) = 2 s; echo time (TE) = 30 ms; flip angel = 82°; 64 slices per TR; multiband acceleration factor = 2; acceleration factor PE (phase encoding) = 2; slice thickness = 2 mm; field of view = 20 cm and a 100 × 100 matrix, resulting in a voxel size of 2 × 2 × 2 mm. Stimuli were presented with MATLAB (MathWorks) and Psychtoolbox (Brainard, 1997; Kleiner et al., 2007) and were displayed on a 32 inch high-definition LCD screen (NordicNeuroLab) viewed by the participants at a distance of 155 cm through a mirror located in the scanner. Anatomical MPRAGE images were collected with 1 × 1 × 1 mm resolution, TE = 2.88 ms, and TR = 2.53 s.

Experimental procedure—experiment 1. The study included a single recording session with six runs of the main experiment and three runs of the functional localizer. Each of the six main-experiment runs included five triads of face, body, and face + body mini-blocks. Figure 3 shows an example of two such triads. The order of face, body, and face + body mini-blocks within each triad was counterbalanced across triads and runs. Each mini-block included eight stimuli, of which seven were of different identities and one identity repeated for the 1-back task. The identities presented in the face, body, and face + body mini-blocks within a triad were different. Thus, each run included face, body, and face + body stimuli of 35 different identities (7 identities × 5 triads). The 35 identities were randomly chosen from the set of 40 identities. Each mini-block lasted 6 s and was followed by 12 s of fixation. A single stimulus display time was 0.325 s, the interstimulus interval was 0.425 s. Subjects performed a 1-back task (one repeated stimulus in each block). Each run began with 6 s (three TRs) of fixation (dummy scan) and lasted a total of 276 s (138 TRs). Subjects were instructed to maintain fixation throughout the run, and their eye movements were recorded with an eye tracker (EyeLink).

Experimental procedure—experiment 2. The experiment included a single recording session with six runs of the main experiment and three runs of localizer. The main experiment included three runs of face, body, and face + body stimuli, identical to experiment 1. In addition, three runs of face, object, and face + object stimuli were presented using the same design used for the face and body runs (Fig. 3). The face + object runs were presented before the face + body runs to avoid the priming of a body in the object and face + object mini-blocks. Subjects were instructed to maintain fixation throughout the run, and their eye movements were recorded with an eye tracker (EyeLink).

Functional localizer. Each run of the functional localizer in both experiments included 21 blocks: 5 baseline fixation blocks and 4 blocks for each of the four experimental conditions, faces, bodies, objects, and persons (analysis of the person condition is not included in this article) in experiment 1, and faces, bodies, objects, and scrambled objects in experiment 2. Each block presented 20 stimuli of 18 different images, of which 2 repeated twice for a 1-back task. Each stimulus was presented for 0.4 s with a 0.4 s interstimulus interval. Each block lasted 16 s. Each run began with a 6 s fixation (three TRs) and lasted a total of 342 s (171 TRs).

Data analyses: fMRI data analysis and preprocessing

fMRI analysis was performed using SPM12, costume MATLAB scripts (MathWorks) and costume R scripts (R Development Core Team, 2011), STAN (Carpenter et al., 2017) for Bayesian model fitting; and Freesurfer (Dale et al., 1999), pysurfer (https://pysurfer.github.io) and costume Python scripts (http://www.python.org) for the surface generation and presentation. The code that was used for data analyses is available at https://github.com/gylab-TAU/multiple_objects_fMRI_analysis. The first three volumes in each run were acquired during a blank screen display and were discarded from the analysis as "dummy scans." The data were then preprocessed using realignment to the mean of the functional volumes and coregistration to the anatomic image (rigid body transformation). For the whole-brain analysis that was performed on data collected in experiment 2 across participants, spatial normalization to MNI space was applied. Otherwise, the data used for all other analyses remained in the subject's native space. Spatial smoothing was performed for the localizer data only (5 mm). A GLM was performed with separate regressors for each run and for each condition, including 24 nuisance motion regressors for each run (6 rigid body motion transformation, 6 motion derivatives, 6 square of motion, and 6 derivatives of square of motion), and a baseline regressor for each run. In addition, a

"scrubbing" method (Power et al., 2012) was applied for every volume with frame displacement >0.9 by adding a nuisance regressor with a value of 1 for that specific volume and zeros for all other volumes. The percentage signal change (PSC) for each voxel was calculated for each experimental condition in each run by dividing the β-weight for that regressor by the β-weight of the baseline for that run.

Data analysis - experiment 1

Region of interest analysis. Based on the functional localizer data, face- and body-selective voxels were defined individually for each subject using contrast *t*-maps. Regions of interest (ROIs) were defined as clusters (>10 voxels) of voxels selective for a given category ($p < 10^{-4}$) within the following specific anatomic locations: (1) fusiform face area (FFA): face>object within the fusiform gyrus; (2) fusiform body area (FBA): body>object within the fusiform gyrus. The overlap area was defined as the conjunction between face- and body-selective ROIs and included all voxels that were both face and body selective, as described above. The 30 most selective voxels from each ROI within the right hemisphere were analyzed with the main experiment data. ROIs with <30 voxels were excluded from further analysis. This criterion resulted in the following number of subjects that were included in the analysis for each ROI: FFA: $N = 13$; FBA: $N = 11$; overlap area: $N = 11$ (see Fig. 5 for the stability of the results across different numbers of voxels even with a very low number of subjects).

Linear model fitting. The mean PSCs across runs to the face, the body, and the face + body conditions from the main experiment data were extracted for each voxel within each ROI of each subject. For each subject and each ROI, we fitted a regression model for the response of the 30 most selective voxels to predict the response to the face + body based on the responses to the isolated face and the isolated body (i.e., the PSC) in each of these voxels, as follows:

$$(\text{face} + \text{body})_{PSC} = \beta_F^{(FB)} \cdot \text{face}_{PSC} + \beta_B^{(FB)} \cdot \text{body}_{PSC} + \varepsilon^{(FB)}. \quad (1)$$

The β-coefficients $\beta_F^{(FB)}$ and $\beta_B^{(FB)}$ indicate the contribution of the face and the body to the face + body response for each area and each subject (the β-coefficients of the multi-category response model are not the same as those derived from the standard fMRI GLM analysis. The β-coefficients from the standard fMRI GLM analysis are used to determine the PSC to each of the single-category and multi-category stimuli as a measure of the fMRI response to that stimuli). We calculated the mean of the β-coefficients of the model, the mean difference between the coefficients and their mean sum across subjects. To examine whether the linear model based on the normalization mechanism (Fig. 1*c*; Eq. 1) is the best fit to the data, we estimated a Bayesian hierarchical model to predict the response to a face + body based on the response to the face and the body including the data from all subjects for each ROI. In addition, we estimated two other Bayesian hierarchical models: one with an addition of an intercept term, and another with the addition of an interaction between the face and the body. We then calculated Bayes factors to compare the models.

Univariate voxelwise analysis. For each voxel within each ROI, we compared the PSC for the face + body to the maximum PSCs for the face and the body, and calculated the proportion of voxels that showed a smaller response to the face + body [i.e., face + body<max(face, body)]. This analysis was done to assure that the weighted mean response is not due to saturation of the BOLD response to face + body.

Searchlight analysis. For the searchlight analysis, we defined a face- and body-selective region based on the localizer data by the contrast [(face + body)/2 > object] ($p < 10^{-4}$) within the ventrotemporal and lateral occipital cortex. In addition, we defined the following two control areas: early visual cortex (EVC) and the parahippocampal place area (PPA). EVC was extracted by performing an inverse normalization from an MNI space Brodmann area 17 mask to each subject's native space. We matched the number of voxels in EVC to the number of voxels within the category-selective region for each subject by randomly choosing voxels from EVC. Because our functional localizer did not include scene images, the PPA was defined by using Neurosynth (Yarkoni et al., 2011; https://neurosynth.org), a meta-analysis tool for extracting cognitive maps. We used an association map with the term "Place" thresholded with a false discovery rate criterion of 0.01. We then masked the image to include only the right parahippocampal cortex. This image then underwent inverse normalization from an MNI space to each subject's native space. The Neurosynth-defined PPA included fewer voxels than the face- and body-selective areas, and therefore all voxels were included in the analysis. For each subject, we defined a moving mask of a sphere of 27 voxels. For each sphere, we fitted a linear model with its voxel data as features to predict the response to the face + body based on the response to the face and the body. The β-coefficients of these models represent the contribution of the face and the body to the response of the face + body of each sphere within the searchlight area. We then plotted a surface map of the β-coefficients of all spheres within the searchlight area to present the spatial distribution of the β-coefficients. We calculated the R^2 value for each sphere and the median R^2 across all spheres. Since the R^2 is calculated to models without intercept, it is possible to get a negative R^2 value (i.e., this model can be worse in predicting the dependent variable compared with a model with only an intercept).

To examine the relationship between the difference between the face and body β-coefficients and the selectivity for a face over a body (i.e., the *t* values of the contrast face>body from the independent functional localizer data), we performed a Pearson correlation across subjects. To assess the level of significance of the correlations, the correlation values were transformed to Fisher's *z*-scores, and a one-sample *t* test was used against a null hypothesis of zero. To reduce statistical dependency because of the overlapping moving mask, we used for the correlation analysis an interleaved mask, taking only spheres that their center is not immediately adjacent to another.

Data analysis - experiment 2

ROI analysis. Based on the functional localizer data, face-, body-, and object-selective voxels were defined individually for each subject. ROIs were defined as clusters (>10 voxels) of category-selective voxels ($p < 10^{-4}$) within specific anatomic locations that show preference to a single category relative to all other categories: (1) FFA: face > body, object, and scrambled object within the fusiform gyrus; (2) FBA: body > face, object, and scrambled object within the fusiform gyrus; and (3) ventral object area: object > face, body, and scrambled object within the medial part of the ventral temporal cortex. Note that we used a modified and well accepted (Peelen and Downing, 2005; Weiner and Grill-Spector, 2010, 2011) version of the ROI definitions relative to experiment 1 (e.g., FFA was defined in experiment 1 with the contrast face > object, as opposed to the current face>body, object, and scrambled object). This modified ROI definition was used to prevent a bias for the body relative to the wardrobe when comparing the face + body and face + objects pairs in areas that were defined by excluding only the object category and not the body category. This modification in the ROI definition results in the absence of an overlap between face-selective and body-selective areas. As in experiment 1, the 30 most selective voxels from each ROI in the right hemisphere were chosen for model fitting. ROIs with <30 voxels were excluded from further ROI analysis. This criterion resulted in the following number of subjects that were included in the analysis for each ROI: FFA: $N = 15$; FBA: $N = 14$; Object-selective area: $N = 13$.

The model fitting described in experiment 1 was used to separately predict the response to the face + body based on the response to the face and the body (Eq. 1) and to predict the response to the face + object based on the response to the face and the object using the following equation:

$$(\text{face} + \text{object})_{PSC} = \beta_F^{(FO)} \cdot \text{face}_{PSC} + \beta_O^{(FO)} \cdot \text{object}_{PSC} + \varepsilon^{(FO)}. \quad (2)$$

Similar to experiment 1, we calculated the β-coefficients of the model, the mean difference between the coefficients and their mean sum for each model across subjects.

To examine whether the pattern of response to face + body and face + object is different, we ran a repeated-measures ANOVA with pair type (face + body, face + object) and ROI (face-selective, body/object selective) as within-subject factors and the difference between the coefficients as a dependent variable. We excluded from this analysis subjects who did not have 30 voxels for all three ROIs (three subjects excluded).

Searchlight analysis. For the searchlight analysis, we defined a category-selective region based on the localizer data by the contrast [(face + body+object)/3 > scrambled object ($p < 10^{-4}$)] within the ventrotemporal cortex and lateral occipital-temporal cortex. A similar analysis that was performed in experiment 1 was performed separately to the face + body runs and the face + object runs.

Whole-brain analysis. To examine whether the relationship between category selectivity and the representation of multiple stimuli is indeed confined to category-selective cortex, we conducted a whole-brain analysis. For this analysis, data were spatially normalized to MNI space in addition to all other preprocessing steps. We performed the same searchlight analysis as described in the previous section for each subject over the whole brain. We used a parcellation based on functional connectivity and anatomy (Schaefer et al., 2018) to divide the brain to 400 parcels. For each parcel and each subject, we calculated a Pearson correlation between the difference in the contribution of the isolated stimuli to the multi-category stimulus of each model and the difference in category selectivity as described in the Searchlight analysis subsection in Materials and Methods. To assess the level of significance of the correlations, the correlation values were transformed to Fisher's *z*-scores, and a one-sample *t* test (one tailed) corrected for multiple comparisons was used to assess whether the correlation value averaged across participants was significantly higher than zero for each brain parcel.

Data availability

The code that was used for data analysis is available at https://github.com/gylab-TAU/multiple_objects_fMRI_analysis. Data that were collected in this study are available at https://openneuro.org.

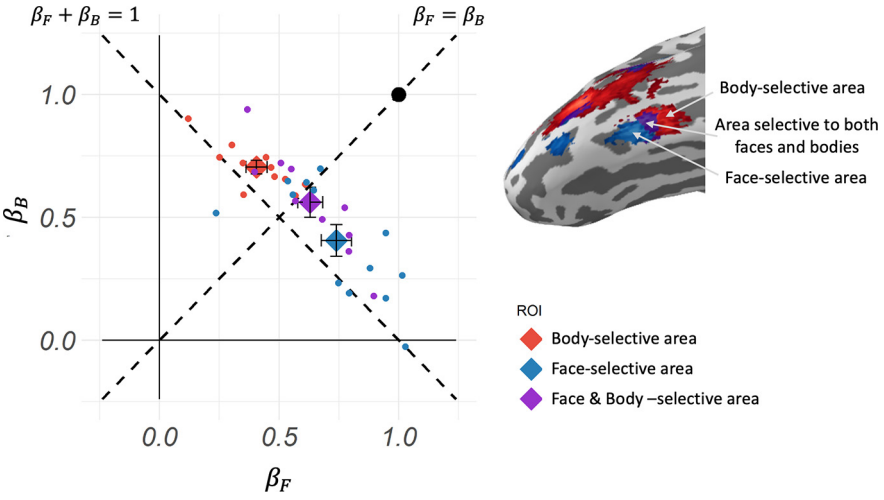

Figure 4. Experiment 1. Left, A scatterplot of the β-coefficients for the face and the body that best fit the response of the 30 most selective voxels within each subject's ROI to the face+body stimulus. Each dot indicates the results of a single subject within an ROI (in the right hemisphere). β_F indicates the contribution of the face to the face+body response, and β_B indicates the contribution of the body to the face+body response. The large diamonds indicate the group mean (error bars indicate the SEM). Right, A brain surface of one representative subject showing the location of the face-selective, body-selective, and the overlap areas in ventrotemporal cortex.

Results

Experiment 1—the representation of multi-category stimuli in category-selective areas

Experiment 1 was designed to test the prediction that the response to multi-category stimuli (face + body) is a weighted mean of the response to each of its components (a face and a body), and that the weights are determined by the magnitude of category selectivity for each of the stimuli and therefore vary continuously along category-selective cortex.

ROI analysis

First, we examined the contribution of the face and the body to the face + body response in the face- and body-selective areas. For each individual subject, we extracted the face-selective area (face>object), body-selective area (body>object), and the overlap between these areas (i.e., areas that are selective for both faces and bodies) using the independent functional localizer data (Fig. 4, example of these areas in a representative subject). For each subject and each area within the right ventrotemporal cortex, we fitted a linear regression model (Eq. 1) to estimate the contribution of the isolated face and body to the response to face + body, indicated by the β-coefficients $\beta_F^{(FB)}$ and $\beta_B^{(FB)}$, respectively. Figure 4 depicts the contribution of the face and the body to the response to the face + body as was derived based on the 30 most selective voxels of each subject's ROI (Fig. 5; similar findings with different numbers of voxels). All areas showed a significant contribution of both the face and the body to the face + body representation across all subjects, indicated by positive nonzero face and body coefficients (β-values = [0.39–0.74], all p values < 0.0001, all Cohen's d values > 1.754).

Based on derivations of the normalization model (Fig. 1), we can further predict that the difference between the coefficients will correspond to the degree of selectivity of a cortical area for the different parts. In other words, the face coefficient should be higher than the body coefficient in face-selective areas, and vice versa for body-selective areas (Fig. 1*d*). Results were consistent with this prediction. We found that in the FFA, which is composed of mainly face-selective neurons, the contribution of the face was larger than the contribution of the body [$\beta_F^{(FB)} - \beta_B^{(FB)}$: mean = 0.334; $t_{(12)}$ = 2.846; p = 0.015; 95% confidence interval (CI) = 0.078, 0.590; Cohen's d = 0.789]. Conversely, in the FBA, which is composed of mainly body-selective neurons, the contribution of the body was larger than the contribution of the face [$\beta_F^{(FB)} - \beta_B^{(FB)}$: mean = −0.298; $t_{(10)}$ = −4.358; p = 0.001; 95% CI = −0.451, −0.146; Cohen's d = 1.314]. In the area of overlap between the FFA and the FBA, which is selective for both faces and bodies, there was no significant difference between the contribution of the face and the body [$\beta_F^{(FB)} - \beta_B^{(FB)}$: mean = 0.070; $t_{(10)}$ = −0.628; p = 0.544; 95% CI = −0.177, 0.316; Cohen's d = 0.189].

Consistent with our predictions (Fig. 1*e*), we found that the sum of the β-coefficients was slightly >1 [mean sum (SEM): FFA: 1.145 (0.049); FBA: 1.110 (0.028); overlap: 1.191 (0.024)]. Note that our model did not limit the sum of the coefficients to 1, but they could take any value. In addition, the response to the face + body is more consistent with a weighted mean response rather than an additive response, as indicated by the coefficients being <1 (all p values <0.01, all Cohen's d values >1.144), and the sum of these coefficients is <2 (all p values <0.001, all Cohen's d values >4.815]. Finally, we rule out an alternative explanation that the weighted mean response is due to saturation of the BOLD response to multiple stimuli. We found that 53.24%

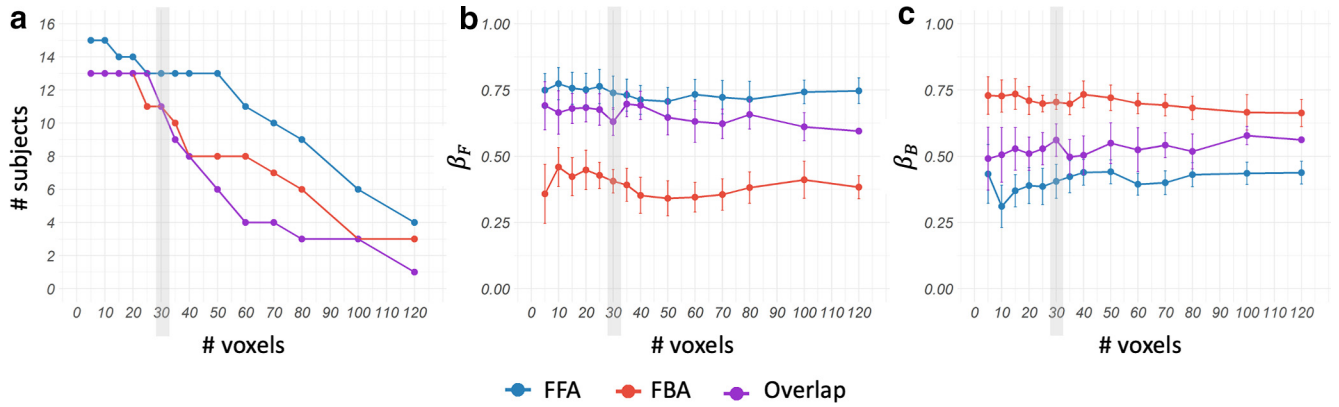

Figure 5. ROI analysis across different number of voxels. The analysis reported is based on 30 voxels for each ROI (marked in gray). ***a***, The number of subjects across different sizes of category-selective ROIs. As the size of the ROI increases, the number of subjects decreases. ***b***, ***c***, Mean β_F (***b***) and mean β_B (***c***) across subjects for each ROI size. Error bars indicate the SEM. These data indicate that the results are highly stable across different ROI sizes and number of subjects, even when analysis includes very small sample sizes.

Table 1. Experiment 1—model comparison

	Comparing models with and without intercept (BF)	Comparing models with and without interaction (BF)
FFA	$2.14 * 10^5$	$1.94 * 10^5$
FBA	$3.45 * 10^7$	$5.36 * 10^4$
Overlap	6.75	$1.15 * 10^4$

To compare the proposed model predicted by the normalization equation (Fig. 1*c*) to other models across all subjects, we used a Bayesian hierarchical model to predict the representation of the face+body stimulus based on the response to the face and the body. For each area, we fitted three models (face and body; adding an intercept; adding an interaction). Values in the table indicate the Bayes factor (BF) for the comparison between the model with only face and body factors to the other models, showing that this model best explains the results within all ROIs.

of the voxels in our data (FFA: 53.33%; FBA: 58.48%; overlap: 47.88%) showed a higher response to one of the single stimuli (a face or a body) relative to the response to the combined stimulus (face + body).

To further assess whether the weighted mean model (i.e., the normalization model; Fig. 1*c*) is the best fit to the data, we compared this model to two other models—one model with a nonzero intercept and another model with an interaction between the face and the body (i.e., a nonlinear relationship between the isolated components and the multi-category stimulus). We found that the model that best explains our results is a linear model with only the face and the body as predictors (Table 1).

Searchlight analysis

Next, we assessed the contribution of the face and the body to the face + body representation along the face and body areas within the right occipitotemporal and lateral-occipital areas. For each individual subject, we measured the response to face, body, and the face + body stimuli of each voxel in these anatomic locations. We then applied a moving mask of a sphere of 27 voxels. For each sphere, we fitted a linear model to the responses of the voxels within the sphere to predict the response to the face + body based on the responses to the face and the body (Fig. 1*c*).

Figure 6, *a* and *b*, depicts the β-coefficients for the face and the body (i.e., the contribution of the face and the body to the face + body response in the face- and body-selective area of a single subject placed on a surface map of his brain). Figure 6, *c* and *d*, shows the distribution of category selectivity for the same subject within the same region for the face and the body, as indicated by the independent functional localizer data. Overall, Figure 6 demonstrates the correspondence between the selectivity and the contribution of the face and the body to the face + body representation throughout the continuum of the face- and body-selective regions: areas with high selectivity for faces and low selectivity for bodies show high contribution of the face to the face + body representation, while areas with low selectivity for faces and high selectivity for bodies show high contribution of the body to the face + body representation.

Figure 7*a* depicts the β-coefficients for the face and the body (i.e., the contribution of the face and the body to the face + body response) of all spheres within the face- and body-selective cortices in the right occipitotemporal and lateral areas of all subjects. The coefficients are scattered along the weighted mean line, indicating a sum of coefficients that is slightly >1 (mean sum = 1.071; 95% CI = 1.036, 1.106), which is consistent with the derivations based on the normalization model (Fig. 1*e*). Figure 7*d* displays the distribution of R^2 of the models for all spheres, indicating a good fit of the linear model to the data (median R^2 = 0.90). The color of each dot indicates the selectivity for the face relative to the body, as measured by the independent functional localizer. Furthermore, consistent with our predictions (Fig. 1*d*), the difference between the contribution of the face and the body to the face + body representation (i.e., the difference between the β-coefficients) is correlated with the face and body selectivity as measured by the independent functional localizer data. To examine the statistical significance of this correlation, the correlation was computed for each subject and transformed to a Fisher's *z*-score, and the mean across subjects was compared with a null hypothesis of a correlation <0 [mean $r = 0.446$; $t_{(14)} = 9.653$; $p < 0.0001$ (one tailed), 95% CI = 0.373, 0.513; Cohen's d = 0.479).

To examine whether the correspondence between category selectivity and the representation of multiple stimuli is restricted to areas that are selective for the stimulus components, we performed a similar searchlight analysis over the following two control areas: EVC (Fig. 7*b*,*e*) and the PPA (Fig. 7*c*,*f*). EVC is sensitive to low-level features of the stimuli, but not to high-level categories. PPA is a part of high-level visual cortex but is selective for places and not for the categories included in the stimuli of this experiment. Results show that the linear model does not fit as well in the EVC and PPA when compared with the face- and body-selective areas, as indicated by the R^2 distributions (median R^2: EVC = 0.722; PPA = 0.487). Moreover, the sum of β-coefficients is slightly <1 (EVC: mean sum = 0.949, 95% CI = 0.897, 1.002; PPA: mean sum = 0.813, 95% CI = 0.739, 0.887), indicating a lower fit to the normalization model predictions. Furthermore, the difference between the contribution of the face and the body to the face + body representation (i.e., the difference between

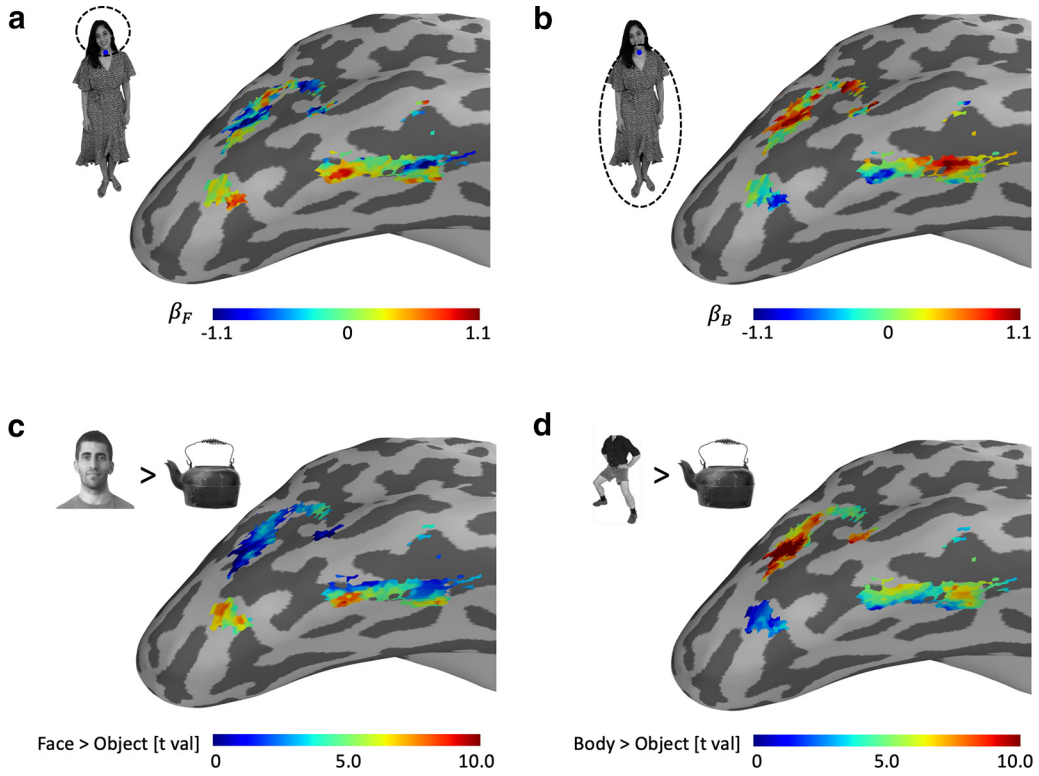

Figure 6. Experiment 1. Results of a representative subject plotted on the cortical surface for voxels that were selective for either faces or bodies. ***a***, The contribution of the face to the face+body representation, as indicated by the face regression coefficients (β_f). ***b***, The contribution of the body to the face+body representation, as indicated by the body regression coefficients (β_b). ***c***, Selectivity for faces (*t*-map of face>object). Selectivity was determined based on independent functional localizer data. ***d***, Selectivity for bodies (*t*-map of body>object). Selectivity was determined based on independent functional localizer data.

the β-coefficients) is not positively correlated with the face and body selectivity as measured by the independent functional localizer data in EVC [mean $r = -0.131$; $t_{(14)} = -3.240$; $p = 0.997$ (one-tailed); 95% CI = −0.201, −0.060; Cohen's d = 0.132] and shows a much lower positive correlation in PPA [mean $r = 0.094$; $t_{(14)} = 1.872$; $p = 0.041$ (one-tailed); 95% CI = 0.006, 0.181; Cohen's d = 0.094]. To directly compare the ROIs, we ran a repeated-measures ANOVA with ROI (face- and body-selective areas, EVC, and PPA) as a within-subject factor and the correlation values (after Fisher's *z*-transformation) as a dependent variable. We found a significant effect for ROI indicating a difference in the correlations between the areas ($F_{(2,28)} = 38.354$; $p < 0.0001$; $\eta^2_G = 0.672$). Thus, the relationship between category selectivity and the contribution of the face and the body to the face + body response was not found in control areas that are not selective for these categories.

Experiment 2—the representation of related and unrelated multi-category stimuli in category-selective areas

Experiment 2 was designed to test whether the correspondence between category selectivity and multi-category representation that we found in experiment 1 applies also to nonrelated pairs of stimuli.

ROI analysis

First, we ran the same analysis reported above to examine the contribution of the face and the body to the face + body response in a face-selective area. We first defined the ROIs in a manner similar to the way they were defined in experiment 1 (FFA: face>object; FBA: body>object, including an overlap area) to assure that we replicate the same findings. Results showed similar findings [FFA: $\beta_F^{(FB)} - \beta_B^{(FB)}$: mean = 0.510, $t_{(14)} = 4.318$, $p < 0.001$, 95% CI = 0.257, 0.7763, Cohen's d = 1.115; FBA: $\beta_F^{(FB)} - \beta_B^{(FB)}$: mean = −0.412, $t_{(12)} = -3.198$, $p = 0.008$, 95% CI = −0.693, −0.131, Cohen's d = 0.887; overlap area: $\beta_F^{(FB)} - \beta_B^{(FB)}$: mean = 0.151, $t_{(10)} = 2.060$, $p = 0.066$, 95% CI = –0.012, 0.315, Cohen's d = 0.621]. To compare between the face + body and face + object findings, in experiment 2 we used a definition of the ROIs modified from the definition used in experiment 1, where each category was subtracted from all other categories (FFA: face>object, body and scrambled object; FBA: body>object, face, and scrambled object) to prevent a bias toward one of the categories (see Materials and Methods). This definition excludes the face–body overlap areas, but still replicates the results of experiment 1 in face- and body-selective areas (Fig. 4*a*), with both the face and the body contributing to the response of the face + body stimulus [$\beta_F^{(FB)}$ and $\beta_B^{(FB)}$ of both FFA and FBA >0, all *p* values <0.001 (except for $p = 0.002$ for $\beta_B^{(FB)}$ in FFA), all Cohen's d values >0.984; Fig. 8*a*]. Furthermore, the relative contribution of the face and the body varied as a function of the face and body selectivity (Fig. 1*d*), replicating the results of experiment 1: in the FFA the contribution of the face was higher than the contribution of the body [$\beta_F^{(FB)} - \beta_B^{(FB)}$: mean = 0.494; $t_{(14)} = 4.169$; $p < 0.001$; 95% CI = 0.240, 0.747; Cohen's d = 1.076], while in the FBA the contribution of the body was higher than the contribution of the face [$\beta_F^{(FB)} - \beta_B^{(FB)}$: mean = −0.382; $t_{(13)} = -3.442$; $p = 0.004$; 95% CI = −0.622, −0.142; Cohen's d = 0.920]. The sum of coefficients in both face and body areas was again slightly over 1 [mean sum (SEM): FFA: 1.042 (0.066); FBA: 1.098 (0.054)] consistent with our model (Fig. 1*e*).

Next, we performed similar analyses for the face + object stimuli. For each subject we fitted a regression model for the 30

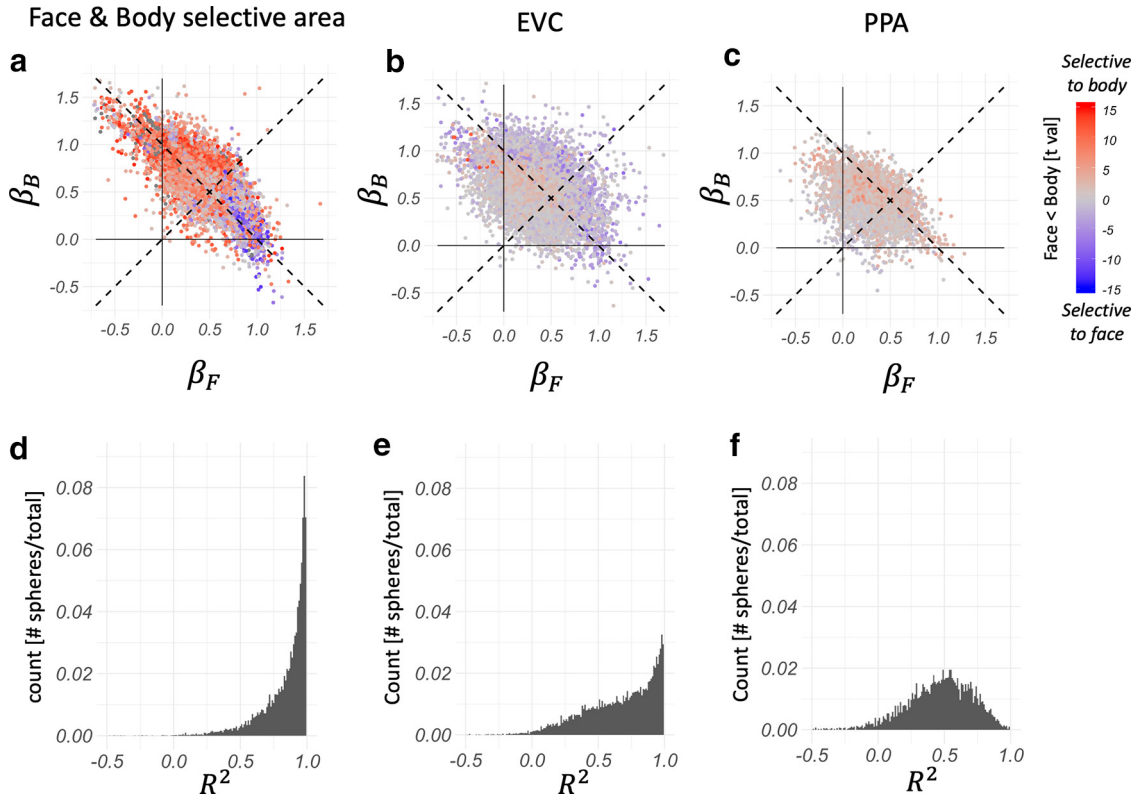

Figure 7. Experiment 1. ***a–c***, The β-coefficients of all spheres of all subjects in a region of interest indicating the contribution of the face (β_F) and the body (β_B) to the response to the face+body (Eq. 1). The color of each dot indicates the selectivity for the face relative to the body based on independent functional localizer data [face- and body-selective area (***a***); EVC (***b***); and PPA (***c***)]. ***d–f***, Histograms of the R^2 values of the linear models accounting for the response to the face+body of all spheres [negative values can be observed for models without intercept; see Materials and Methods data; face- and body-selective area (***d***); EVC (***e***); PPA (***f***)].

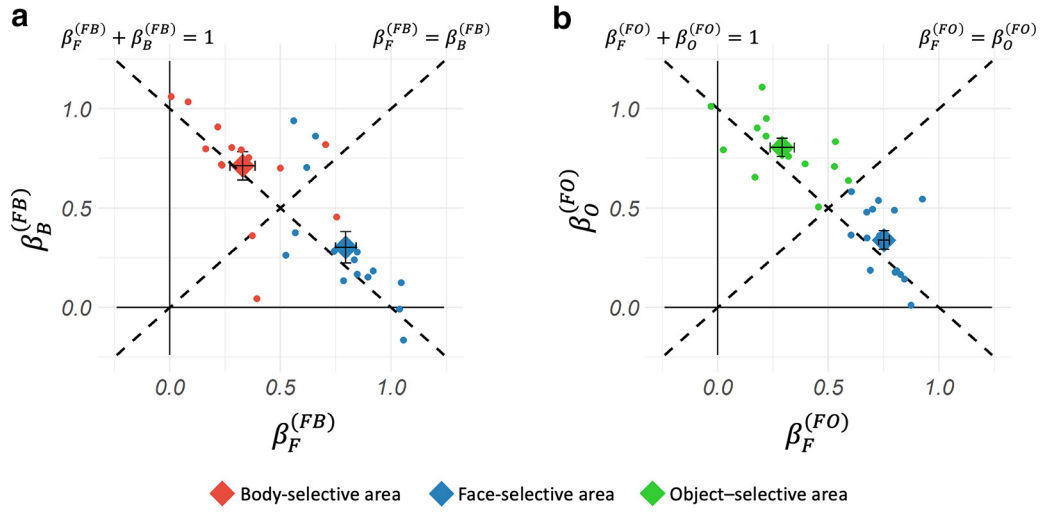

Figure 8. Experiment 2. ***a***, β-Coefficients for the face and the body predicting the response of the 30 most selective voxels within each subject's ROIs to the face+body stimulus. $\beta_F^{(FB)}$ is the contribution of the face to the face+body response, and $\beta_B^{(FB)}$ is the contribution of the body to the face+body response. Each dot indicates the results of a single subject within an ROI. The large diamonds indicate the group mean (error bars indicate the SEM). ***b***, β-Coefficients for the face and the object predicting the response of the 30 most selective voxels within each subject's ROIs to the face+object stimulus. $\beta_F^{(FO)}$ indicates the contribution of the face to the face+object response, and $\beta_O^{(FO)}$ indicates the contribution of the object to the face+object response. Each dot indicates the results of a single subject within an ROI. The large diamonds indicate the group mean (error bars indicate the SEM).

most selective voxels within the face-selective area (face > object, body, and scrambled object) and the object-selective area (object > face, body, and scrambled object) to predict the response to the face + object based on the responses to the face and the object (Eq. 2). Similar to the face + body findings, the face- and object-selective areas showed a significant contribution of both the face and the object to the face + object representation across all subjects, indicated by positive, nonzero coefficients of both the face and the object [$\beta_F^{(FO)}$ and $\beta_O^{(FO)}$ of both FFA and object-selective area >0, all p values <0.001, all Cohen's d values >1.266; Fig. 8*b*]. In addition, the selectivity of the area determined the relative contribution of the face and the object to the

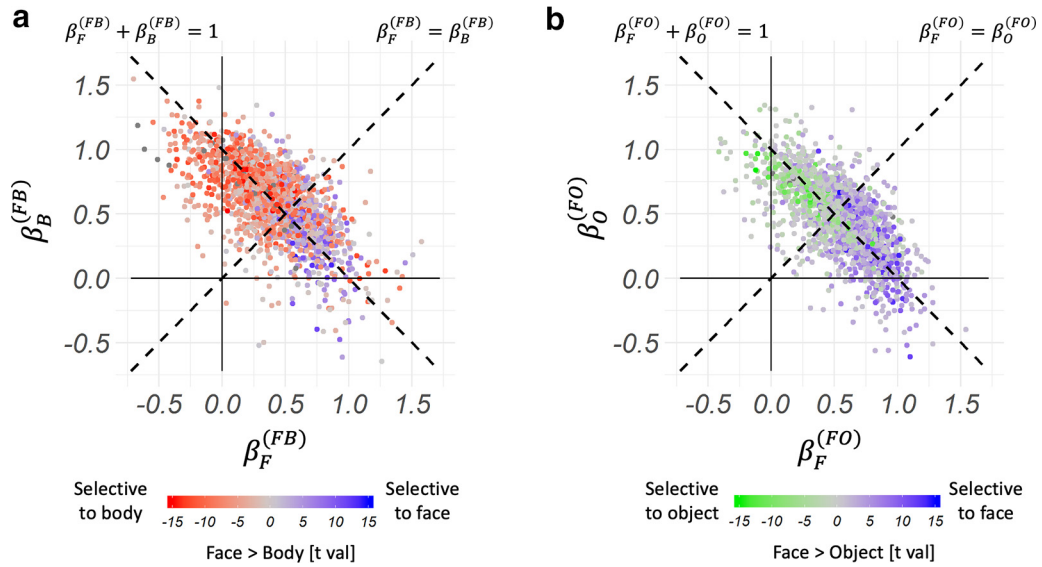

Figure 9. Results of a searchlight analysis in experiment 2. ***a***, The β-coefficients of all spheres in the category-selective cortices of all subjects indicating the contribution of the face ($\beta_F^{(FB)}$) and the body ($\beta_B^{(FB)}$) to the response to the face+body (Eq. 1). The color of each dot indicates the selectivity for the face relative to the body based on independent functional localizer data. There was a positive correlation between category selectivity (face>body) and difference between β-coefficients ($\beta_F^{(FB)} - \beta_B^{(FB)}$). ***b***, The β-coefficients of all spheres in the category-selective cortices (same as ***a***) of all subjects, indicating the contribution of the face ($\beta_F^{(FO)}$) and the object ($\beta_O^{(FO)}$) to the response to the face+object (Eq. 2). The color of each dot indicates the selectivity for the face relative to the object based on independent functional localizer data. There was a positive correlation between category selectivity (face>object) and difference between β-coefficients ($\beta_F^{(FO)} - \beta_O^{(FO)}$).

face + object representation (Fig. 1*d*). Specifically, we found that in the FFA, which is mainly selective for faces, the contribution of the face was higher than the contribution of the object [$\beta_F^{(FO)} - \beta_O^{(FO)}$: mean = 0.413; $t_{(14)} = 6.737$; $p < 0.001$; 95% CI = 0.282, 0.545; Cohen's d = 1.740], while in the object-selective area, the contribution of the object was higher than the contribution of the face [$\beta_F^{(FO)} - \beta_O^{(FO)}$: mean = −0.512; $t_{(12)} = -5.753$; $p < 0.001$; 95% CI = −0.706, −0.318; Cohen's d = 1.596]. The sum of coefficients, again, was slightly >1, which is consistent with our model [Fig. 1*e*; mean sum (SEM): FFA: 1.090 (0.043); object area: 1.096 (0.047)].

The face + body stimuli are different from the face + object stimuli in that the former are a familiar combination, whereas the latter are not. Previous studies have predicted different patterns of representations to familiar than nonfamiliar object combinations (Song et al., 2013; Baldassano et al., 2016; Kaiser and Peelen, 2018), whereas others did not find such a difference (Baeck et al., 2013; Kaiser et al., 2014). To examine whether the patterns of response to face + body and face + object are different, we ran a repeated-measures ANOVA with pair type (face + body, face + object) and ROI (face-selective, body/object selective) as within-subject factors, and the difference between the coefficients as a dependent variable. We excluded from this analysis subjects who did not have 30 voxels in each of the three ROIs (three subjects). As expected, the main effect of the ROI was significant ($F_{(1,11)} = 54.382$, $p < 0.0001$), indicating that the selectivity of the ROI accounts for the relative contribution of each of the single categories to their multi-category stimuli. Importantly, we found no support for differences between pair type ($F_{(1,11)} = 1.361$, $p = 0.268$, $\eta_G^2 = 0.030$), as well as no interaction between the ROI and pair type ($F_{(1,11)} = 0.024$, $p = 0.808$, $\eta_G^2 = 0.0003$). Thus, the same normalization framework accounts for the two types of multi-category stimuli.

Searchlight analysis

A searchlight analysis similar to that described in experiment 1 was performed for the face + body (Eq. 1) and the face + object (Eq. 2) stimuli in ventrotemporal and lateral-occipital areas that are selective for faces, bodies, or objects relative to scrambled objects (i.e., category-selective cortex). Figure 9*a* depicts the β-coefficients for the face and the body (i.e., the contribution of the face and the body to the face + body response of all spheres within the category-selective cortices of all subjects). Although this area also contains voxels that are selective for objects, the results are similar to experiment 1. Specifically, the difference in the contribution of the face and the body to the face + body representation (i.e., the difference between the β-coefficients) is positively correlated with the selectivity for the face relative to the body as predicted [mean $r = 0.386$, $t_{(14)} = 8.444$, $p < 0.0001$ (one-tailed), 95% CI = 0.312, 0.456, Cohen's d = 2.180], and the sum of coefficients is slightly >1 (mean sum = 1.013, 95% CI = 0.970, 1.056), replicating the results of experiment 1.

We performed the same analysis for the face + object model over the same searchlight area and found results similar to the face + body findings (Fig. 9*b*): The β-coefficients are scattered along the weighted mean line with a sum of coefficients that is slightly >1 (mean sum = 1.015, 95% CI = 0.993, 1.038), and the difference in the contribution of the face and the object to the face + object representation (i.e., the difference between the coefficients) is correlated with the selectivity for the face relative to the object as expected [mean $r = 0.395$, $t_{(14)} = 11.193$, $p < 0.0001$ (one tailed), 95% CI = 0.338, 0.449, Cohen's d = 2.890; Fig. 1*d*,*e*].

To compare the spatial distribution of the β-coefficients and category selectivity, we plotted the difference between the coefficients and the difference between the selectivity for each pair of categories on brain surface maps of one representative subject along his category-selective cortex (Fig. 10*a*–*d*). Figure 10*a* shows the difference between the face and body coefficients (i.e., difference between the contribution of the face and the contribution of the body to the face + body representation). Figure 10*b* shows the selectivity for the face relative to the selectivity for the body as measured by the independent functional localizer data. It can be seen that cortical areas that show a higher contribution

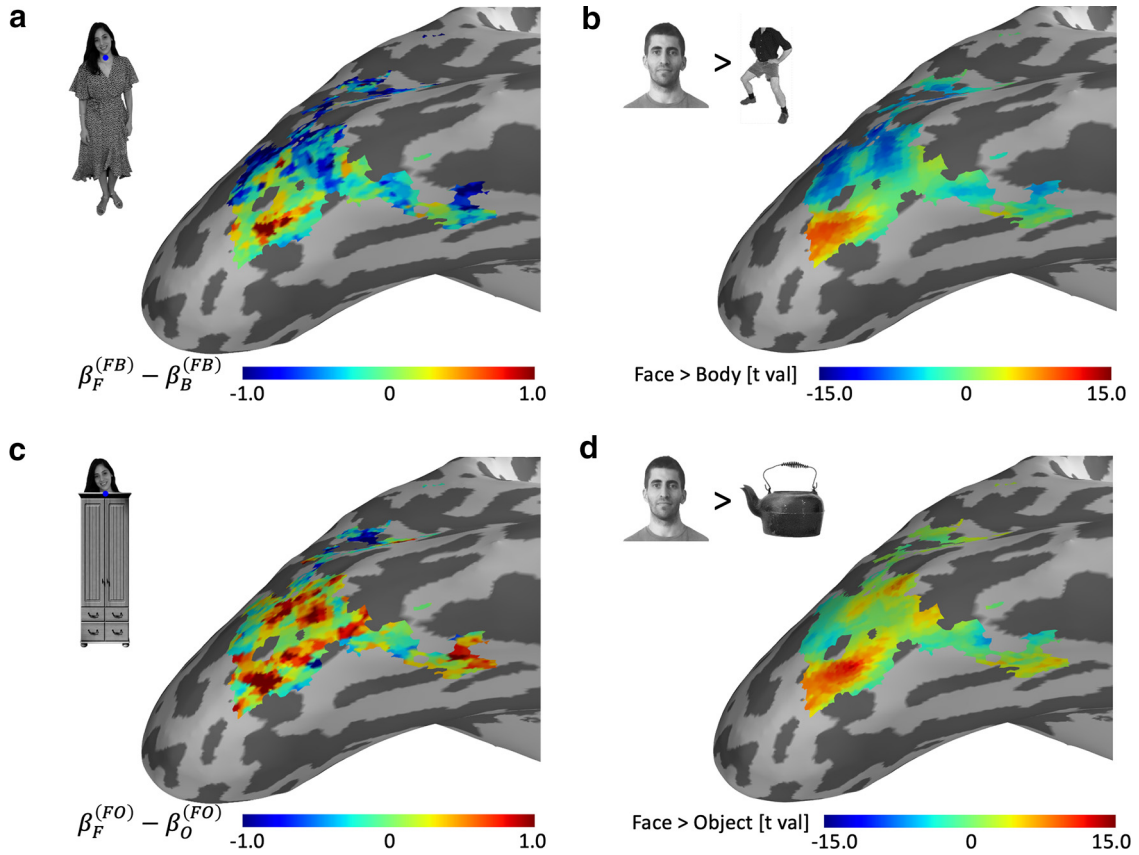

Figure 10. Experiment 2. Results of searchlight analysis of one representative subject plotted on the cortical surface show the correspondence of the difference between the coefficients of the two categories and the magnitude of their selectivity in category-selective cortex. Note that Figure 3 shows a map of the coefficients, and here we show a map of the difference between the coefficients. ***a***, The difference between the contribution of the face and the body to the face+body representation, as indicated by the difference between the regression coefficients. A larger difference corresponds to a higher contribution of the face than the body to the representation of the face+body stimulus. ***b***, Selectivity for faces relative to bodies (*t*-map of face>-body). Selectivity was determined based on independent functional localizer data. ***c***, The difference between the contributions of the face and the object to the face+object representation, as indicated by the difference between the regression coefficients. A larger difference corresponds to a higher contribution of the face than the object to the representation of the face+object stimulus. ***d***, Selectivity for faces relative to objects (*t*-map of face<object) based on independent functional localizer data.

of the face to the face + body representation correspond to face-selective clusters (Fig. 10*a*,*b*, red), and that areas that show a higher contribution of the body to the face + body representation correspond to body-selective clusters (Fig. 10*a*,*b*, blue). Figure 10*c* shows the difference between the contribution of the face and the object to the face + object representation for the same category-selective area. Figure 10*d* shows the selectivity for the face relative to the object based on the functional localizer data. Similar to the face + body results, areas that show a higher contribution of the face to the face + object representation correspond to face-selective clusters (Fig. 10*c*,*d*, red), and areas that show a higher contribution of the object to the face + object representation correspond to object-selective clusters (Fig. 10*c*,*d*, blue).

Whole-brain analysis

To reveal whether the correspondence between category selectivity and multi-category representation is a unique property of category-selective visual cortex, we performed an unconstrained whole-brain searchlight analysis, similar to the searchlight analysis described in the previous section. We used a parcellation of 400 parcels (Schaefer et al., 2018), and for each parcel and each subject we calculated the Pearson correlation between category selectivity and the difference between the β-coefficients in our model. Figure 11, *a* and *c*, depicts the correlation for each parcel of the right hemisphere for the face + body and the face + object models, respectively, averaged across subjects (after Fisher's *z*-transformation). Figure 11, *b* and *d*, depicts parcels that show significant correlation across subjects for the two models (one-tailed *t* test with $N = 15$, $p < 0.05$ corrected for multiple comparisons). Only parcels within high-level visual cortex (ventrotemporal and lateral-occipital areas) showed significant correlations. Moreover, the pattern of correlations is different for the face + body model and the face + object model. The ventromedial areas, which are typically selective for inanimate stimuli show a positive correlation for the face + object model but not for the face + body model, further indicating the correspondence between components of the multi-category stimuli and the selectivity for its components.

Discussion

The current fMRI study demonstrated a remarkable correspondence between the spatial distribution of category selectivity and the representation of multi-category stimuli across high-level, category-selective cortex (Figs. 5-11). We further showed that this correspondence is restricted to category-selective visual cortex (Figs. 7, 11). Consistent with our predictions (Fig. 1), we found that the relative contributions of each category (i.e., the model coefficients) to the multi-category response are determined by the magnitude of category selectivity in a given cortical area, and therefore vary across different areas of category-

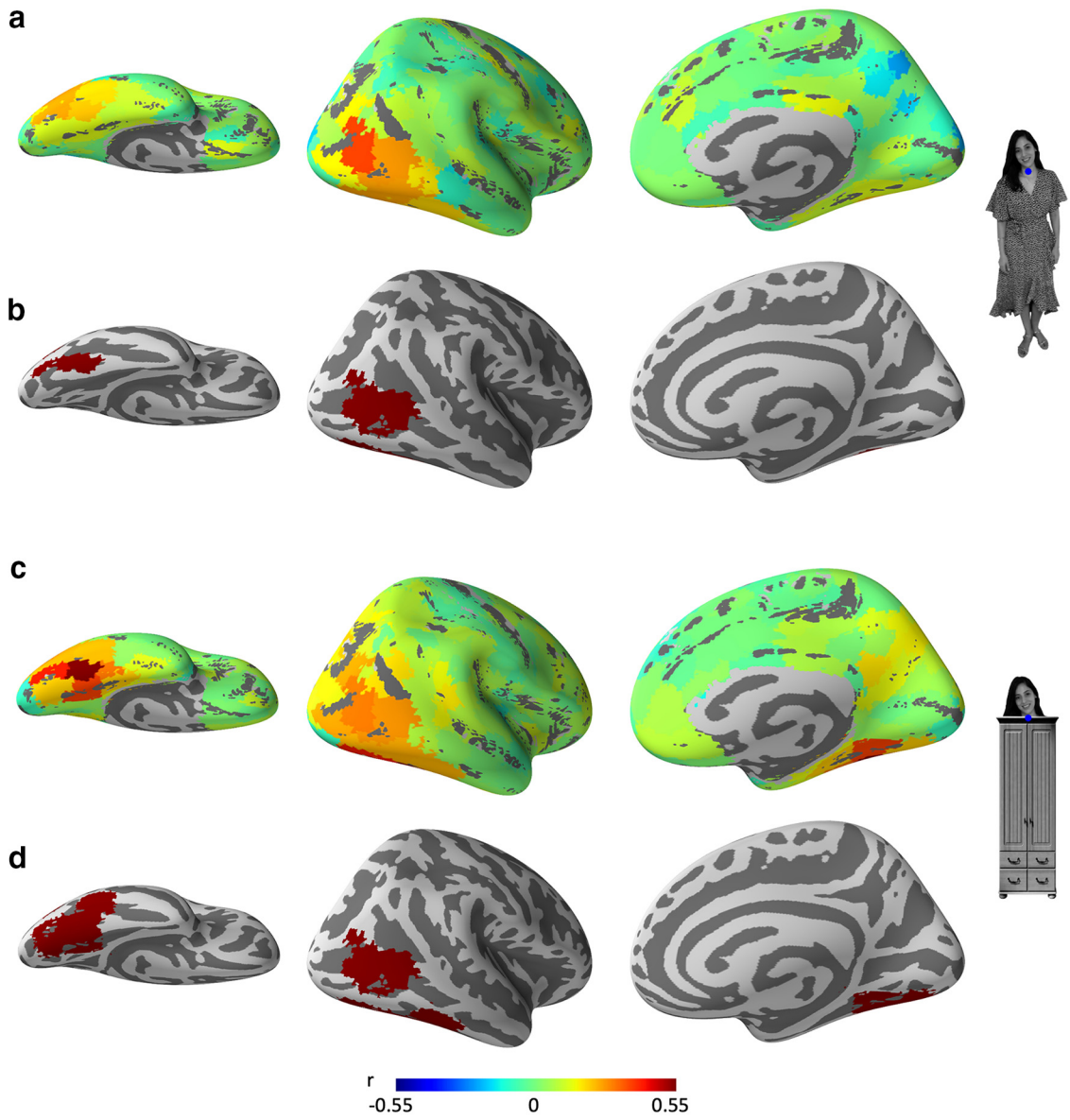

Figure 11. Experiment 2. Results of whole-brain searchlight analysis. ***a***, Correlation of the differences between the contributions of the face and the body to the face+body representation and difference between category selectivity (face>body of localizer data) in each parcel, averaged across subjects. ***b***, Parcels showing significant positive correlation as described in ***a*** (one-tailed *t* test across subjects of Fisher's *z*-transformed correlations, $p < 0.05$ corrected for multiple comparisons of 400 parcels). ***c***, Correlation of the differences between the contributions of the face and the object to the face+object representation and the difference between category selectivity (Face>Object of localizer data) in each parcel, averaged across subjects. ***d***, Parcels showing significant positive correlation, as described in ***c*** (one-tailed *t* test across subjects of Fisher's *z*-transformed correlations, $p < 0.05$ corrected for multiple comparisons of 400 parcels).

selective cortex. These findings are consistent with a normalization mechanism (MacEvoy and Epstein, 2009; Reddy et al., 2009; Bao and Tsao, 2018) but go beyond previous reports in the following ways: (1) by showing that the representations of multi-category stimuli are determined by the category selectivity for their component stimuli, we provide a general framework for the various findings reported in previous studies that showed either a mean or a maximum response in different areas of category-selective cortex; (2) by using fMRI, we can show this principle of operation across a large, continuous region of category-selective visual cortex and that it is restricted to this cortical region; and (3) we found that this weighted linear model accounts for the representations of both related (face + body) and nonrelated (face+wardrobe) stimuli.

Our findings are consistent with a recent single-unit recording study (Bao and Tsao, 2018) that proposed that the response of a neuron to multi-category stimuli may vary as a function of the homogeneity of category selectivity of the surrounding neurons. If the surrounding neurons are selective for the same category (i.e., homogeneous normalization pool) as the recorded neuron (i.e., a face neuron in a face-selective area), the normalization pool is unresponsive to the nonpreferred stimulus and therefore does not reduce the response of the recorded neuron to its preferred stimulus, yielding a maximum response. Thus, areas with a high concentration of neurons selective for a single category give priority to the preferred stimulus, filtering out the nonpreferred stimuli, resulting in a maximum response (Reddy et al., 2009; Bao and Tsao, 2018; Fig. 1*b*). This operation enables hard-

wired decluttering at early stages of visual processing (Bao and Tsao, 2018) in category-selective areas. In contrast, in areas with a mixed population of category-selective neurons, the surrounding neurons respond to the nonpreferred stimuli, yielding similar, possibly competitive, representations of different categories, resulting in a mean response. By generating a response to multiple stimuli that ranges from a mean to a maximum response, the normalization mechanism keeps the neuronal response within the dynamic range, preventing saturation of the neural response (Carandini and Heeger, 2011). The fMRI results reported in the current study add to the neuronal findings by demonstrating the correspondence between the functional organization of high-level visual cortex and the representation of multi-category stimuli across a large area of cortex with varying degrees of category selectivity that cannot be obtained in neurophysiological studies. This is enabled by the following two features of the fMRI signal: first, the magnitude of category selectivity measured with fMRI provides a measure of the homogeneity of the normalization pool, an important factor in the representation of multiple categories as derived from the normalization equation (Fig. 1); and second, fMRI enables exploring the pattern of response across a large, continuous area of cortex with different mixtures of category-selective neurons. This pattern of response indicates that the representation of the multi-category stimulus changes gradually in a way that corresponds to the profile of category selectivity (Figs. 6, 7, 9, 10, 11). These results propose a continuous mode of organization of high-level visual cortex, rather than the more common, discrete-like depiction of category-selective cortex.

Nevertheless, fMRI cannot determine whether the response of neurons to a face and a body in the overlap area that is selective for both faces and bodies reflects neuronal saturation of neurons that are selective for either a face or a body, or a mean response of two populations of face-selective and body-selective neurons. Based on single-unit recording studies, we believe that the latter alternative is more likely. First, Bao and Tsao (2018) showed that the response of face-selective neurons to two simultaneously presented faces is the mean response to the two isolated faces, indicating no evidence for neuronal saturation. Thus, even if neurons that are selective for either faces or bodies exist, they are more likely to show a mean response to a face and a body rather than to neuronal saturation. Second, the normalization mechanism functions as a "gain control" mechanism, preventing neurons from reaching saturation even when presented with more than one preferred stimulus (Carandini and Heeger, 2011).

Previous neuroimaging and single-unit recording studies reported mixed findings of a mean response (Zoccolan et al., 2005; MacEvoy and Epstein, 2009), a weighted mean response (Baeck et al., 2013), or a maximum response (Reddy et al., 2009; Bao and Tsao, 2018) to multiple stimuli in different areas of category-selective cortex. Our study proposes a general framework that accounts for these various findings by showing that the representation of multiple stimuli vary across high-level visual cortex as a function of the category selectivity in different cortical regions. Other neuroimaging studies that examined the representation of multiple stimuli have asked whether the response to a pair of stimuli deviates from a simple mean model, in particular for pairs of stimuli that show a meaningful relationship between them (MacEvoy and Epstein, 2011; Song et al., 2013; Kaiser et al., 2014; Fisher and Freiwald, 2015; Baldassano et al., 2016; Kaiser and Peelen, 2018). In these studies, a deviation from a simple mean response was considered as evidence for integration or a holistic representation of the complex stimulus. The main advantage of the linear model we used here is that it provides us with a direct measure of the type of deviation from the mean that the data show and can therefore decide between a weighted mean response, an additive response, or a nonadditive response. Our findings show that the deviation from the mean reflects a weighted mean response. We found no evidence for a nonadditive response to the combined stimulus and therefore no support for a holistic representation. This was the case both for the meaningful pair of face + body stimuli as well as for the nonmeaningful face+wardrobe pair that generated similar representations. Similar results were reported by Baeck et al. (2013), who found the same representations for related and unrelated pairs of objects. Thus, the normalization mechanism operates in a similar manner for related and unrelated pairs of stimuli in object category-selective cortex. Finally, although we refer to the model as a weighted mean model (i.e., sum of weights of 1), derivations of the normalization model, as detailed in Figure 1, predict that the sum of coefficients will be slightly >1. Indeed, our results reveal that the sum of the coefficients is slightly >1, which is consistent with predictions of the normalization model as well as with previous findings (Reddy et al., 2009).

Three additional studies that examined the representation of the whole person are noteworthy. Kaiser et al. (2014) reported no deviation from the mean in the response to a face and a body in a person-selective area (area defined by a whole person>objects). This area is likely to correspond to the overlap area reported in our study that is selective for both faces and bodies, and therefore is consistent with our findings (Fig. 4). Song et al. (2013) reported that only the right FFA showed a deviation from the mean for the response of the whole person and interpreted that as evidence for a holistic representation. This deviation, however, may reflect a weighted mean response rather than a nonadditive response. Finally, Fisher and Freiwald (2015) examined the contribution of the face and body to the whole person in a monkey fMRI study and found a superadditive (more than the sum) response in anterior but not posterior face areas, in particular, in area AF in the dorsal bank of the superior temporal sulcus. The human analog of area AF is likely to be in the superior temporal sulcus (Yovel and Freiwald, 2013), an area that we did not examine in the current study that may apply a different mode of operation than the ventral visual cortex (see also Baldassano et al., 2016).

To summarize, our findings reveal a general framework of operation according to which the contribution of each stimulus to the representation of multiple stimuli in a given cortical area is determined by its profile of category selectivity. We therefore suggest that the functional organization of neighboring patches of neurons, each selective for a single or more categories, enables a flexible representation of complex visual scenes, where both decluttering and competition operate in different cortical areas, using the same type of neurons and the same mechanism of normalization. This type of organization may permit high-level cognitive processes to bias the response to any of these different representations according to task demands (Desimone and Duncan, 1995; Reynolds and Heeger, 2009), making the taxing operation of understanding complex visual scenes dynamic and flexible.

References

Baeck A, Wagemans J, de Beeck HP (2013) The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: the weighted average as a general rule. Neuroimage 70:37–47.

Baldassano C, Beck DM, Fei-Fei L (2016) Human–object interactions are more than the sum of their parts. Cereb Cortex 27:2276–2288.

Bao P, Tsao DY (2018) Representation of multiple objects in macaque category-selective areas. Nat Commun 9:1774.
Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.
Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. Nat Rev Neurosci 13:51–62.
Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017) Stan: a probabilistic programming language. J Stat Soft 76:1–32.
Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. Neuroimage 9:179–194.
Desimone R, Duncan J (1995) Neural mechanism of selective visual attention. Annu Rev Neurosci 18:193–222.
Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. Science 293:2470–2473.
Fisher C, Freiwald WA (2015) Whole-agent selectivity within the macaque face-processing system. Proc Natl Acad Sci U S A 112:14717–14722.
Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. Nature Reviews. Neuroscience 15:536–548.
Kaiser D, Peelen MV (2018) Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. Neuroimage 169:334–341.
Kaiser D, Strnad L, Seidl KN, Kastner S, Peelen MV (2014) Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns. J Neurophysiol 111:82–90.
Kanwisher N, Yovel G (2006) The fusiform face area: a cortical region specialized for the perception of faces. Philos Trans R Soc Lond B Biol Sci 361:2109–2128.
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.
Kleiner M, Brainard DH, Pelli DG, Broussard C, Wolf T, Niehorster D (2007) What's new in Psychtoolbox-3? Perception 36:1–16.
MacEvoy SP, Epstein RA (2009) Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. Curr Biol 19:943–947.
MacEvoy SP, Epstein RA (2011) Constructing scenes from objects in human occipitotemporal cortex. Nat Neurosci 14:1323–1329.
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.
Peelen MV, Downing PE (2005) Selectivity for the human body in the fusiform gyrus. J Neurophysiol 93:603–608.
Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59:2142–2154.
R Development Core Team (2011) R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
Reddy L, Kanwisher NG, Vanrullen R (2009) Attention and biased competition in multi-voxel object representations. Proc Natl Acad Sci U S A 106:21447–21452.
Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron 61:168–185.
Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo X-N, Holmes AJ, Eickhoff SB, Yeo BTT (2018) Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. Cereb Cortex 28:3095–3114.
Song Y, Luo YLL, Li X, Xu M, Liu J (2013) Representation of contextually related multiple objects in the human ventral visual pathway. J Cogn Neurosci 25:1261–1269.
Weiner KS, Grill-Spector K (2010) Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. Neuroimage 52:1559–1573.
Weiner KS, Grill-Spector K (2011) Not one extrastriate body area: using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. Neuroimage 56:2183–2199.
Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011) Large-scale automated synthesis of human functional neuroimaging data. Nat Methods 8:665–670.
Yovel G, Freiwald WA (2013) Face recognition systems in monkey and human: are they the same thing? F1000Prime Rep 5:10.
Zoccolan D, Cox DD, DiCarlo JJ (2005) Multiple object response normalization in monkey inferotemporal cortex. J Neurosci 25:8150–8164.