The functional organization of high-level visual cortex determines the representation of complex visual stimuli

Libi Kliger[1,*] & Galit Yovel[1,2]

[1] The School of Psychological Sciences and [2] Sagol School of Neuroscience,

Tel Aviv University, P.O. Box 39040, Tel Aviv 6997801, Israel

* Correspondence: libikl@mail.tau.ac.il

Conflict of interest statement:

The authors declare that there is no conflict of interest.

Acknowledgments:

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.

Abstract

A hallmark of high-level visual cortex is its functional organization of neighboring areas that are selective to single categories such as faces, bodies and objects. However, visual scenes are typically composed of multiple categories. How does 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 category-selective cortex as a function of the magnitude of category selectivity to 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 degree of its category-selectivity. This was the case for both related (face+body) and unrelated (face+wardrobe) 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.

Significance Statement

It is well established that 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 to 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 of 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 primates' high-level visual cortex is its division to category-selective areas, such as face, body or object-selective regions that reside in nearby locations (Downing, Jiang, Shuman, & Kanwisher, 2001; Kanwisher, McDermott, & Chun, 1997; Kanwisher & Yovel, 2006; Malach et al., 1995; Grill-Spector & Weiner, 2014). This division to 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 contribution 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 IT in monkeys (Zoccolan, Cox, & DiCarlo, 2005) or LOC in humans (Baeck, Wagemans, & de Beeck, 2013; Macevoy & Epstein, 2009) was a mean or a weighted mean response of the isolated stimuli, the response in category-selective areas, such as face- or scene-areas (Bao & Tsao, 2018; Reddy, Kanwisher, & Vanrullen, 2009) was similar to the response to the preferred category (i.e., a max 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 & Heeger, 2012; Fig. 1a) and therefore 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 max) 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 to the same category as the recorded neuron (i.e., a face-neuron in a

face-selective area), the normalization pool is unresponsive to the non-preferred stimulus and therefore does not reduce the response of the recorded neuron to its preferred stimulus, yielding a max response.

Here we provide a general framework for the relationship between category-selectivity and the representation of multi-category stimuli, as detailed below (see Figure 1), by showing this correspondence with functional MRI across a large continuous area of cortex. Category-selectivity, as measured with fMRI, can provide an estimate of the proportion of neurons that are selective to each of the measured categories and therefore with a measure of the homogeneity of the normalization pool. A voxel that shows high selectivity to a given category has a larger proportion of neurons selective to 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 max response in category-selective areas to a mean response in areas that show similar response to multiple categories, such as in the borders between two category-selective areas (Fig. 1b). 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 to each of the stimuli determines its weights (Fig. 1c-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.

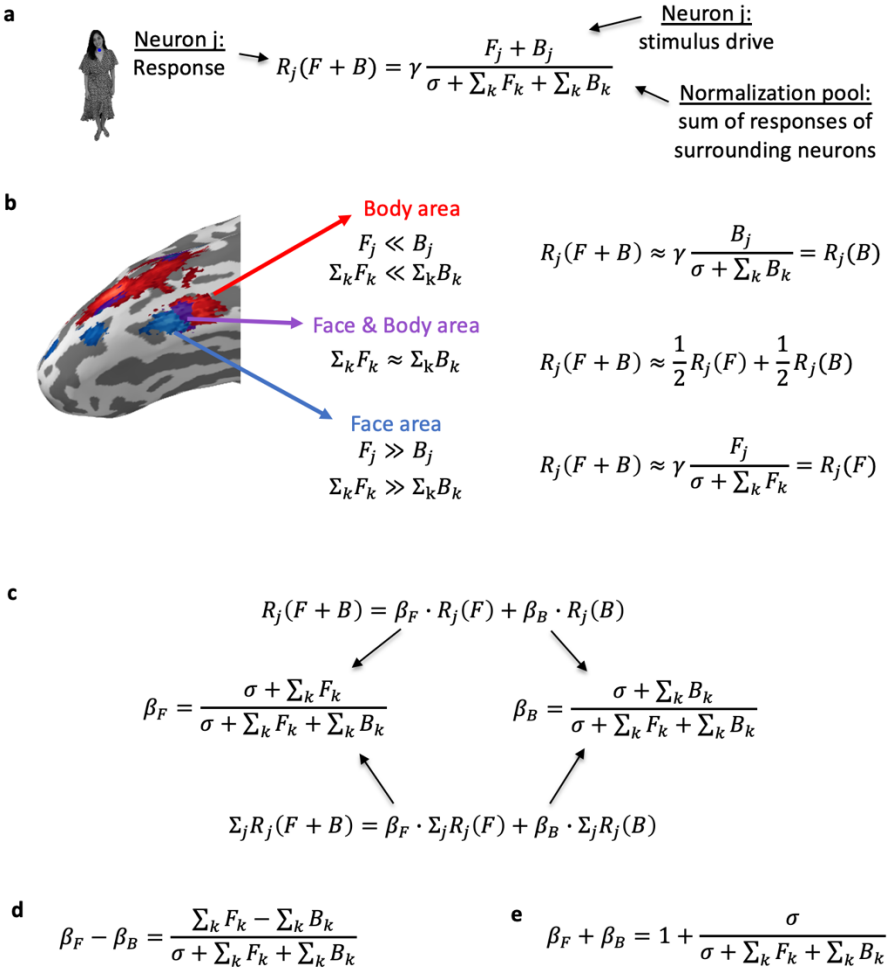

Figure 1: (a) The normalization equation (Reynolds & Heeger, 2009). The response of a neuron is divided (normalized) by the sum of the responses of the surrounding neurons. Here we show the response to a face (F) and a body (B) presented together. (b) A surface map of face- and body-selective areas with 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 to the same category, and therefore resulting in a max-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 contribution 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) The normalization equation further predicts that the difference between the weights corresponds to the difference in the proportions of face and body selective neurons, (e) and that the sum of weights is slightly higher than 1 (i.e., 1 plus a small positive term). Formal derivations can be found at https://github.com/LibiKl/multiple_objects_fMRI_analysis.

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. 2a) 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. 2b).

Participants

Thirty-Two healthy volunteers with normal or corrected-to-normal vision participated in both experiments. Fifteen volunteers (6 women, ages 19-37, 13 right-handed) participated in Experiment 1 and seventeen healthy volunteers (11 women, ages 20-30, 14 right-handed) that did not participate in Experiment 1 participated in Experiment 2. Two participants were excluded form analysis of Experiment 2 due to technical difficulties. Participants were paid $15/hr. 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 performed in accordance with relevant guidelines and regulations. The sample size for each experiment (N=15) chosen for this study was similar to sample size of other fMRI studies that examined the representation of multiple objects in high-level visual cortex (10-15 subjects per experiment) (see for example: Baeck et al., 2013; Baldassano et al., 2016; Kaiser & Peelen, 2017; Kaiser et al., 2014; Macevoy & Epstein, 2009; MacEvoy & Epstein, 2011; Reddy et al., 2009; Song et al., 2013)

Stimuli

Face+body stimuli. The face+body stimuli set was used in both Experiment 1 and Experiment 2. Stimuli consisted of 40 grey-scale images of a whole person standing in a straight frontal posture with their background removed 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 (Figure. 2a). 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 (at the center and upper part of the display) (Figure 2a). The size of the whole person image was approximately 3.5X12.2 degrees 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 (Figure 2b). 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 grey-scale wardrobes with their background 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) (Figure 2b). The size of the face+object pair was similar to the face+body and was approximately 3.5X12.2 degrees of visual angle.

Functional localizer stimuli. Functional localizer stimuli of Experiment 1 were grey-scale images of faces, headless-bodies, non-leaving objects (Figure 2C), and images of the whole person that were not included in analyses of this study. Functional localizer stimuli of Experiment 2 were grey-scale pictures of faces, headless-bodies, non-leaving objects, and scrambled objects (Figure 2C). The size of the stimuli was approximately 5.5X5.5 degrees of visual angle.

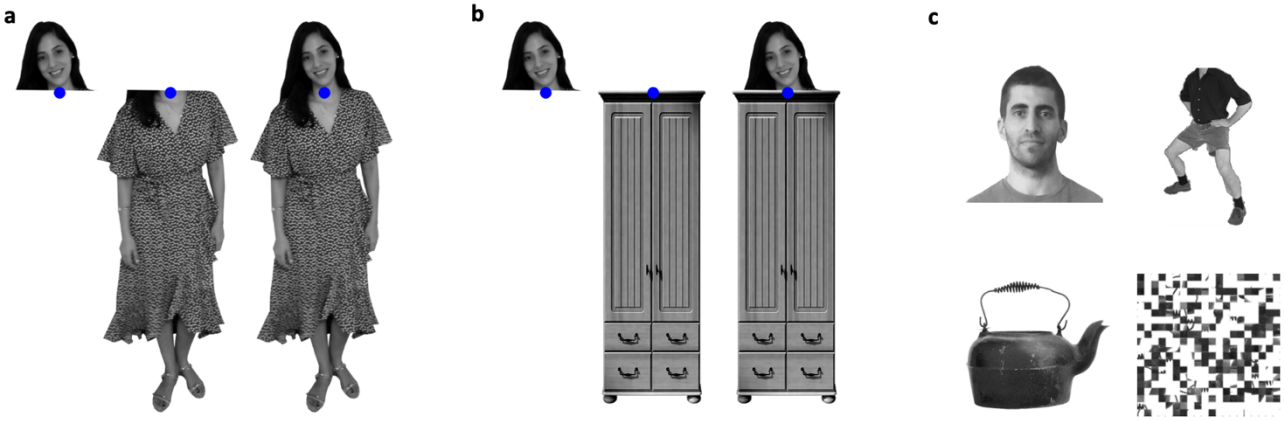

Figure 2: (a) A Face-Body stimulus set: face, body, and face+body stimuli, taken from the same images. The fMRI response to these stimuli was used to estimate the contribution of the face and the body to the face+body representation. Participants were asked to fixate on the blue-dot and perform a one-back task (see 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 one-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 voxel-wise selectivity to specific categories, independently from the data that were used to estimate the contribution of each part to the multi-category representation.

Apparatus and Procedure

fMRI acquisition parameters. fMRI data were acquired in a 3T Siemens MAGNETOM Prisma MRI scanner in Tel Aviv University, using a 64-channel head coil. Echo-planar volumes were acquired with the following parameters: repetition time (TR) = 2 s, echo time = 30 ms, flip angel = 82°, 64 slices per TR, multi-band acceleration factor = 2, acceleration factor PE = 2, slice thickness = 2 mm, field of view = 20 cm and 100 × 100 matrix, resulting in a voxel size of 2 × 2 × 2 mm. Stimuli were presented with Matlab (The MathWorks Inc.) and Psychtoolbox (Brainard, 1997; Kleiner et al., 2007) and displayed on a 32" 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, echo time = 2.88 ms, 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 functional localizer. Each of the six main-experiment runs included 5 triads

of face, body and face+body mini-blocks. Fig. 3 shows an example of two such triads. The order of face, body and face+body mini-blocks within each triad was counter-balanced 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 x 5 triads). The 35 identities were randomly chosen from the set of 40 identities. Each mini-block lasted 6 seconds and was followed by 12 seconds of fixation. A single stimulus display time was 0.325 seconds, inter-stimulus-interval was 0.425 seconds. Subjects performed a 1-back task (one repeated stimulus in each block). Each run began with six seconds (3 TRs) of fixation (dummy scan) and lasted a total of 276 seconds (138 TRs). Subjects were instructed to maintain fixation throughout the run and their eye movements were recorded with an Eye tracker (EyeLink®).

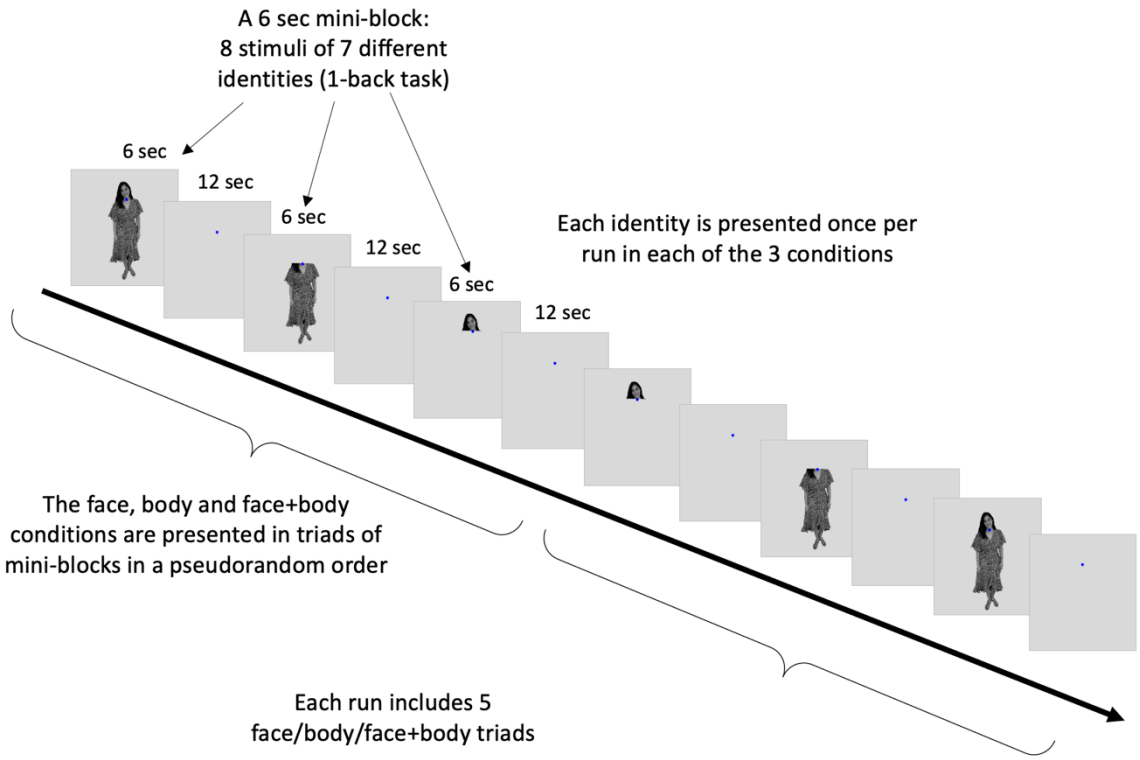

Figure 3: Experimental procedure. Each run had 15 blocks of 3 conditions (5 blocks each). See Methods for a full description of the procedure.

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 3 runs of face, body and face+body stimuli identical to Experiment 1. In addition, 3 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 person condition is not included in this paper) in Experiment 1 and faces, bodies, objects and scrambled objects in Experiment 2. Each block presented 20 stimuli of 18 different images of which two repeated twice for a 1-back task. Each stimulus was presented for 0.4 sec with 0.4 sec Inter-stimulus interval. Each block lasted 16 seconds. Each run began with a six-seconds fixation (3 TRs) and lasted a total of 342 seconds (171 TRs).

Data analyses

fMRI Data Analysis and preprocessing

fMRI analysis was performed using SPM12 software, Matlab (The MathWorks Inc.) and R (R Development Core Team, 2011) costumed scripts, STAN (Carpenter et al., 2017) for Bayesian model fitting and Freesurfer (Dale, Fischl, & Sereno, 1999), pysurfer (https://pysurfer.github.io) and Python (http://www.python.org) costumed scripts for the surface generation and presentation. The code that was used for data analyses is available at https://github.com/LibiKl/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 co-registeration to the anatomical 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, data used for all

other analyses remained in 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, Barnes, Snyder, Schlaggar, & Petersen, 2012) was applied for every volume with frame-displacement (FD) > 0.9 by adding a nuisance regressor with a value of 1 for that specific volume and zeros for all other volumes. Percent signal change (PSC) for each voxel was calculated for each experimental condition in each run by dividing the beta weight for that regressor by the beta weight of the baseline for that run.

Experiment 1:

Region of interest (ROI) 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 (ROI) were defined as clusters (>10 voxels) of voxels selective to a given category ($p<10^{-4}$) within specific anatomical 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 less than 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 Figure 5 for the stability of the results across different number of voxels even with very low number of subjects).

Linear model fitting. The mean percent signal change (PSC) 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 percent signal change, PSC) in each of these voxels:

$$(Face + Body)_{PSC} = \beta_F^{(FB)} \cdot Face_{PSC} + \beta_B^{(FB)} \cdot Body_{PSC} + \varepsilon^{(FB)} \quad (1)$$

The beta 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 beta coefficients of the multi-category response model are not the same as the betas derived from the standard fMRI GLM analysis. The betas from the standard fMRI GLM analysis are used to determine the percent signal change (PSC) to each of the single- and multi-category stimuli as a measure of the fMRI response to that stimuli). We calculated the mean of the beta 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. 1c, equation 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 voxel-wise analysis. For each voxel within each ROI we compared the PSC to the face+body to the maximum PSC to the face and the body, and calculated the proportion of voxels that showed smaller response to the face+body, i.e., $face + body < \max(face, body)$. This analysis was done to assure that 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 ventro-temporal and lateral occipital cortex. In addition, we defined 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, Poldrack, Nichols, Van Essen, & Wager, 2011, https://neurosynth.org), a meta-analysis tool for extracting cognitive maps. We used an association map with the term “Place” and thresholded with FDR criterion of 0.01. We than

masked the image to include only the right parahipocampal cortex. This image was than underwent inverse normalization from an MNI space to each subject's native space. The Neurosynth-defined PPA included less 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 beta 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 beta coefficients of all spheres within the searchlight area to present the spatial distribution of the beta coefficients. We calculated R^2 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. that this model can be worse in predicting the dependent variable compared to a model with only an intercept.

To examine the relationship between the difference between the face and body beta coefficients and the selectivity to 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, and a one-sample t-test was used against a null-hypothesis of zero. To reduce statistical dependency in our dataset 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.

Experiment 2:

ROI Analysis. Based on the functional localizer data, face- body- and object-selective voxels were defined individually for each subject. Regions of interest (ROI) were defined as clusters (>10 voxels) of category selective voxels ($p<10^{-4}$) within specific anatomical locations that show preference to a single category relative to all other categories: (1) Fusiform face area (FFA): Face > Body, Object & Scrambled-object within the Fusiform gyrus; (2) Fusiform body area (FBA):): Body > Face, Object & Scrambled-object within the Fusiform gyrus; (3) Ventral object area: Object > Face, Body & Scrambled-object within the medial part of the ventral temporal cortex. Note that we used a modified and also well-accepted (e.g., Peelen & Downing, 2005; Weiner

& Grill-Spector, 2010, 2011) version of the ROI definitions relative to Experiment 1 (for example, FFA were defined in Experiment 1 with the contrast Face>Object, in opposed to the current Face>Body, Object & 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. However, 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 less than 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 (equation 1) and to predict the response to the face+object based on the response to the face and the object using the following equation:

$$(Face + Object)_{PSC} = \beta_F^{(FO)} \cdot Face_{PSC} + \beta_O^{(FO)} \cdot Object_{PSC} + \varepsilon^{(FO)} \quad (2)$$

Similar to Experiment 1, we calculated the beta 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 measure 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 that did not had 30 voxels for all three ROIs (3 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 Ventro-temporal 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 was 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 method section. To assess the level of significance of the correlations, the correlation values were transformed to fisher's Z, and a one-sample t-test (one tailed) corrected for multiple comparisons was used to assess if the correlation value averaged across participants was significantly higher than zero for each brain parcel.

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) will be a weighted mean of the response to each of its components (a face and a body), and that the weights will be determined by the magnitude of category-selectivity to each of the stimuli and will therefore vary continuously along category-selective cortex.

Region of interest (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 to both faces and bodies) using the independent functional localizer data (see Fig. 4 for an example of these areas in a representative subject). For each subject and each area within the right ventro-temporal cortex we fitted a linear regression model (equation 1) to estimate the contribution of the isolated face and body to the response to face+body, indicated by the beta 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 (see Figure 5 for 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 non-zero face and body coefficients ($\beta's$ = [0.39-0.74], all p values < .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 to 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. 1d). 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 (C.I.): (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% C.I. (-0.451, -0.146), Cohen's d=1.314]. In the area of overlap between the FFA and the FBA, which is selective to 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% C.I. (-0.177, 0.316), Cohen's d=0.189].

Consistent with our predictions (Fig. 1e), we found that the sum of the beta coefficients was slightly larger than 1 [mean sum (s.e.m.): 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 smaller than 1 [all p-values <0.01, all Cohen's d values > 1.144], and the sum of these coefficients is lower than 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% of the voxels in our data [FFA: 53.33%, FBA: 58.48%, Overlap: 47.88%] showed higher

response to one of the single stimuli (a face or a body) relative to the response to the combined stimulus (face+body).

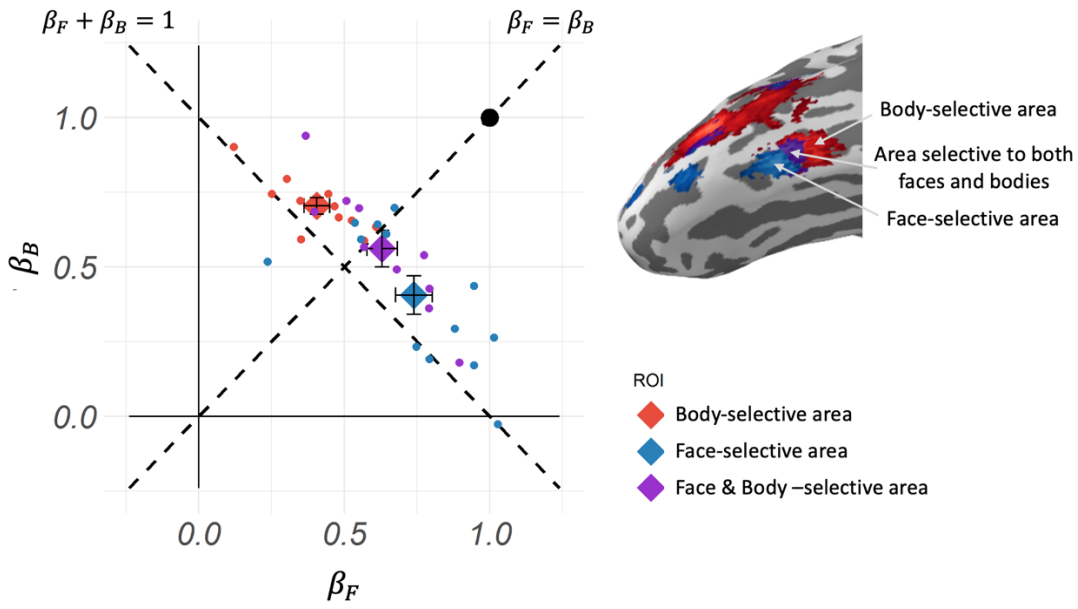

Figure 4: Experiment 1: Left: A scatterplot of the beta coefficients for the face and the body that best fit the response of the 30 most selective voxels within each subject's ROI to the face+body stimulus. Each dot indicates the results of a single subject within an ROI (in the right hemisphere). β_F 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 s.e.m.). Right: a brain surface of one representative subject showing the location of the face-selective, body-selective and the overlap areas in ventro-temporal cortex.

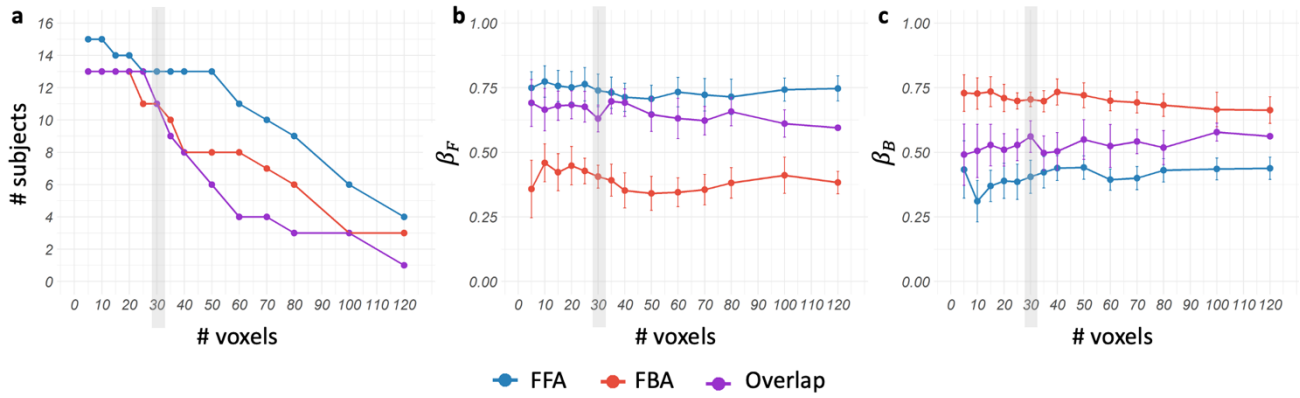

Figure 5: ROI analysis across different number of voxels. Analysis reported is based on 30 voxels for each ROI (marked in grey) (a) The number of subjects across different sizes of category-selective ROIs. As the size of the ROI increases the number of subjects decreases. (b) Mean β_F and (c) mean β_B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

To further assess if the weighted mean model (i.e., the normalization model, Fig. 1c) is the best fit to the data, we compared this model to two other models – one model with a non-zero intercept and another model with an interaction between the face and the body (i.e., a non-linear 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 (see Table 1).

	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

Table 1: Experiment 1 – Model comparison. In order to compare the proposed model predicted by the normalization equation (Fig. 1c) 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 explain the results within all ROIs.

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 occipito-temporal 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 anatomical 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. 1c).

Figure 6a-b depicts the beta 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 6c-d shows the distribution of category selectivity of 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 to faces and low selectivity to bodies show high contribution of the face to the face+body representation, while areas with low selectivity to faces and high selectivity to bodies show high contribution of the body to the face+body representation.

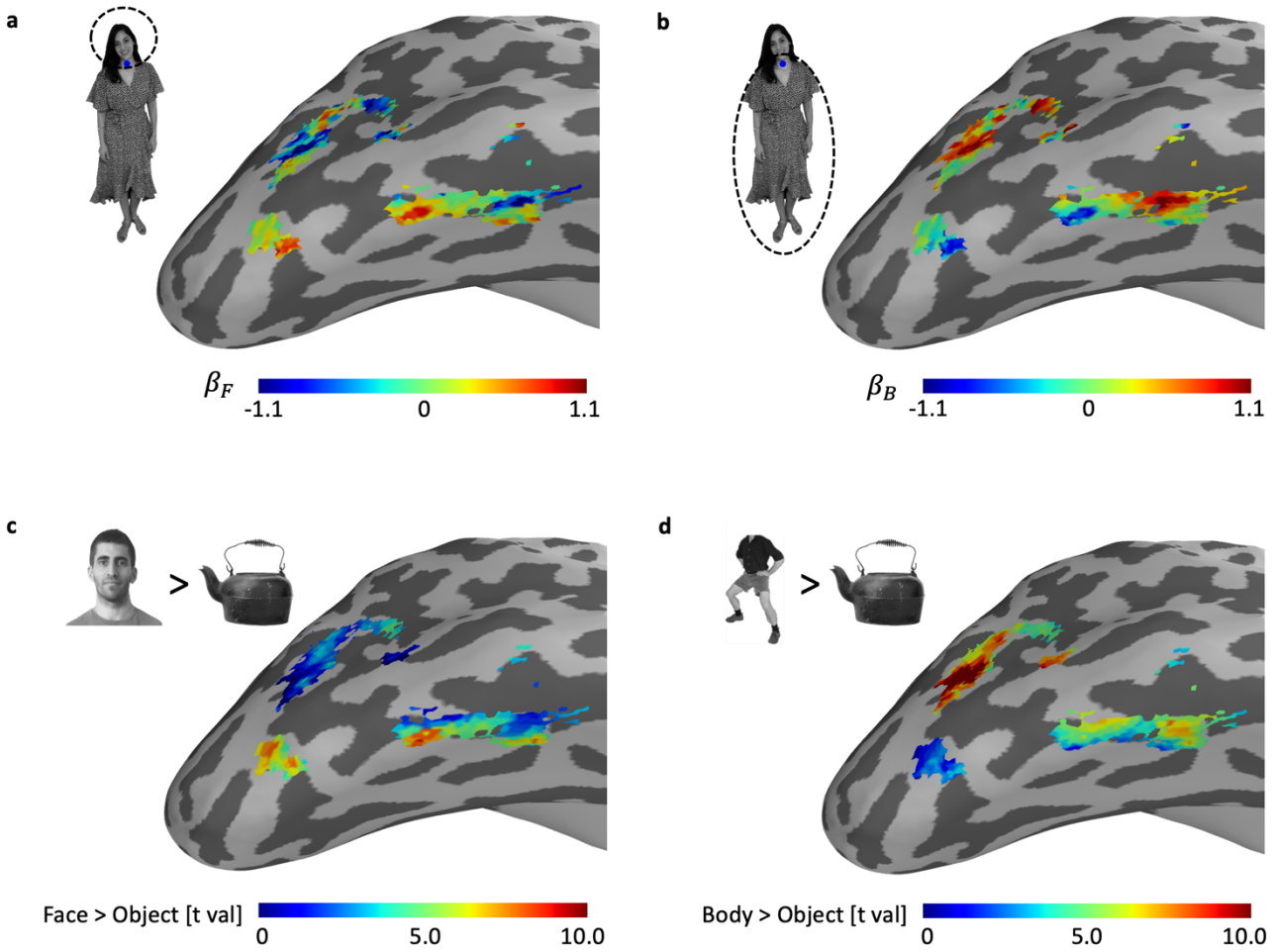

Figure 6: Experiment 1: Results of a representative subject plotted on the cortical surface for voxels that were selective to either faces or bodies: (a) The contribution of the face to the face+body representation as indicated by the face regression coefficients (β_F). (b) The contribution of the body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

Figure 7a depicts the beta 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 occipito-temporal and lateral areas of all subjects. The coefficients are scattered along the weighted mean line,

indicating a sum of coefficients that is slightly higher than 1 [mean sum=1.071, 95% C.I. (1.036, 1.106)], consistent with the derivations based on the normalization model (Fig. 1e). Figure 7d 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 to the face relative to the body, as measured by the independent functional localizer. Furthermore, consistent with our predictions (Fig. 1d), the difference between the contribution of the face and the body to the face+body representation, (i.e. the difference between the beta 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 to a null hypothesis of a correlation lower than zero [mean r=0.446, t(14)=9.653, p<0.0001 (one tailed), 95% C.I. (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 to the stimulus components, we performed a similar searchlight analysis over two control areas: Early visual cortex (EVC) (Figure 7b,e) and the Parahippocampal place area (PPA) (Figure 7c,f). EVC is sensitive to low-level features of the stimuli but not to high-level categories. PPA is part of high-level visual cortex, but is selective to places and not to 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 to the face- & body-selective areas, as indicated by the R^2 distributions (median R^2: EVC=0.722; PPA=0.487). Moreover, the sum of beta coefficient is slightly lower than 1 [EVC: mean sum=0.949, 95% confidence interval (C.I.): (0.897, 1.002); PPA: mean sum=0.813, 95% C.I.:(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 the beta 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% C.I.: (-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% C.I.: (0.006, 0.181), Cohen's d=0.094]. In order to directly compare between the ROIS, we ran a repeated measure ANOVA with ROI (face

and Body selective area, EVC, PPA) as a within-subject factor and the correlation values (after Fisher z transformation) as a dependent variable. We found a significant effect for ROI indicating a difference in the correlations between the areas [$F(2,28)=38.354$, $p<0.0001$, $\eta_G^2=0.672$]. Taken together, 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 to these categories.

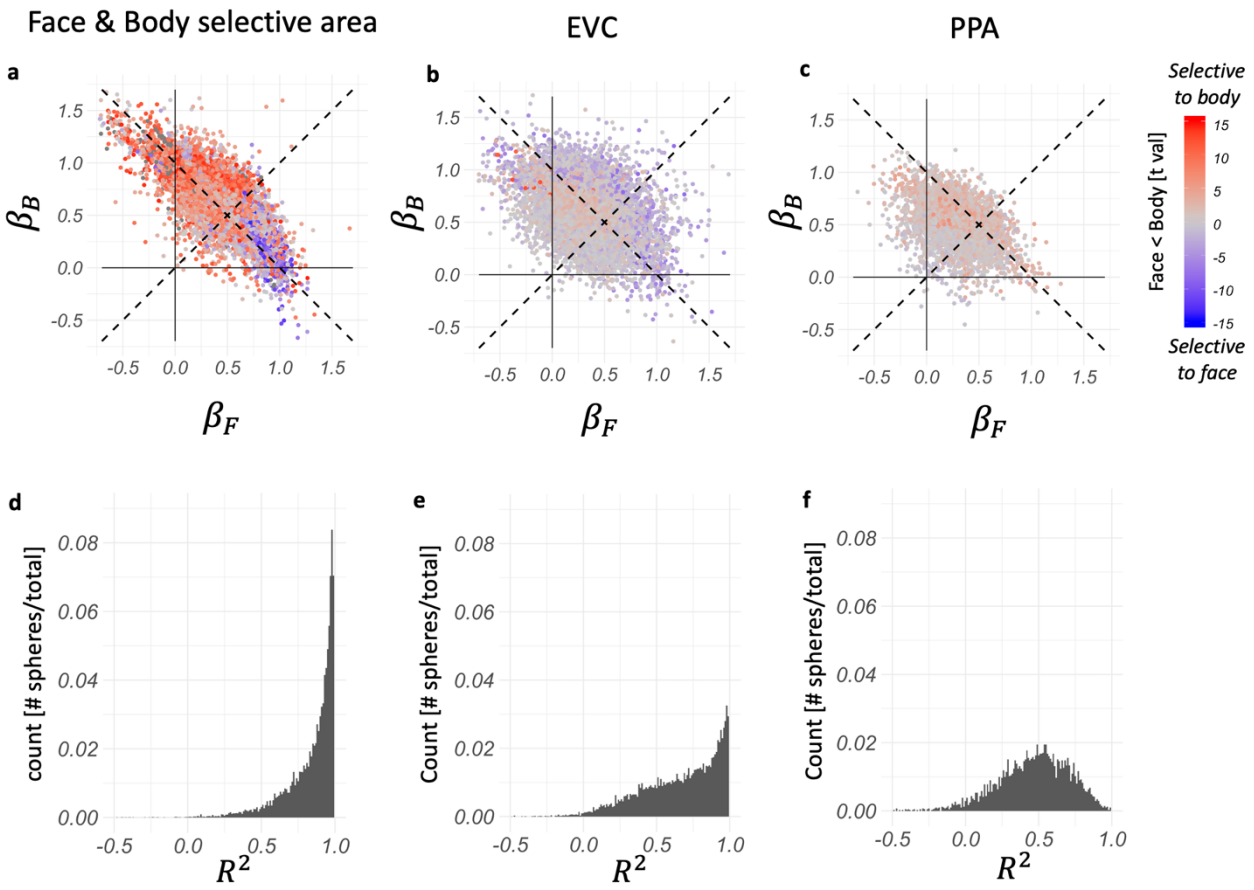

Figure 7: Experiment 1: (a-c) The beta coefficients of all spheres of all subjects in a region of interest indicating the contribution of the face (β_F) and the body (β_B) to the response to the face+body (equation (1)). The color of each dot indicates the selectivity for the face relative to the body based on independent functional localizer data [(a) Face and body selective area; (b) early visual cortex (EVC); (c) Parahippocampal place area (PPA)]. (d-f) Histograms of the R2 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 Methods data [(d) Face and body selective area; (e) EVC; (f) PPA].

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 non-related 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 face-selective area. We first defined the ROIs 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% C.I. (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% C.I. (-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% C.I. (--0.012, 0.315), Cohen's d=0.621]). However, in order to compare between the face+body and face+object findings, in Experiment 2 we used a modified definition of the ROIs than the definition used in Experiment 1, where each category was subtracted from all other categories (FFA: Face > Object, Body & Scrambled-object; FBA: Body > Object, Face & Scrambled-object) to prevent a bias toward one of the categories (see Methods). This definition excludes the face-body overlap areas, but still replicates results of Experiment 1 in face- and body-selective areas (Fig. 4a), 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, see Fig. 8a]. Furthermore, the relative contribution of the face and the body varied as a function of the face and body selectivity (Fig. 1d), 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% C.I. (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% C.I. (-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 (s.e.m.): FFA: 1.042 (0.066); FBA: 1.098 (0.054)] consistent with our model (Fig. 1e).

Next, we performed similar analyses for the face+object stimuli. For each subject we fitted a regression model for the 30 most selective voxels within the face-selective area (Face > Object, Body& Scrambled-object) and

the object-selective area (Object > Face, Body & Scrambled-object) to predict the response to the face+object based on the responses to the face and the object (equation 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, non-zero 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. see Figure 8b]. In addition, the selectivity of the area determined the relative contribution of the face and the object to the face+object representation (Fig. 1d). Specifically, we found that in the FFA, which is mainly selective to 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% C.I. (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% C.I. (-0.706, -0.318), Cohen's d=1.596]. The sum of coefficients, again, was slightly over 1 consistent with our model (Fig. 1e) [mean sum (s.e.m.): 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 non-familiar object combinations (Baldassano et al., 2016; Kaiser & Peelen, 2018; Song et al., 2013) whereas others did not find such difference (Baeck et al., 2013; Kaiser et al., 2014). To examine whether the pattern of response to face+body and face+object is different, we ran a repeated measure 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 that did not had 30 voxels in each of the three ROIs (3 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, η_G^2=0.030], as well as no interaction between the ROI and Pair Type [F(1,11)=0.024, p=0.808, η_G^2=0.0003]. Thus, the same normalization framework accounts for the two types of multi-category stimuli.

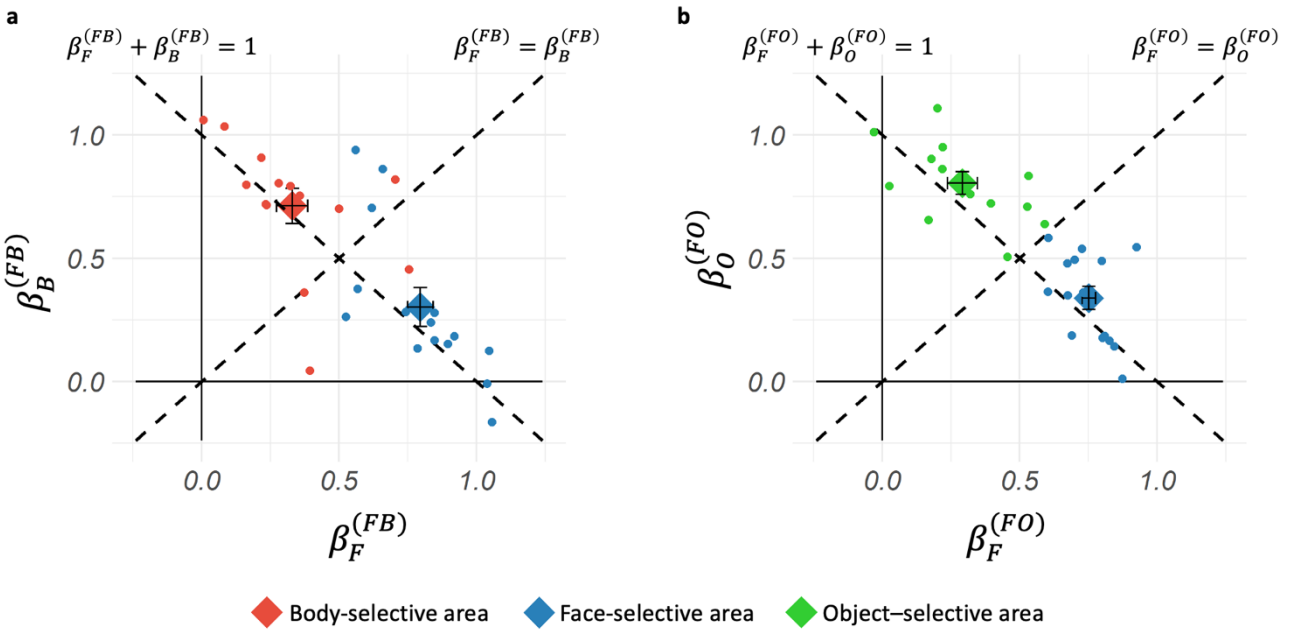

Figure 8: Experiment 2: (a) Beta coefficients for the face and the body predicting the response of the 30 most selective voxels within each subject's ROIs to the face+body stimulus. $\beta_F^{(FB)}$ is the contribution of the face to the face+body response and $\beta_B^{(FB)}$ is the contribution of 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 s.e.m.). (b) Beta 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 s.e.m.).

Searchlight analysis

A similar searchlight analysis as described in Experiment 1 was performed for the face+body (equation (1)) and the face+object stimuli (equation (2)) in ventrotemporal and lateral-occipital areas that are selective to faces, bodies or objects relative to scrambled objects (i.e., category-selective cortex). Figure 9a depicts the beta 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 contains also voxels that are selective to objects, 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 beta coefficients) is positively correlated with the selectivity to the face relative to the body as predicted [mean r=0.386, t(14)=8.444, p<0.0001 (one-tailed), 95% C.I.=(0.312, 0.456), Cohen's d=2.180], and the sum of

coefficients is slightly higher than 1 [mean sum=1.013, 95% C.I.=(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 similar results to the face+body findings (Fig. 9b): The beta-coefficients are scattered along the weighted mean line with a sum of coefficients that is slightly higher than 1 [mean sum=1.015, 95% C.I.=(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 to the face relative to the object as expected [mean r=0.395, t(14)=11.193, p<0.0001 (one tailed), 95% C.I.=(0.338, 0.449) , Cohen's d=2.890] (Fig. 1d-e).

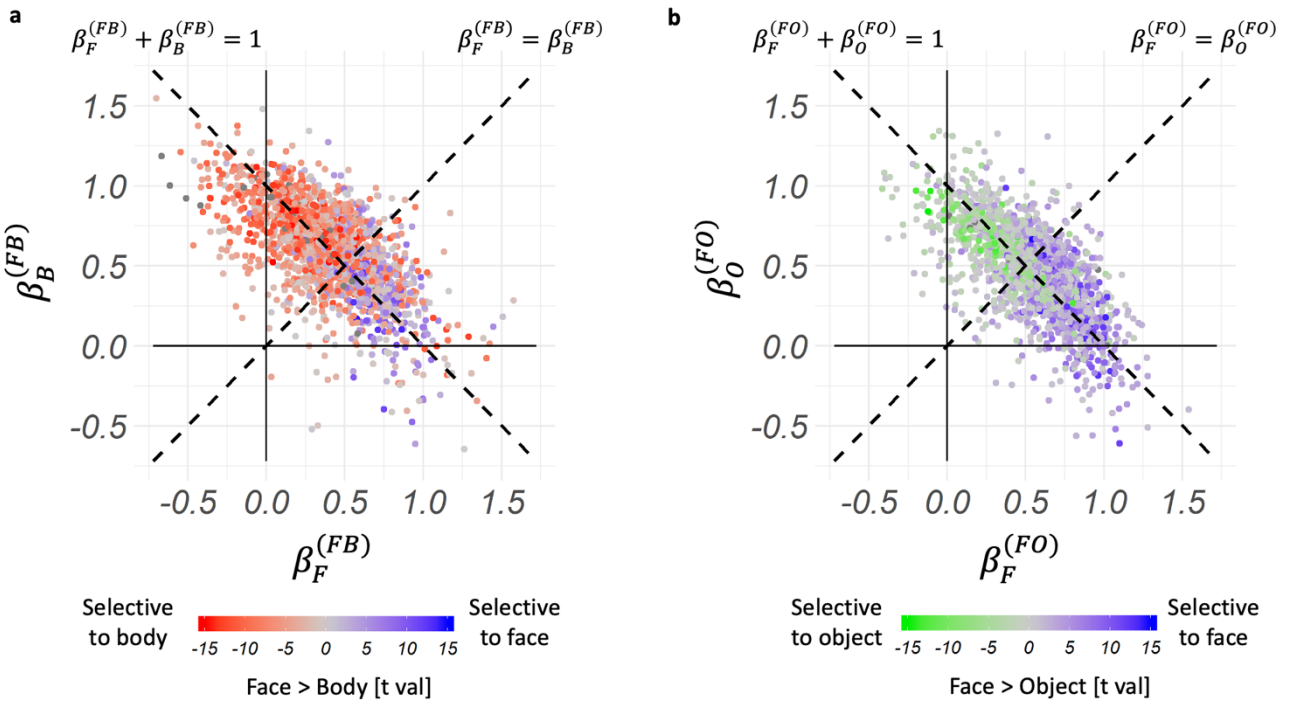

Figure 9: Results of searchlight analysis in Experiment 2. (a) The beta coefficients of all spheres in the category selective cortices of all subjects indicating the contribution of the face ($\beta_F^{(FB)}$) and the body ($\beta_B^{(FB)}$) to the response to the face+body (equation (1)). The color of each dot indicates the selectivity to 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 betas ($\beta_F^{(FB)} - \beta_B^{(FB)}$) (b) The beta 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 (equation (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 betas ($\beta_F^{(FO)} - \beta_O^{(FO)}$).

To compare the spatial distribution of the beta-coefficients and category selectivity, we plotted the difference between the coefficients and the difference between the selectivity to each pair of categories on brain surface maps of one representative subject along his category-selective cortex. (Fig. 10a-d). Figure 10a 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 10b shows the selectivity to the face relative to the selectivity to the body as measured by the independent functional localizer data. It can be seen that cortical areas that show higher contribution of the face to the face+body representation correspond to face-selective clusters (red in both figures), and that areas that show higher contribution of the body to the face+body representation correspond to body-selective clusters (blue in both figures). Figure 10c shows the difference between the contribution of the face and the object to the face+object representation for the same category-selective area. Figure 10d shows the selectivity to the face relative to the object based on the functional localizer data. Similar to the face+body results, areas that show higher contribution of the face to the face+object representation correspond to face-selective clusters (red in both figures), and areas that show higher contribution of the object to the face+object representation correspond to object-selective clusters (blue in both figures).

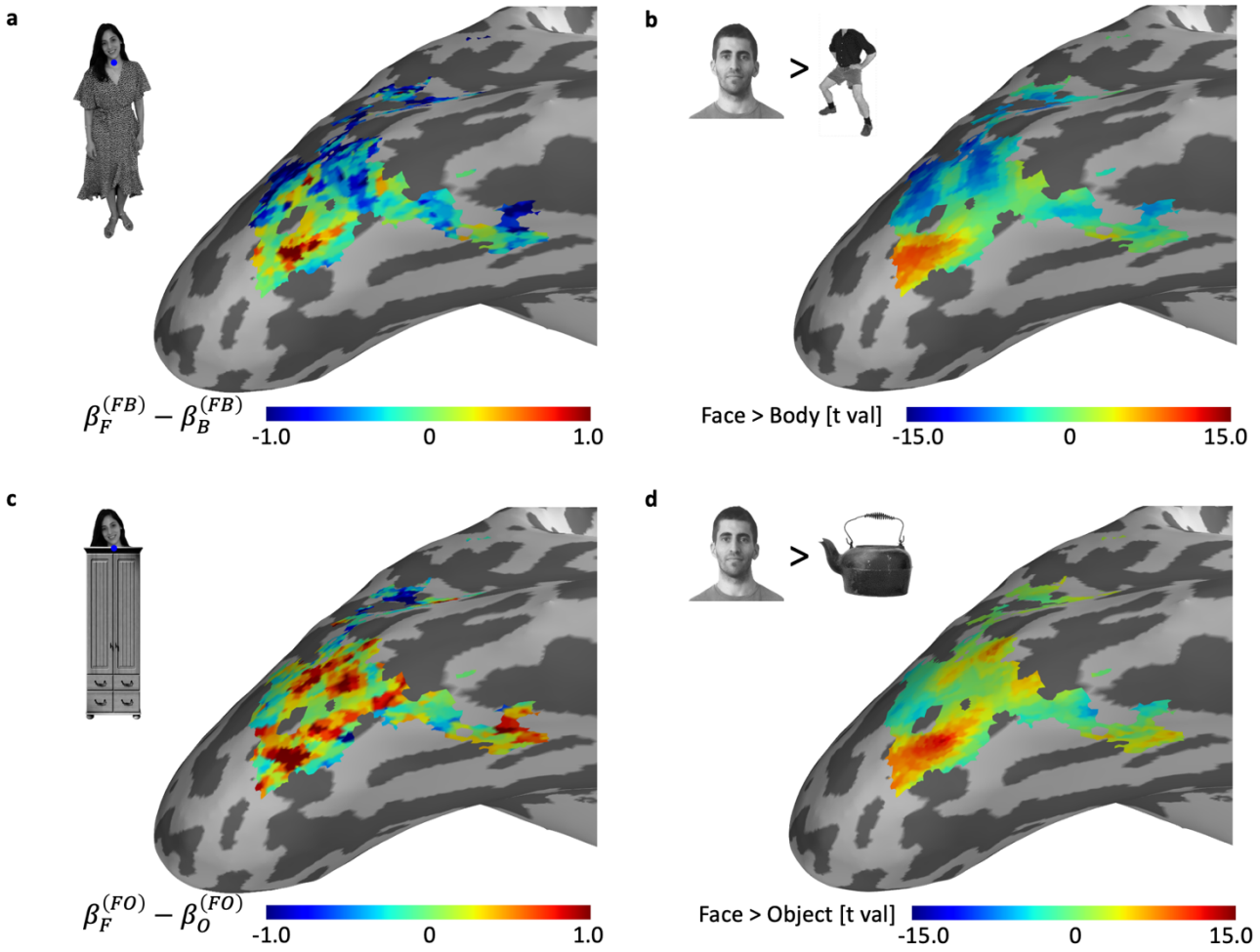

Figure 10: Experiment 2: Results of searchlight analysis of one representative subject plotted on the cortical surface show the correspondence between the difference between the coefficients of the two categories and the magnitude of their selectivity in category-selective cortex. Note that Figure 3 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 to faces relative to bodies (t map of Face>Body). Selectivity was determined based on independent functional localizer data. (c) The difference between the contribution 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 to faces relative to objects (t map of Face<Object) based on independent functional localizer data.

Whole-brain analysis

To reveal whether the correspondence between category-selectivity and multi-category representation is a unique property of the category-selective visual cortex, we performed an unconstrained whole-brain

575 searchlight analysis, similar to the searchlight analysis described in the previous section. We used a
576 parcellation of 400 parcels (Schaefer et al., 2018) and for each parcel and each subject we calculated the
577 Pearson correlation between category selectivity and the difference between the betas in our model. Figure
578 11a,c depict the correlation for each parcel of the right hemisphere for the face+body and the face+object
579 model, respectively, averaged across subjects (after Fisher's z transformation). Figure 11b,d depict parcels
580 that show significant correlation across subjects for the two models (one-tailed t-test with N=15, $p<0.05$
581 corrected for multiple comparisons). Only parcels within high-level visual cortex (ventro-temporal and lateral-
582 occipital areas) showed significant correlations. Moreover, the pattern of correlations is different for the
583 face+body model and the face+object model. The ventro-medial areas, which are typically selective to
584 inanimate stimuli show a positive correlation for the face+object model but not for the face+body model,
585 further indicating the correspondence between components of the multi-category stimuli and the selectivity
586 to its components.

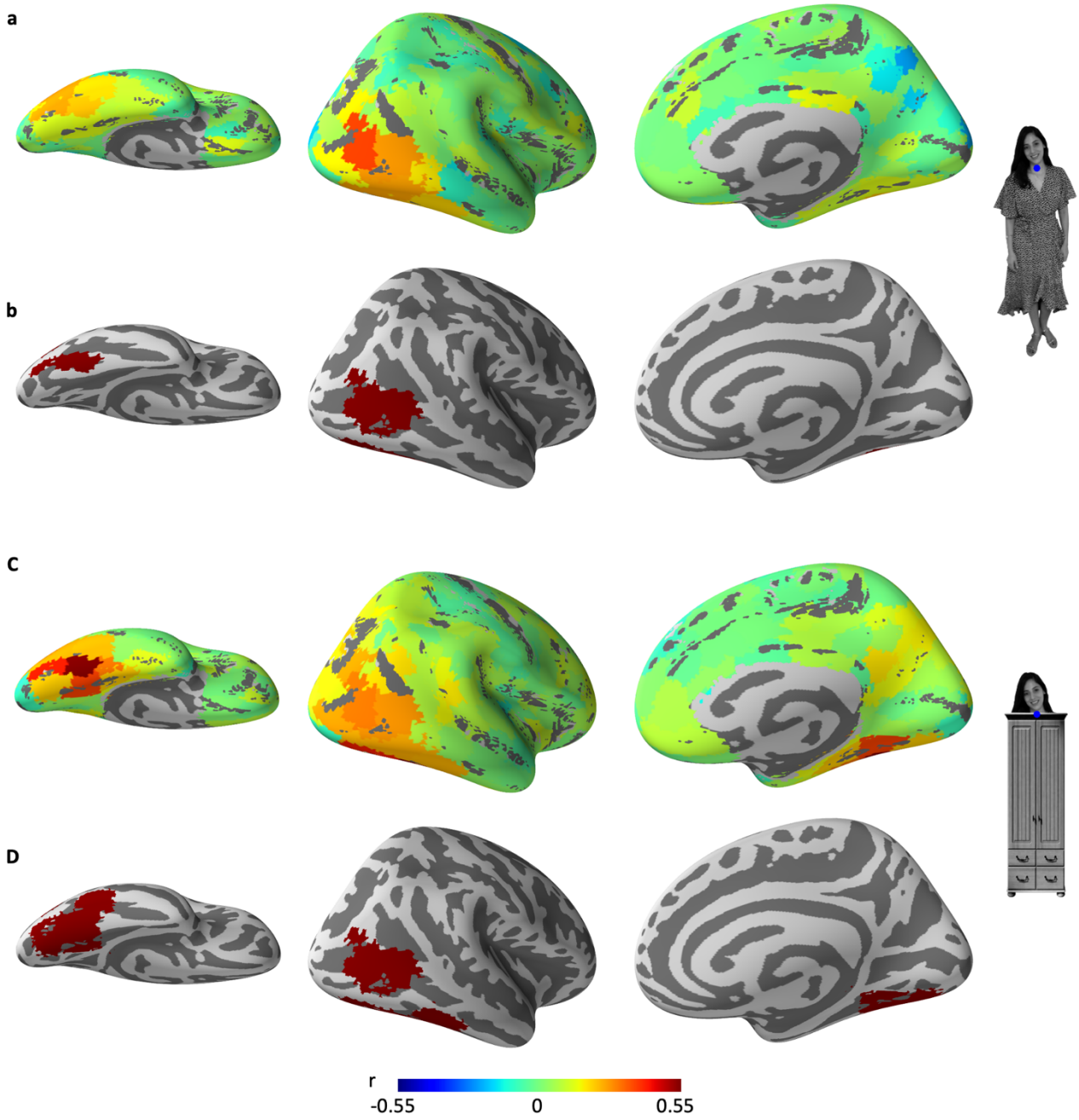

Figure 11: Experiment 2: Results of whole-brain searchlight analysis. (a) Correlation between the difference between the contribution of the face and the body to the face+body representation and difference between category selectivity (Face>Body of localizer data) in each parcel, averaged across subjects. (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). Selectivity to faces relative to bodies (t map of Face>Body). Selectivity was determined based on independent functional localizer data. (c) Correlation between the difference between the contribution of the face and the object to the face+object representation and 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).

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 (Figures 5-11). We further showed that this correspondence is restricted to category-selective visual cortex (Figure 7,11). Consistent with our predictions (Figure 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-selective cortex. These findings are consistent with a normalization mechanism (Bao & Tsao, 2018; Macevoy & Epstein, 2009; Reddy et al., 2009) but go beyond previous reports in several ways: (1) By showing that the representations of multi-category stimuli is determined by the category selectivity to their component stimuli, we provide a general framework to the various findings reported in previous studies that showed either a mean or a max 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 (3) We found that this weighted linear model accounts for the representations of both related (face+body) and non-related (face+wardrobe) multi-category stimuli.

Our findings are consistent with a recent single unit recording study (Bao & 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 to 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 non-preferred stimulus and therefore does not reduce the response of the recorded neuron to its preferred stimulus, yielding a max response. Thus, areas with high concentration of neurons selective to a single category give priority to the preferred stimulus, filtering out the non-preferred stimuli, resulting in a max response (Bao & Tsao, 2018; Reddy et al., 2009) (See Figure 1b). This operation enables hard-wired de-cluttering at early stages of visual processing (Bao & Tsao, 2018) in category-selective areas. In contrast, in areas with a mixed population of category-selective neurons, the

surrounding neurons respond to the non-preferred stimuli, yielding similar, possibly competitive, representations to different categories, resulting in a mean response. By generating a response to multiple stimuli that ranges from a mean to a max response, the normalization mechanism keeps the neuronal response within the dynamic range preventing saturation of the neural response (Carandini & Heeger, 2012). 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 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). 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 (Fig. 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, functional MRI cannot determine whether the response of neurons to a face and a body in the overlap area that is selective to both faces and bodies reflects neuronal saturation of neurons that are selective to 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 & 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 to either faces or bodies exist, they are more likely to show a mean response to a face and a body rather than neuronal saturation. Secondly, the normalization mechanism function as a “gain control” mechanism, preventing neurons from reaching saturation even when presented with more than one preferred stimulus (Carandini & Heeger 2012).

Previous neuroimaging and single unit recording studies reported mixed findings of either a mean response (Macevoy & Epstein, 2009; Zoccolan et al., 2005), a weighted mean response (Baeck et al., 2013) or a max response (Bao & Tsao, 2018; Reddy et al., 2009) 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 (Baldassano et al., 2016; Fisher & Freiwald, 2015; Kaiser & Peelen, 2018; Kaiser et al., 2014; MacEvoy & Epstein, 2011; Song et al., 2013). 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 non-additive response. Our findings show that the deviation from the mean reflects a weighted mean response. We found no evidence for a non-additive 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 non-meaningful face+wardrobe pair that generated similar representations. Similar results were reported by Baeck et al. (2013) that 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 higher than 1. Indeed, our results reveal that the sum of the coefficients is slightly higher than 1, 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 to both faces and bodies, and therefore 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 non-additive response. Finally, Fisher & Freiwald (2015) examined the contribution of the face and body to the whole person in a monkey fMRI study and found a super-additive (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 & 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 to a single or more categories, enables a flexible representation of complex visual scenes, where both de-cluttering 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 & Duncan, 1995; Reynolds & Heeger, 2009) making the taxing operation of understanding complex visual scenes dynamic and flexible.

Data Availability

The code that was used for data analysis is available at https://github.com/LibiKl/multiple_objects_fMRI_analysis. Data that was collected in this study will be available at https://openneuro.org after publication.

Author Contributions

L.K. and G.Y. designed the experiments, interpreted the data, and wrote the paper. L.K. conducted the experiments and analyzed the data.

References

Baeck, A., Wagemans, J., & de Beeck, H. P. (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. https://doi.org/10.1016/j.neuroimage.2012.12.023

Baldassano, C., Beck, D. M., & Fei-Fei, L. (2016). Human-object interactions are more than the sum of their parts. *Cerebral Cortex*, 1–13. https://doi.org/10.1093/cercor/bhw077

Bao, P., & Tsao, D. Y. (2018). Representation of multiple objects in macaque category-selective areas. *Nature Communications*, *9*(1), 1–16. https://doi.org/10.1038/s41467-018-04126-7

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*. https://doi.org/10.1163/156856897X00357

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*(1), 51–62. https://doi.org/10.1038/nrn3136

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v076.i01

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Desimone, R., & Duncan, J. (1995). Neural mechanism of selective visual attention. *Annu Rev Neurosci*, *18*,

193–222.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A Cortical Area Selective for Visual Processing of the Human Body. *Science*, *293*(5539), 2470–2473. https://doi.org/10.1126/science.1063414

Fisher, C., & Freiwald, W. A. (2015). Whole-agent selectivity within the macaque face-processing system. *Proceedings of the National Academy of Sciences*, *112*(47), 201512378. https://doi.org/10.1073/pnas.1512378112

Kaiser, D., & Peelen, M. V. (2018). Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2017.12.065

Kaiser, D., Strnad, L., Seidl, K. N., Kastner, S., & Peelen, M. V. (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. *Journal of Neurophysiology*, *111*(1), 82–90. https://doi.org/10.1152/jn.00371.2013

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*(11), 4302–4311. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9151747

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1476), 2109–2128. https://doi.org/10.1098/rstb.2006.1934

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in Psychtoolbox-3? *Perception*. https://doi.org/10.1068/v070821

MacEvoy, S. P., & Epstein, R. A. (2011). Constructing scenes from objects in human occipitotemporal cortex. *Nature Neuroscience*, *14*(10), 1323–1329. https://doi.org/10.1038/nn.2903

Macevoy, S. P., & Epstein, R. A. (2009). Decoding the representation of multiple simultaneous objects in

741 human occipitotemporal cortex. *Current Biology : CB*, *19*(11), 943–947.
742 https://doi.org/10.1016/j.cub.2009.04.020

743 Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., … Tootell, R. B. (1995).
744 Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex.
745 *Proceedings of the National Academy of Sciences*, *92*(18), 8135 LP – 8139.
746 https://doi.org/10.1073/pnas.92.18.8135

747 Peelen, M. V, & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of*
748 *Neurophysiology*, *93*(1), 603–608. https://doi.org/10.1152/jn.00513.2004

749 Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., & Kastner, S. (2009). Neural
750 Representations of Faces and Body Parts in Macaque and Human Cortex: A Comparative fMRI Study.
751 *Journal of Neurophysiology*. https://doi.org/10.1152/jn.91198.2008

752 Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic
753 correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*(3),
754 2142–2154. https://doi.org/10.1016/J.NEUROIMAGE.2011.10.018

755 R Development Core Team, R. (2011). *R: A Language and Environment for Statistical Computing*. *R*
756 *Foundation for Statistical Computing*. https://doi.org/10.1007/978-3-540-74686-7

757 Reddy, L., Kanwisher, N. G., & Vanrullen, R. (2009). Attention and biased competition in multi-voxel object
758 representations. *Proceedings of the National Academy of Sciences of the United States of America*,
759 *106*(50), 21447–21452.

760 Reynolds, J. H., & Heeger, D. J. (2009). The Normalization Model of Attention. *Neuron*, *61*(2), 168–185.
761 https://doi.org/10.1016/j.neuron.2009.01.002

762 Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., … Yeo, B. T. T. (2018). Local-
763 Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral*

764 *Cortex*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

765 Song, Y., Luo, Y. L. L., Li, X., Xu, M., & Liu, J. (2013). Representation of Contextually Related Multiple Objects
766 in the Human Ventral Visual Pathway. *Journal of Cognitive Neuroscience*, *25*(8), 1261–1269.
767 https://doi.org/10.1162/jocn_a_00406

768 Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans.
769 *Proceedings of the National Academy of Sciences*, *105*(49), 19514–19519.
770 https://doi.org/10.1073/PNAS.0809662105

771 Weiner, K. S., & Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in
772 human ventral temporal cortex. *NeuroImage*, *52*(4).
773 https://doi.org/10.1016/j.neuroimage.2010.04.262

774 Weiner, K. S., & Grill-Spector, K. (2011). Not one extrastriate body area: Using anatomical landmarks, hMT+,
775 and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex.
776 *NeuroImage*, *56*(4). https://doi.org/10.1016/j.neuroimage.2011.03.041

777 Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated
778 synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670.
779 https://doi.org/10.1038/nmeth.1635

780 Yovel, G., & Freiwald, W. A. (2013). Face recognition systems in monkey and human: Are they the same
781 thing? *F1000Prime Reports*. https://doi.org/10.12703/P5-10

782 Zoccolan, D., Cox, D. D., & DiCarlo, J. J. (2005). Multiple object response normalization in monkey
783 inferotemporal cortex. *The Journal of Neuroscience : The Official Journal of the Society for*
784 *Neuroscience*, *25*(36), 8150–8164. https://doi.org/10.1523/JNEUROSCI.2058-05.2005

785