

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, Tel Aviv, Israel

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

Keywords: Neuroimaging; fMRI; Normalization model; Category-selective visual cortex; High level vision; Face; Body; Object.

1 **Summary**

2 A hallmark of high-level visual cortex is its functional organization of neighboring clusters of neurons that are
3 selective to single categories such as faces, bodies and objects. However, visual scenes are typically composed
4 of multiple categories. How does category-selective cortex represent such complex stimuli? According to a
5 normalization mechanism, the response of a single neuron to multiple stimuli is normalized by the response
6 of its neighboring neurons (normalization pool). Here we show that category-selectivity, measured with fMRI,
7 can provide an estimate for the heterogeneity of the normalization pool, which determines the response to
8 multiple stimuli. These results provide a general framework for the varying representations of multiple stimuli
9 that were reported in different regions of category-selective cortex in neuroimaging and single-unit recording
10 studies. This type of organization may enable a dynamic and flexible representation of complex visual scenes
11 that can be modulated by higher-level cognitive systems according to task demands.

12 Introduction

13 A fundamental feature of primates' high-level visual cortex is its division to category-selective areas, such as
14 face, body or object-selective regions (Downing, Jiang, Shuman, & Kanwisher, 2001; Kanwisher, McDermott,
15 & Chun, 1997; Kanwisher & Yovel, 2006; Malach et al., 1995; Peelen & Downing, 2005; Pinsk et al., 2009; Tsao,
16 Moeller, & Freiwald, 2008). These category-selective regions reside in adjacent locations along the lateral
17 occipital and ventral temporal cortex (Fisher & Freiwald, 2015; Grill-Spector & Weiner, 2014; Premereur,
18 Taubert, Janssen, Vogels, & Vanduffel, 2016; Schwarzlose, Baker, & Kanwisher, 2005; Weiner & Grill-Spector,
19 2013). This division to areas selective to certain categories has led to numerous studies that have examined
20 the profile of response of these category-selective areas to different categories when presented in isolation
21 (e.g., Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Op de Beeck, Brants, Baeck, & Wagemans, 2010;
22 Tsao, Freiwald, Tootell, & Livingstone, 2006; Yovel & Kanwisher, 2004). Nevertheless, visual scenes are
23 typically composed of multiple objects. In the present study we demonstrate that this functional organization
24 of neighboring clusters of category-selective neurons determines the representation of multi-category visual
25 scenes.

26 The neural representation of multiple stimuli has been initially examined in single unit recording studies in
27 low-level visual cortex. These studies have shown that the response to a preferred stimulus is reduced when
28 presented with a non-preferred stimulus (e.g., Reynolds, Chelazzi, & Desimone, 1999, for review, Reynolds &
29 Heeger, 2009). A normalization mechanism was proposed to account for these results. According to the
30 normalization model, the response of a neuron to a stimulus is normalized by the response of its surrounding
31 neurons to this stimulus (normalization pool) (Carandini & Heeger, 2012). When a preferred stimulus is
32 presented together with a non-preferred stimulus, neighboring neurons that are selective to the non-
33 preferred stimulus normalize the response of the neuron, resulting in a lower response to the pair of stimuli
34 relative to the response to the preferred stimulus when presented alone.

35 Whereas the normalization model was initially developed based on the response of neurons in early visual
36 cortex, findings supporting the normalization mechanism were also found in high-level visual cortex in both

37 single-unit recording and fMRI studies. These studies examined the relative contribution of the isolated stimuli
38 to the response of multiple stimuli and found different patterns of response in different areas of high-level
39 visual cortex. The response of single neurons in inferotemporal cortex (IT) of the monkey (Zoccolan, Cox, &
40 DiCarlo, 2005), as well as the fMRI response in object-selective cortex in humans (Baeck, Wagemans, & de
41 Beeck, 2013; Macevoy & Epstein, 2009) to multiple stimuli, was the mean or a weighted mean response of the
42 component stimuli. Unlike the mean response of general object areas, category-selective areas such as the
43 face-selective or scene-selective areas in humans fMRI studies (Reddy, Kanwisher, & Vanrullen, 2009) as well
44 as the face and body-selective neurons in monkeys (Bao & Tsao, 2018) showed a max response. In particular,
45 the response to a multi-category stimulus composed of a preferred and non-preferred stimulus, was not
46 reduced by the non-preferred stimulus but was similar to the response to the preferred category when
47 presented alone (i.e. a max response). Bao & Tsao (2018) suggested that such pattern of response is consistent
48 with the normalization model and can be explained by the degree of homogeneity of the normalization pool.
49 If the surrounding neurons are selective to the same category as the recorded neuron (i.e., a face neuron in a
50 face-selective area), the normalization pool is unresponsive to the non-preferred stimulus and therefore does
51 not reduce the response of the recorded neuron to its preferred stimulus, yielding a max response. Taken
52 together, previous single unit and neuroimaging studies have found either a mean response, a weighted mean
53 response or a max response to multiple stimuli. These representations of multi-category stimuli may vary with
54 the degree of homogeneity of the population of category-selective neurons (i.e. the homogeneity of the
55 normalization pool) within a given cortical region and therefore reflect the operation of the same
56 normalization mechanism in different areas of category-selective cortex (Bao & Tsao, 2018).

57 In the current study, we propose that category-selectivity, as measured with fMRI, can provide an estimate of
58 the proportion of neurons that are selective to each of the measured categories and therefore with a measure
59 of the homogeneity of the normalization pool. For example, a voxel that shows higher response to faces than
60 bodies or objects has a larger proportion of face-selective than body or object-selective neurons (i.e.
61 homogeneous normalization pool) (Tsao et al., 2006). A voxel that shows similar response to faces and bodies
62 has roughly similar proportion of neurons that are selective to faces and bodies (i.e., heterogeneous

63 normalization pool). Figure 1 shows the predictions of the normalization model for the response to a face and
 64 a body presented together in different cortical areas that are composed of face-selective neurons, body-
 65 selective neurons or with two populations of face-selective and body-selective neurons, as typically found in
 66 the borders between face- and body-selective areas. The response to multiple stimuli is expected to vary from
 67 a max response in areas with a homogeneous population of category-selective neurons to a mean response
 68 in an area with a mixed population of category-selective neurons (Fig. 1b). More generally, the normalization
 69 model predicts that the response to multiple stimuli is a weighted mean of the response to each of the stimuli,
 70 and that the weights are determined by the magnitude of category-selectivity to each of the stimuli (Fig. 1c).
 71 Thus, by using fMRI we can examine the variations in the representation of multi-category stimuli and their
 72 correspondence with the magnitude of category-selectivity across a large, continuous area of cortex.

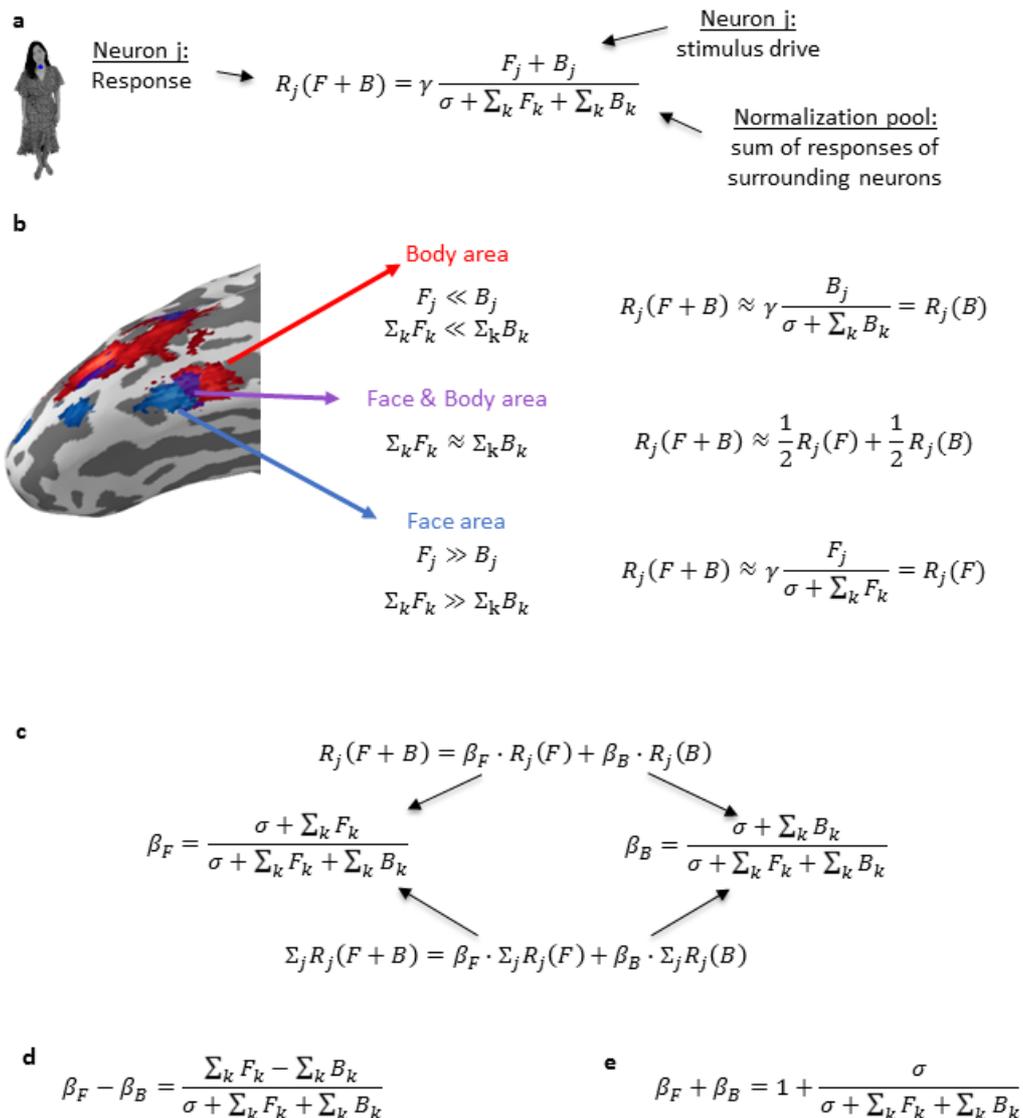
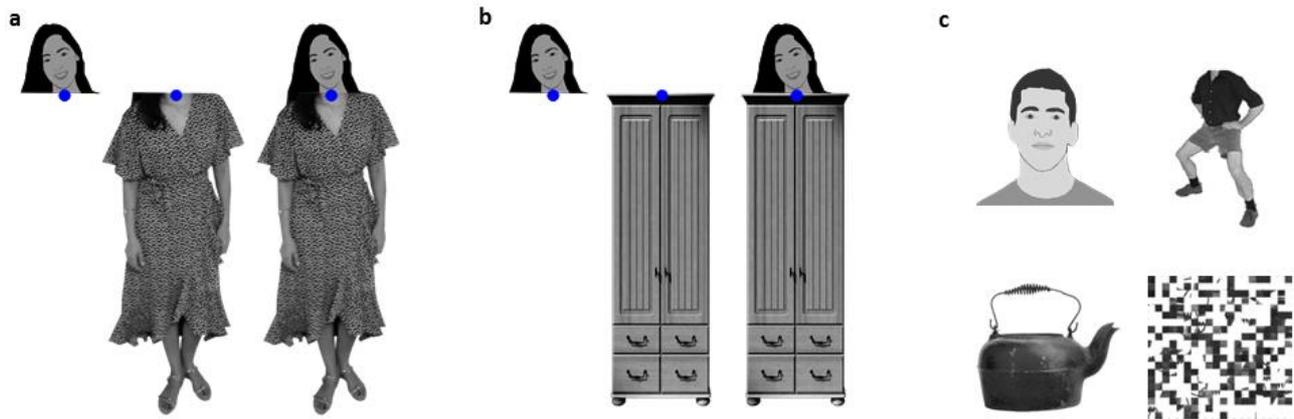


Figure 1: (a) The normalization equation (Reynolds & Heeger, 2009). The response of a neuron is divided (normalized) by the sum of the responses of the surrounding neurons. Here we show the response to a face (F) and a body (B) presented together. (b) A surface map of face- and body-selective areas with 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) (see Figure 1–figure supplement 1a for detailed derivation), 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 (see Figure 1–figure supplement 1b). (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 approximately 1 (see Figure 1–figure supplement 1c,d). **Face image was replaced by illustrations due to bioRxiv's policy on not including human faces within posted manuscripts.**

74

75 To test the correspondence between category-selectivity and the representation of multi-category stimuli in
76 category-selective cortex, we ran two fMRI studies. In the first study we presented a face, a body and a
77 face+body stimuli (Fig. 2a) and estimated the response to the combined stimulus based on the response to
78 the isolated components by fitting a linear model to the data (Reddy et al., 2009). We found that category-
79 selectivity to the isolated stimuli determines their contribution to the multi-category stimulus consistent with
80 the predications stated above of the normalization model. In a second experiment, we replicated these
81 findings and generalized them to a face+object stimulus (Fig. 2b).



82

Figure 2: (a) A Face-Body stimulus set: face, body, and face+body stimuli, taken from the same images. **Face images were replaced by illustrations in this manuscript due to bioRxiv's policy on not including human faces within posted manuscripts. The experiment stimuli included real human photos.** 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. The face and person images shown in the figures were not presented during the experiments but of individuals who gave consent to publish their images in this publication.

83 Results

84 Experiment 1 – The representation of a face+body in face- and body-selective areas

85 In an fMRI study, 15 subjects were presented with face, body and face+body stimuli, all taken from the same
86 images (see Fig. 2a). They were instructed to fixate on the blue dot. In addition, these subjects were presented
87 with a functional localizer, which included faces, bodies, objects (see Fig. 2c) and images of the whole person
88 not used for the purpose of this study. The functional localizer data were used to define face and body-
89 selective regions of interest (ROIs) as well as voxel-wise selectivity maps for faces and bodies. We estimated
90 the contribution of the face and the body to the face+body response in these different category-selective
91 areas by fitting a linear model to the data. Importantly, we did not limit the sum of the coefficients to 1 so
92 they could take any value that best fits the data. Results support our predictions (Fig. 1) that category

93 selectivity to the face and body in a given cortical area determines their contribution to the response to the
94 combined face+body stimulus (Fig. 3). Moreover, we show that the entire cortical area that is selective to
95 either faces or bodies follows the same principal normalization framework with the specific parameters
96 determined by the local profile of category selectivity (Fig. 4).

97 **Region of interest (ROI) analysis**

98 First, we examined the contribution of the face and the body to the face+body response in the face and body-
99 selective areas. For each individual subject, we extracted the face-selective area, body-selective area and the
100 overlap between these areas (i.e. areas that are selective to both faces and bodies) (see Fig. 3 for an example
101 of these areas in a representative subject). For each subject and each area within the right occipito-ventral
102 cortex, we fitted a regression model for the response of the 30 most selective voxels (see Figure 3–figure
103 supplement 1 for similar findings with different numbers of voxels) to predict the response to the face+body
104 based on the responses to the face and the body (i.e., the percent signal change, PSC) in each of these voxels:

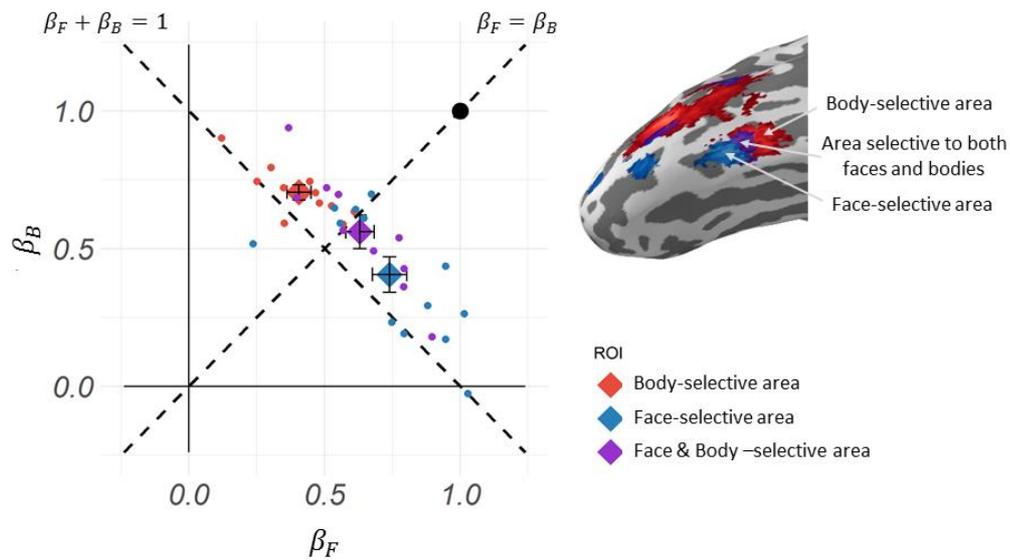
$$(Face + Body)_{PSC} = \beta_F^{(FB)} \cdot Face_{PSC} + \beta_B^{(FB)} \cdot Body_{PSC} + \varepsilon^{(FB)} \quad (1)$$

105 The beta coefficients $\beta_F^{(FB)}$ and $\beta_B^{(FB)}$ indicate the contribution of the face and the body to the face+body
106 response for each area and each subject (The beta coefficients of the multi-category response model are not
107 the same as the betas derived from the standard fMRI GLM analysis. The betas from the standard fMRI GLM
108 analysis are used to determine the percent signal change (PSC) to each of the single- and multi-category stimuli
109 as a measure of the fMRI response to that stimuli). All areas showed a significant contribution of both the face
110 and the body to the face+body representation across all subjects, indicated by positive non-zero face and body
111 coefficients ($\beta = [0.39-0.74]$, $p < .0001$, see Fig. 2 and Figure 3–table supplement 1). Figure 3–figure
112 supplement 2 shows similar results for the lateral occipital face and body areas, the Occipital Face Area (OFA)
113 and Extrastriate Body Area (EBA).

114 Based on derivations of the normalization model we can further predict that the difference between the
115 coefficients will correspond to the degree of selectivity of a cortical area to the different parts. In other words,
116 the face coefficient should be higher than the body coefficient in face-selective areas, and vice versa for body-

117 selective areas. (Fig. 1d. See Figure 1–figure supplement 1c for detailed derivation). Results were consistent
118 with this prediction. We found that in the FFA, which is composed of mainly face-selective neurons, the
119 contribution of the face was larger than the contribution of the body [$\beta_F^{(FB)} - \beta_B^{(FB)}$: mean=0.334,
120 $t(12)=2.846$, $p=0.015$]. Conversely, in the FBA, which is composed of mainly body-selective neurons, the
121 contribution of the body was larger than the contribution of the face [$\beta_F^{(FB)} - \beta_B^{(FB)}$: mean=-0.298, $t(10)=-$
122 4.358, $p=0.001$]. In the area of overlap between the FFA and the FBA, which is selective to both faces and
123 bodies, there was no significant difference between the contribution of the face and the body [$\beta_F^{(FB)} - \beta_B^{(FB)}$:
124 mean=0.070, $t(10)=-0.628$, $p=0.544$].

125 Additionally, we can further predict that the sum of the beta coefficients will be approximately 1 (Fig. 1e. See
126 Figure 1–figure supplement 1d for detailed derivation). Indeed, the sum of weights was slightly over 1 [mean
127 sum (s.e.m.): FFA: 1.145 (0.049); FBA: 1.110 (0.028); Overlap: 1.191 (0.024)] consistent with the normalization
128 model predictions (Fig. 1e). In addition, the response to the face+body is consistent with a weighted mean
129 response rather than an additive response, as indicated by the coefficients being smaller than 1 [all p-values
130 <0.01], and the sum of these coefficients is lower than 2 [all p values <0.001]. Finally, we rule out an alternative
131 explanation that the weighted mean response is due to saturation of the BOLD response to multiple stimuli.
132 We found that 53.24% of the voxels in our data [FFA: 53.33%, FBA: 58.48%, Overlap: 47.88%.] showed higher
133 response to one of the single stimuli (a face or a body) relative to the response to the combined stimulus
134 (face+body).



135

Figure 3: 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 lateral-ventral occipito-temporal cortex.

136

To further assess if the weighted mean model (i.e., the normalization model, Fig. 1c) is the best fit to the data,

137

we compared this model to two other models – one model containing a non-zero intercept and another model

138

containing an interaction between the face and the body. We found that the model that best explains our

139

results is a 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	$2.14 \cdot 10^5$	$1.94 \cdot 10^5$
FBA	$3.45 \cdot 10^7$	$5.36 \cdot 10^4$
Overlap	6.75	$1.15 \cdot 10^4$

140

Searchlight analysis

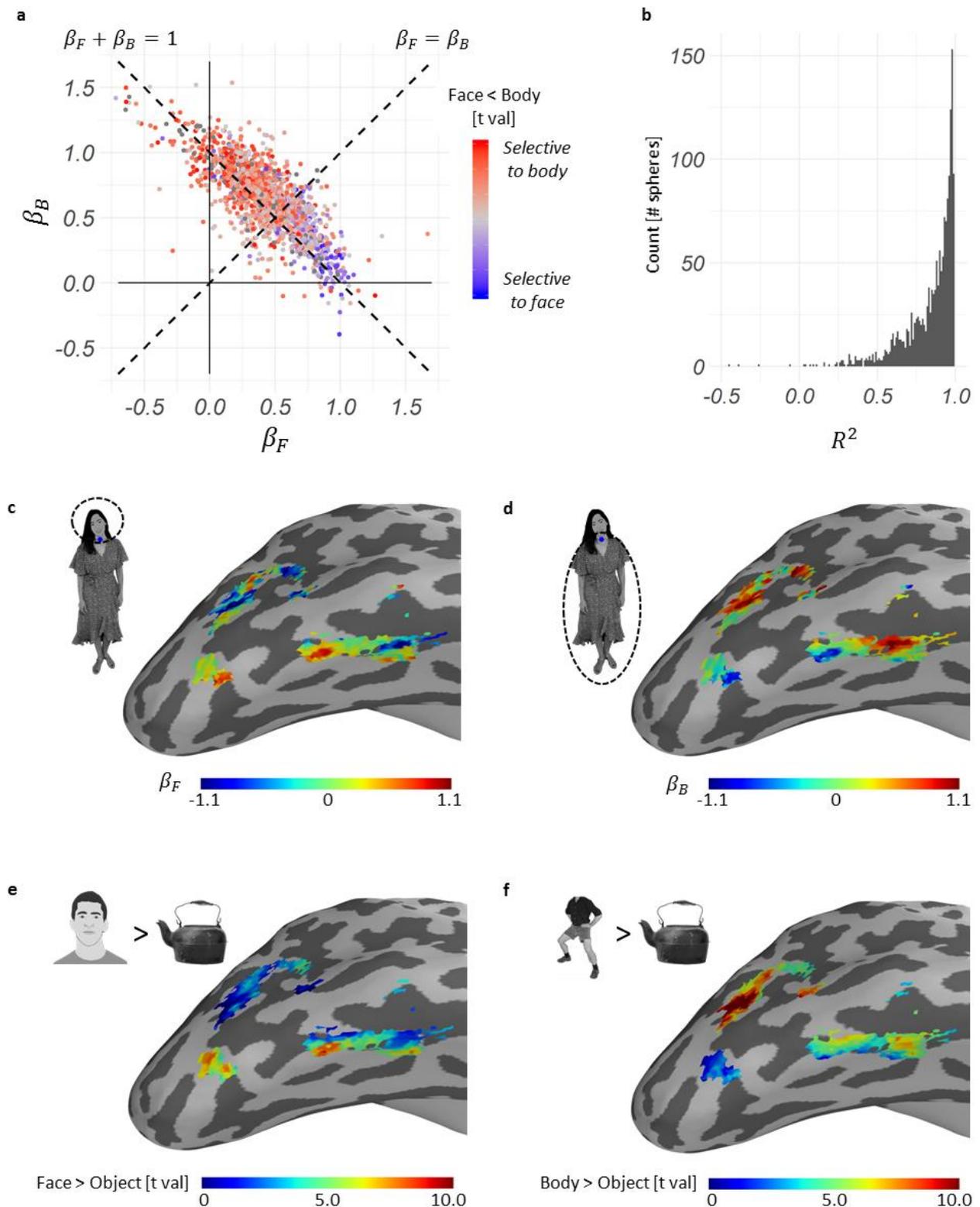
Table 1: Experiment 1 – Model comparison. In order to compare the proposed model predicted by the normalization equation (Fig. 1) to other models across all subjects, we used a Bayesian hierarchical model to predict the representation of the face+body stimulus based on the response to the face and the 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.

141 Next, we assessed the contribution of the face and the body to the face+body representation along the face
142 and body areas within the right occipito-temporal and lateral areas (see Figure 4–figure supplement 1 for
143 similar results of the left hemisphere). For each individual subject, we measured the response to face, body
144 and the face+body stimuli of each voxel in these anatomical locations. We then applied a moving mask of a
145 sphere of 27 voxels. For each sphere, we fitted a linear model to the responses of the voxels within the sphere
146 to predict the response to the face+body based on the responses to the face and the body (Fig. 1c).

147 Figure 4a depicts the beta coefficients for the face and the body, i.e. the contribution of the face and the body
148 to the face+body response, of all spheres within the face and body-selective cortices in the right occipito-
149 temporal and lateral areas of all subjects. The coefficients are scattered along the weighted mean line,
150 indicating a sum of coefficients of approximately 1 [mean sum=1.071, 95% confidence interval (C.I.): [1.036,
151 1.106], consistent with the normalization model prediction (Fig. 1e). Figure 4b displays the distribution of R^2
152 of the models for all spheres indicating a good fit of the linear model to the data [median $R^2=0.90$]. The color
153 of each dot indicates the selectivity to the face relative to the body, as measured by the independent
154 functional localizer. Furthermore, consistent with our predictions (Fig. 1d), the difference between the
155 contribution of the face and the body to the face+body representation, (i.e. the difference between the beta
156 coefficients) is correlated with the face and body-selectivity as measured by the independent functional
157 localizer data. To examine the statistical significance of this correlation, the correlation was computed for each
158 subject and transformed to a Z-Fisher score and the mean across subjects was compared to a null hypothesis
159 of a correlation of zero [mean fisher $z=0.458$, $t(14)=8.058$, $p<0.001$, 95% C.I. (0.321, 0.595)]. To reduce
160 statistical dependency in our dataset because of the overlapping moving mask, we used for the correlation
161 analysis an interleaved mask, taking only spheres that their center is not immediately adjacent to another.

162 Figure 4c-d shows the same face and body coefficients presented in Fig. 4a of a single subject placed on a
163 surface map of his brain. Figure 4e-f shows the distribution of category selectivity of the same subject within
164 the same region for the face and the body as indicated by the independent functional localizer data (indicated
165 in Fig. 4a by the color of the dots). Overall, Fig. 4 shows the correspondence between the selectivity and the
166 contribution of the face and the body to the face+body representation throughout the continuum of the face-

167 and body-selective regions: areas with high selectivity for faces and low selectivity for bodies show high
 168 contribution of the face to the face+body representation, while areas with low selectivity for faces and high
 169 selectivity for bodies show high contribution of the body to the face+body representation.



170

Figure 4: Experiment 1: (a) The beta coefficients of all spheres of all subjects in the face and body selective areas 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. (b) A histogram 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 Methods) (c-f) Results of a representative subject plotted on the cortical surface for voxels that were selective to either faces or bodies: (c) The contribution of the face to the face+body representation as indicated by the face regression coefficients (β_F). (d) The contribution of the body to the face+body representation as indicated by the body regression coefficients (β_B). (e) Selectivity to faces (t map of Face>Object). Selectivity was determined based on independent functional localizer data. (f) Selectivity to bodies (t map of Body>Object). Selectivity was determined based on independent functional localizer data. **Face images were replaced by illustration due to bioRxiv's policy on not including human faces within posted manuscripts.**

171 **Experiment 2 – The representation of a face+body and face+object in category-selective areas**

172 In a second fMRI study, 15 subjects were presented with face, body and face+body stimuli, as well as face,
173 object (wardrobe) and face+object stimuli forming a composite multi-category stimulus (see Fig. 2b). Similar
174 to Experiment 1, we estimated the contribution of the face and the body to the face+body response as well
175 as the contribution of the face and the object to the face+object response in different category-selective areas.
176 In this study we first show a replication of findings of Experiment 1 with the face and body stimuli based on
177 half of the data that was collected in Experiment 1 (3 runs instead of 6 runs). Then we show that the entire
178 category-selective area follows the same normalization framework both for face+body and for face+object
179 with the specific parameters determined by the local profile of category selectivity to the relevant categories.
180 For each individual subject we extracted the face-selective area, body-selective area and the object-selective
181 area based on independent functional localizer data that presented faces, bodies, objects and scrambled
182 objects. Each area was defined by voxels that show a significantly higher response to one category relative to
183 the combined response of the three categories. Thus, in this ROI analysis, areas that are selective to two
184 categories are excluded.

185 **ROI analysis:**

186 First, we ran the same analysis reported above to examine the contribution of the face and the body to the
187 face+body response in face- and body-selective areas. Findings replicated the results of Experiment 1 (Fig. 4A),
188 with both the face and the body contributing to the response to the face+body [FFA: $\beta_B^{(FB)}$: $p=0.002$. all other

189 p-values <0.001, see Fig. 5a and Figure 5–table supplement 1]. Furthermore, the relative contribution of the
190 face and the body varied as a function of the face and body selectivity (Equation 9), replicating the results of
191 Experiment 1: in the FFA the contribution of the face was higher than the contribution of the body [$\beta_F^{(FB)} -$
192 $\beta_B^{(FB)}$: mean=0.494, t(14)=4.169, p<0.001], while in the FBA the contribution of the body was higher than the
193 contribution of the face [$\beta_F^{(FB)} - \beta_B^{(FB)}$: mean=-0.382, t(13)=-3.442, p=0.004]. The sum of coefficients in both
194 face and body areas was again approximately 1 [mean sum (s.e.m.): FFA: 1.042 (0.066); FBA: 1.098 (0.054)]
195 consistent with the normalization model predictions (Equation 10).

196 Next, we performed similar analyses for the face+object stimuli. For each subject we fitted a regression model
197 for the 30 most selective voxels within the face-selective area and the object-selective area to predict the
198 response to the face+object based on the responses to the face and the object using the following equation:

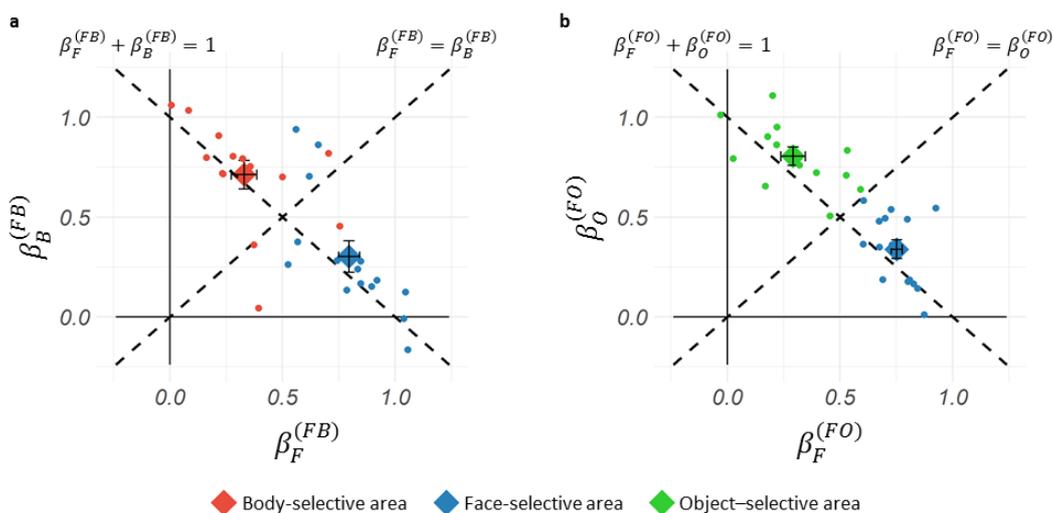
$$(Face + Object)_{PSC} = \beta_F^{(FO)} \cdot Face_{PSC} + \beta_O^{(FO)} \cdot Object_{PSC} + \varepsilon^{(FO)} \quad (2)$$

199

200 Similar to the face+body findings, the face- and object-selective areas showed a significant contribution of
201 both the face and the object to the face+object representation across all subjects, indicated by positive non-
202 zero coefficients of both the face and the object [all p-values<0.001, see Fig. 5b and Figure 5–table supplement
203 2]. In addition, the selectivity of the area determined the relative contribution of the face and the object to
204 the face+object representation (Fig. 1d). Specifically, we found that in the FFA, which is mainly selective to
205 faces, the contribution of the face was higher than the contribution of the object [$\beta_F^{(FO)} - \beta_O^{(FO)}$: mean=0.413,
206 t(14)=6.737, p<0.001], while in the object-selective area, the contribution of the object was higher than the
207 contribution of the face [$\beta_F^{(FO)} - \beta_O^{(FO)}$: mean=-0.512, t(12)=-5.753, p<0.001]. The sum of coefficients, again,
208 was 1 consistent with the normalization model (Fig. 1e) [mean sum (s.e.m.): FFA: 1.090 (0.043); Object area:
209 1.096 (0.047)].

210 The face+body stimuli are different from the face+object stimuli in that the former are a familiar combination
211 whereas the latter are not. Previous studies have predicted different patterns of representations to familiar
212 than non-familiar object combinations (Baldassano, Beck, & Fei-Fei, 2016; Kaiser & Peelen, 2018; Song, Luo,

213 Li, Xu, & Liu, 2013) whereas others did not find such difference (Baeck et al., 2013; Kaiser, Strnad, Seidl,
 214 Kastner, & Peelen, 2014). To examine whether the pattern of response to face+body and face+object is
 215 different, we ran a repeated measure ANOVA with Pair Type (face+body, face+object) and ROI (face-selective,
 216 body/object selective) as within-subject factors and the difference between the coefficients as a dependent
 217 variable. We excluded from this analysis subjects that did not had 30 voxels in each of the three ROIs (3
 218 subjects). As expected, the main effect of the ROI was significant [$F(1,11)=54.382$, $p<0.0001$], indicating that
 219 the selectivity of the ROI accounts for the relative contribution of each of the single categories to their multi-
 220 category stimuli. Importantly, we found no support for differences between Pair Type $F(1,11)=1.361$, $p=0.268$,
 221 as well as no interaction between the ROI and Pair Type $F(1,11)=0.024$, $p=0.808$]. Thus, the same normalization
 222 framework accounts for the two types of multi-category stimuli.



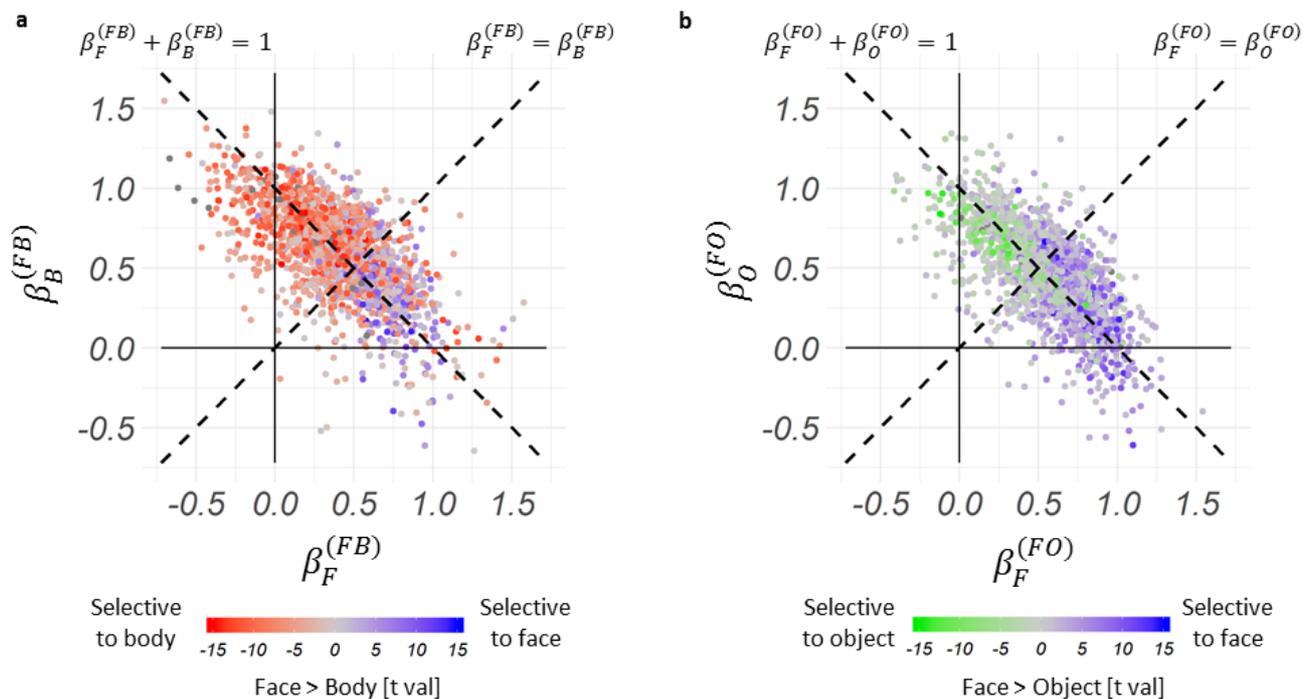
223

Figure 5: 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.).

224 Searchlight analysis

225 A similar searchlight analysis as described in Experiment 1 was performed for the face+body (equation (1))
226 and the face+object stimuli (equation (2)) in occipitotemporal and lateral areas that are selective to faces,
227 bodies or objects relative to scrambled objects. Figure 5A depicts the beta coefficients for the face and the
228 body, i.e. the contribution of the face and the body to the face+body response of all spheres within the
229 category-selective cortices of all subjects. Although this area contains also voxels that are selective to objects,
230 results are similar to Experiment 1. Specifically, the sum of coefficients is not significantly different from 1
231 [mean sum=1.013, $t(14)=0.638$, $p=0.534$, 95% C.I.=(0.970, 1.056)], and the difference in the contribution of
232 the face and the body to the face+body representation, (i.e. the difference between the beta coefficients) is
233 correlated with the selectivity to the face relative to the body as predicted [mean fisher $z=0.407$, $t(14)=8.444$,
234 $p<0.001$, 95% C.I.=(0.304, 0.511)], replicating the results of Experiment 1.

235 We performed the same analysis for the face+object model over the same searchlight area and found similar
236 results to the face+body findings (Fig. 6b): The beta-coefficients are scattered along the weighted mean line
237 with a sum of coefficients that is not significantly different from 1 [mean sum=1.015, $t(14)=1.490$, $p=0.158$,
238 95% C.I.=(0.993, 1.038)] and the difference in the contribution of the face and the object to the face+object
239 representation (i.e., the difference between the coefficients) is correlated with the selectivity to the face
240 relative to the object as expected [mean fisher $z=0.418$, $t(14)=11.193$, $p<0.001$, 95% C.I.=(0.338, 0.498)] (Fig.
241 1 d-e).



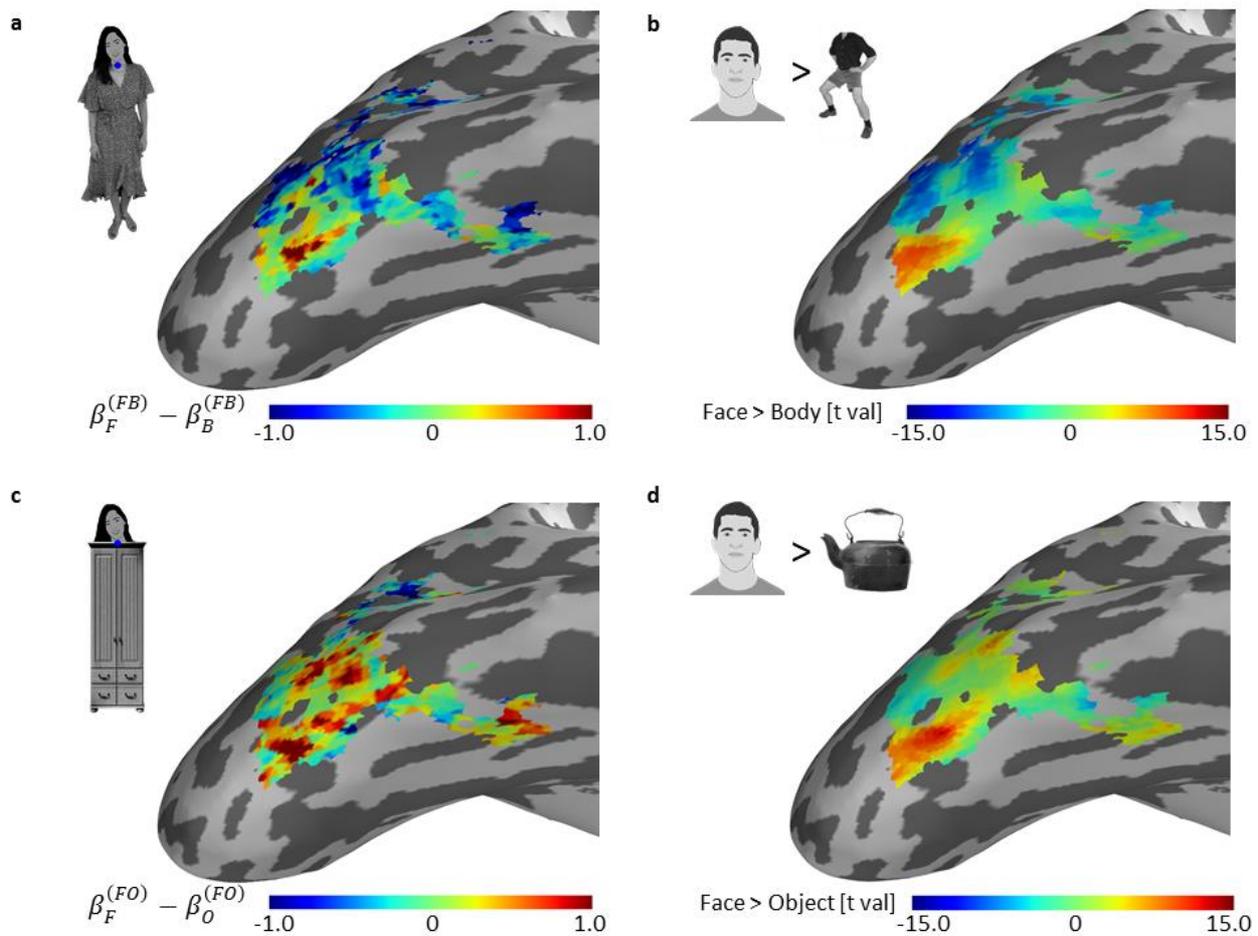
242

Figure 6: 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 for the face relative to the body based on independent functional localizer data. (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.

243

244 To compare the spatial distribution of the beta-coefficients and category selectivity, we plotted the difference
 245 between the coefficients and the difference between the selectivity to each pair of categories on brain surface
 246 maps (Fig. 7a-d). Figure 7a shows the difference between the face and body coefficients (i.e., difference
 247 between the contribution of the face and the contribution of the body to the face+body representation) of
 248 one representative subject along his category-selective cortex. Figure 7b shows the selectivity to the face
 249 relative to the selectivity to the body for the same subject as measured by the independent functional localizer
 250 data. It can be seen that cortical areas that show higher contribution of the face to the face+body
 251 representation correspond to face-selective clusters (red in both figures), and that areas that show higher
 252 contribution of the body to the face+body representation correspond to body-selective clusters (blue in both
 253 figures). Figure 7c shows the difference between the contribution of the face and the object to the face+object

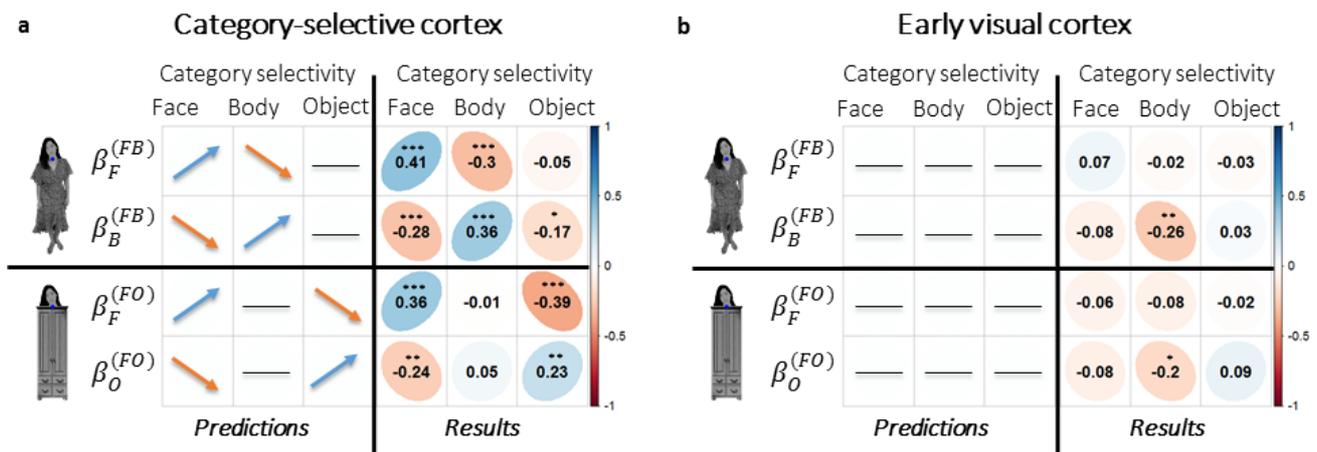
254 representation for the same subject. Figure 7d shows the selectivity to the face relative to the object based
255 on the functional localizer data. Similar to the face+body results, areas that show higher contribution of the
256 face to the face+object representation correspond to face-selective clusters (red in both figures), and areas
257 that show higher contribution of the object to the face+object representation correspond to object-selective
258 clusters (blue in both figures).



259

Figure 7: 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. 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. **Face images were replaced by illustrations due to bioRxiv's policy on not including human faces within posted manuscripts.**

260 Finally, we computed the correlation between the beta coefficients and category selectivity for each category
 261 in category-selective cortex as well as a control area - early visual cortex. As expected, correlations between
 262 the beta coefficients and category-selectivity were found in category-selective areas but not in early visual
 263 cortex (see Fig. 8).



264

Figure 8: Experiment 2 - Correlations between category-selectivity and model coefficients. Predictions and results of the correlations between the contributions of the single category to the multi-category response (beta coefficients of the models) and the category selectivity (based on the functional localizer data). (a) Results for category-selective cortex (Face, Body and Objects > Scrambled objects). A weighted mean model predicts that the contribution of each single category to the multi-category representation will be: (1) positively correlated to the selectivity to that same category; (2) negatively correlated to the selectivity to the other category that is present in the multi-category stimulus; and (3) not correlated with the selectivity to a category that is not present in the multi-category stimulus. Results are consistent with these predictions. (b) Results for early visual cortex (EVC). Since EVC does not show selectivity to object categories, we predict that category-selectivity will not be correlated with the contribution of the single object categories to the multi-category stimulus response. Values presented in the figure are the mean across subjects for the fisher z. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ significant correlations corrected for multiple comparisons. **Face images were replaced by illustrations due to bioRxiv's policy on not including human faces within posted manuscripts.**

265 Discussion

266 Our findings show that the functional organization of category-selective cortex determines the representation
 267 of multiple-category stimuli. Whereas previous studies have primarily focused on the role of this functional
 268 organization in the representation of single isolated stimuli (Grill-Spector & Weiner, 2014), visual scenes are
 269 typically composed of multiple stimuli. Thus, a great challenge of the visual system is to resolve competition
 270 among multiple stimuli (Cohen, Konkle, Rhee, Nakayama, & Alvarez, 2014; Kastner & Ungerleider, 2001;

271 Peelen & Kastner, 2014) and generate a veridical representation of the objects that compose a complex visual
272 scene. Here we show that by applying a normalization mechanism, the functional organization of neighboring
273 clusters of category-selective neurons generates different representations along high-level visual cortex
274 according to the profile of category selectivity. Areas with high concentration of neurons selective to a single
275 category give priority to the preferred stimulus, filtering out the non-preferred stimuli. This operation enables
276 hard-wired de-cluttering at early stages of visual processing (see also Bao & Tsao, 2018). Areas with a mixed
277 population of category-selective neurons, enable similar, possibly competitive, representation to different
278 categories that can be modulated by higher-level cognitive mechanisms according to task demands (Desimone
279 & Duncan, 1995; Reynolds & Heeger, 2009).

280 The fMRI results reported in the current study are consistent with predictions derived from the normalization
281 model that were developed based on single unit recording data (Fig. 1). The fMRI findings add to the neuronal
282 findings by demonstrating the correspondence between the functional organization of high-level visual cortex
283 and the representation of multi-category stimuli. This is enabled by two features of the fMRI signal: First, the
284 magnitude of category-selectivity measured with fMRI provides a measure of the homogeneity of the
285 normalization pool, an important factor in the normalization equation. Second, fMRI enables exploring the
286 pattern of response across a large, continuous area of cortex. This pattern of response indicates that the
287 representation of the multi-category stimulus changes gradually in a way that corresponds to the profile of
288 category-selectivity (Fig. 4, 6, 7). These results propose a continuous mode of organization of high-level visual
289 cortex, rather than the more common, discrete-like depiction of category-selective cortex.

290 Our findings provide a general framework that accounts for previous reports of single cell recording and
291 neuroimaging studies that reported either a mean response (Macevoy & Epstein, 2009; Zoccolan et al., 2005),
292 a weighted mean response (Baeck et al., 2013) or a max response (Bao & Tsao, 2018; Reddy et al., 2009) to
293 multiple stimuli in different areas of category-selective cortex. We show that the relative contribution of each
294 stimulus to the response of the compound stimulus varies along the weighted mean line and that this variation
295 is accounted for by variation in category selectivity (Fig. 4, 6, 7). It is noteworthy that our findings **do not** imply
296 that the representation of a multi-category stimulus of a single neuron is determined solely by *its own*

297 selectivity to each of the stimulus categories (the nominator in the normalization equation). Category-
298 selectivity measured with fMRI estimates the selectivity of the surrounding neurons (the denominator in the
299 normalization equation) and therefore provides an estimate of the selectivity of the normalization pool and
300 its effect on the response to multiple stimuli.

301 The normalization model was confirmed in our study by several measures: First, the model predicts a specific
302 correspondence between the coefficients of the model and the selectivity of a cortical area, which was
303 confirmed both in an ROI and a searchlight analysis. Second, we fit the data to alternative models, including a
304 model with an interaction term and a model with a non-zero intercept, and found that the normalization
305 model best accounts for the response to a multi-category stimulus (Table 1). Third, we performed the same
306 analysis in early-visual cortex (Fig. 8), a control area that shows no selectivity to these object categories and
307 therefore category-selectivity is not expected to explain the contribution of each of the single categories to
308 the multi-category response, and indeed found no such relationship. Last, we ruled out an alternative
309 explanation for the results, suggesting that the weighted mean is a result of the saturation of the BOLD signal
310 to the multi-category stimuli, leaving the normalization model as the most probable explanation of our results.

311 Several neuroimaging studies that examined the representation of multiple stimuli have asked whether the
312 response to a pair of stimuli deviates from a simple mean model, in particular for pairs of stimuli that show a
313 meaningful relationship between them (Baldassano et al., 2016; Fisher & Freiwald, 2015; Kaiser & Peelen,
314 2017; Kaiser et al., 2014; MacEvoy & Epstein, 2011; Song et al., 2013). In these studies, a deviation from a
315 simple mean response was considered as evidence for integration or a holistic representation of the complex
316 stimulus. The main advantage of the linear model we used here is that it provides us with a direct measure of
317 the type of deviation from the mean that the data show and can therefore decide between a weighted mean
318 response, an additive response or a non-additive response. Our findings show that the deviation from mean
319 reflects a weighted mean response. We found no evidence for a non-additive response to the combined
320 stimulus and therefore no support for a holistic representation. This was the case both for the meaningful pair
321 of face+body stimuli as well as for the non-meaningful face+wardrobe pair that generated similar
322 representations. Similar results were reported by Baeck et al. (2013) that found the same representations for

323 related and unrelated pairs of objects. Thus, the normalization mechanism operates in a similar manner for
324 related and unrelated pairs of stimuli in object-category selective cortex.

325 Three additional studies that examined the representation of the whole person are noteworthy. Kaiser et al.
326 (2014) reported no deviation from the mean in the response to a face and a body in a person-selective area
327 (area defined by a whole person > objects). This area is likely to correspond to the overlap area reported in
328 our study that is selective to both faces and bodies, and therefore consistent with our findings (Figure 2). Song
329 et al. (2013) reported that only the right FFA showed a deviation from the mean for the response of the whole
330 person and interpreted that as evidence for a holistic representation. This deviation, however, may reflect a
331 weighted mean response rather than a non-additive response. Finally, Fisher & Freiwald (2015) examined the
332 contribution of the face and body to the whole person in a monkey fMRI study and found a super-additive
333 (more than the sum) response in anterior but not posterior face areas, in particular, in area AF in the dorsal
334 bank of the superior temporal sulcus. The human analog of area AF is likely to be in the superior temporal
335 sulcus (Yovel & Freiwald, 2013) an area that we did not examine in the current study that may apply a different
336 mode of operation than the ventral visual cortex.

337 To summarize, our findings reveal a general framework of operation according to which the contribution of
338 each stimulus to the representation of multiple stimuli in a given cortical area is determined by its profile of
339 category selectivity, in line with a normalization mechanism. We therefore suggest that the functional
340 organization of neighboring patches of neurons, each selective to a single or more categories, enables a
341 flexible representation of complex visual scenes, where both de-cluttering and competition operate in
342 different cortical areas, using the same type of neurons and the same mechanism of normalization. This type
343 of organization may permit high-level cognitive processes to bias the response to any of these different
344 representations according to task demands making the taxing operation of understanding complex visual
345 scenes dynamic and flexible.

346 **Methods**

347 **Participants**

348 **Experiment 1**

349 Fifteen healthy volunteers (6 women, ages 19-37, 13 right-handed) with normal or corrected-to-normal vision
350 participated in Experiment 1. Participants were paid \$15/hr. All participants provided written informed
351 consent to participate in the study, which was approved by the ethics committees of the Sheba Medical Center
352 and Tel Aviv University, and performed in accordance with relevant guidelines and regulations.

353 The sample size (N=15) chosen for this study was similar to sample size of other fMRI studies that examined
354 the representation of multiple objects in high-level visual cortex (10-15 subjects per experiment) (see for
355 example: Baeck et al., 2013; Baldassano et al., 2016; Kaiser & Peelen, 2017; Kaiser et al., 2014; Macevoy &
356 Epstein, 2009; MacEvoy & Epstein, 2011; Reddy et al., 2009; Song et al., 2013)

357 **Experiment 2**

358 Seventeen healthy volunteers (11 women, ages 20-30, 14 right-handed) that did not participate in Experiment
359 1, with normal or corrected-to-normal vision participated in Experiment 2. Two participants were excluded
360 from analysis due to technical difficulties. Participants were paid \$15/hr. All participants provided a written
361 informed consent to participate in the study, which was approved by the ethics committees of the Sheba
362 Medical Center and Tel Aviv University, and performed in accordance with relevant guidelines and regulations.

363 **Stimuli**

364 **Experiment 1:**

365 **Main Experiment**

366 Stimuli consisted of 40 grey-scale images of a whole person standing in a straight frontal posture with their
367 background removed downloaded from the internet (20 men and 20 women identities). Each image of a
368 person was cut into two parts approximately in the neck area resulting in a face stimulus and a headless body
369 stimulus for each identity (Figure. 1A). The isolated face and body stimuli were presented in the same location
370 they occupied in the whole person stimulus. A blue fixation dot was presented at a constant location around

371 the neck on the screen across all conditions (at the center and upper part of the display) (Figure 1A). The size
372 of the whole person image was approximately 3.5X12.2 degrees of visual angle.

373 ***Functional Localizer***

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

377 ***Experiment 2:***

378 ***Main Experiment***

379 Experiment 2 contained two main parts: a face-body part and a face-object part. For the face-body part we
380 used the same stimuli as in Experiment 1. For the face-object part we used pictures of faces, wardrobes and
381 faces-above-wardrobes (Figure 1B). The face stimuli were the same 40 images of faces used in Experiment 1.
382 For the object stimuli we used 40 images of grey-scale wardrobes with their background removed that were
383 taken from the internet. We digitally manipulated the images of the wardrobes so that the object location,
384 size (number of pixels on the screen), contrast and luminance will be matched to the 40 pictures of headless
385 bodies from Experiment 1. The face+object stimuli were created by placing the wardrobe images right below
386 the face in the same location of the body, i.e. a face above a wardrobe with no gap between them. A blue
387 fixation dot was presented at a constant location on the screen across all conditions right over the neck in the
388 same location as in Experiment 1) (Figure 1B). The size of the face+body stimuli as well as the face+object pair
389 was approximately 3.5X12.2 degrees of visual angle.

390 ***Functional Localizer***

391 Localizer stimuli were grey-scale pictures of faces, headless-bodies, non-leaving objects, and scrambled
392 objects (Figure 1C). The size of the stimuli was approximately 5.5X5.5 degrees of visual angle.

393 ***Apparatus and Procedure***

394 ***fMRI acquisition parameters***

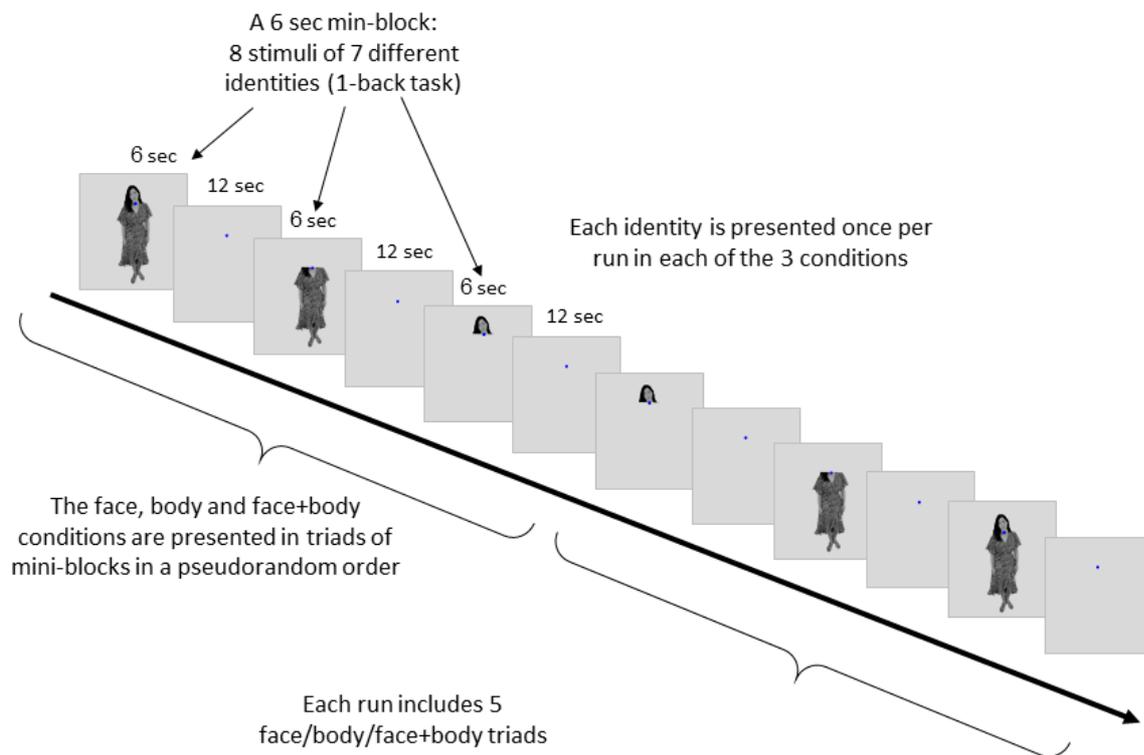
395 fMRI data were acquired in a 3T Siemens MAGNETOM Prisma MRI scanner in Tel Aviv University, using a 64-
396 channel head coil. Echo-planar volumes were acquired with the following parameters: repetition time
397 (TR) = 2 s, echo time = 30 ms, flip angle = 82°, 64 slices per TR, multi-band acceleration factor = 2, acceleration
398 factor PE = 2, slice thickness = 2 mm, field of view = 20 cm and 100 × 100 matrix, resulting in a voxel size of
399 2 × 2 × 2 mm. Stimuli were presented with Matlab (The MathWorks Inc.) and Psychtoolbox (Brainard, 1997;
400 Kleiner et al., 2007) and displayed on a 32" high definition LCD screen (NordicNeuroLab) viewed by the
401 participants at a distance of 155 cm through a mirror located in the scanner. Anatomical MPRAGE images were
402 collected with 1 × 1 × 1 mm resolution, echo time = 2.88 ms, TR = 2.53 s.

403 ***Experiment 1:***

404 The study included a single recording session with six runs of the main experiment and three runs of functional
405 localizer.

406 ***Main Experiment:*** Each of the six runs included 5 triads of face, body and face+body mini-blocks. Fig. 9 shows
407 an example of two such triads. The order of face, body and face+body mini-blocks within each triad was
408 counter-balanced across triads and runs. Each mini-block included eight stimuli of which 7 were of different
409 identities and one identity repeated for the 1-back task. The identities presented in the face, body and
410 face+body mini-blocks within a triad were different. Thus, each run included face, body and face+body stimuli
411 of 35 different identities (7 identities × 5 triads). The 35 identities were randomly chosen from the set of 40
412 identities. Each mini-block lasted 6 seconds was followed by 12 seconds of fixation. A single stimulus display
413 time was 0.325 s, inter-stimulus-interval was 0.425 s. Subjects performed a 1-back task (one repeated stimulus
414 in each block). Each run began with a six seconds (3 TRs) fixation (dummy scan) and lasted a total of 276
415 seconds (138 TRs).

416 Subjects were instructed to maintain fixation throughout the run and their eye movements were recorded
417 with an Eye tracker (EyeLink®).



418

Figure 9: Experimental procedure. Each run had 15 blocks of 3 conditions (5 blocks each). See Methods for a full description of the procedure. **Face images were replaced by illustrations in this manuscript due to bioRxiv's policy on not including human faces within posted manuscripts. The experiment stimuli included real human photos.**

419 **Functional Localizer:** Each run of the functional localizer included 21 blocks: 5 baseline fixation blocks and 4
420 blocks for each of the four experimental conditions: faces, bodies, objects and persons (analysis of person
421 condition is not included in this paper). Each block presented 20 stimuli of 18 different images of which two
422 repeated twice for a 1-back task. Each stimulus was presented for 0.4 sec with 0.4 sec Inter-stimulus interval.
423 Each block lasted 16 seconds. Each run began with a six seconds fixation (3 TRs) and lasted a total of 342
424 seconds (171 TRs).

425 **Experiment 2:**

426 The experiment included a single recording session with six runs of the main experiment and three runs of
427 localizer.

428 **Main experiment:** The main experiment included 3 runs of face, body and face+body stimuli identical to
429 Experiment 1. In addition, 3 runs of face, object and face+object stimuli were presented using the same design
430 used for the face and body runs (Fig. 9). The face+object runs were presented before the face+body runs to
431 avoid the priming of a body in the object and face+object mini-blocks. Subjects were instructed to maintain
432 fixation throughout the run and their eye movements were recorded with an Eye tracker (EyeLink®). (See
433 Figure 5 - Supplementary Figure 1 for the comparison of eye tracker data between conditions).

434 **Functional Localizer:** The functional localizer included four conditions of faces, bodies, objects and scrambled
435 objects. All other parameters of the functional localizer runs were identical to Experiment 1.

436 **Data analyses**

437 ***fMRI Data Analysis and preprocessing***

438 fMRI analysis was performed using SPM12 software, Matlab (The MathWorks Inc.) and R (R Development Core
439 Team, 2011) costumed scripts, STAN (Carpenter et al., 2017) for Bayesian model fitting and Freesurfer (Dale,
440 Fischl, & Sereno, 1999), pysurfer (<https://pysurfer.github.io>) and Python (<http://www.python.org>) costumed
441 scripts for the surface generation and presentation. The code that was used for data analyses is available at
442 https://github.com/LibiKI/multiple_objects_fmri_analysis. The first three volumes in each run were acquired
443 during a blank screen display and were discarded from the analysis as “dummy scans”. The data were then
444 preprocessed using realignment to the mean of the functional volumes and co-registration to the anatomical
445 image (rigid body transformation). Spatial smoothing was performed for the localizer data only (5 mm). A GLM
446 was run with separate regressors for each run and for each condition, including 24 nuisance motion regressors
447 for each run (6 rigid body motion transformation, 6 motion derivatives, 6 square of motion and 6 derivatives
448 of square of motion), and a baseline regressor for each run. In addition, a "scrubbing" method (Power, Barnes,
449 Snyder, Schlaggar, & Petersen, 2012) was applied for every volume with frame-displacement (FD) > 0.9 by
450 adding a nuisance regressor with a value of 1 for that specific volume and zeros for all other volumes. Percent
451 signal change (PSC) for each voxel was calculated for each experimental condition in each run by dividing the
452 beta weight for that regressor by the beta weight of the baseline for that run.

453 **Experiment 1:**

454 ***Region of interest (ROI) analysis***

455 Based on the functional localizer data, face- and body-selective voxels were defined individually for each
456 subject using contrast t -maps. Regions of interest (ROI) were defined as clusters (>10 voxels) of voxels
457 selective to a given category ($p < 10^{-4}$) within specific anatomical locations: (1) Fusiform face area (FFA):
458 Face $>$ Object within the Fusiform gyrus; (2) Fusiform body area (FBA): Body $>$ Object within the Fusiform gyrus.
459 The overlap areas were defined as the conjunction between face and body selective ROIs, included all voxels
460 that were both face- and body-selective as described above. The 30 most selective voxels from each ROI within
461 the right hemisphere were analyzed with the main experiment data. ROIs with less than 30 voxels were
462 excluded from further analysis (see Figure 3–figure supplement 1 for the stability of the results across different
463 number of voxels even with very low number of subjects).

464 ***Linear model fitting***

465 The mean percent signal change (PSC) across runs to the face, the body and the face+body conditions from
466 the main experiment data were extracted for each voxel within each ROI of each subject. For each subject, we
467 fitted a linear model to predict the response to the face+body based on the response to the face and the
468 body (Reddy et al., 2009) (equation 1). The features of the model were the response of a single voxel to a single
469 condition. The sum of the coefficients was not pre-determined to sum up to 1 but was determined solely by
470 fitting of the model to the data. We calculated the mean of the beta coefficients of the model, the mean
471 difference between the coefficients and their mean sum across subjects.

472 To examine whether the linear model based on the normalization mechanism (Fig. 1c, equation 1) is the best
473 fit to the data, we estimated a Bayesian hierarchical model to predict the response to a face+body based on
474 the response to the face and the body including the data from all subjects for each ROI. In addition, we
475 estimated two other Bayesian hierarchical models: one with an addition of an intercept term, and another with
476 the addition of an interaction between the face and the body. We then calculated Bayes factors to compare
477 the models.

478 ***Univariate voxel-wise analysis***

479 For each voxel within each ROI we compared the PSC to the face+body to the maximum PSC to the face and
480 the body, and calculate the proportion of voxels that showed smaller response to the face+body, i.e., $face +$
481 $body < \max(face, body)$. This analysis was done to assure that weighted mean response is not due to
482 saturation of the BOLD response to face+body.

483 ***Searchlight analysis***

484 For the searchlight analysis, we defined a face and body-selective region based on the localizer data by the
485 contrast $[(Face+Body)/2 > Object]$ ($p < 10^{-4}$) within the ventro-temporal and lateral occipital cortex. For each
486 subject we defined a moving mask of a sphere of 27 voxels. To reduce statistical dependency in our dataset
487 because of the overlapping moving mask, we used for the correlation analysis an interleaved mask, taking only
488 spheres that their center is not immediately adjacent to another. For each sphere we fitted a linear model
489 with its voxel data as features to predict the response to the face+body based on the response to the face and
490 the body. The beta coefficients of these models represent the contribution of the face and the body to the
491 response of the face+body of each sphere within the searchlight area. We then plotted a surface map of the
492 beta coefficients of all spheres within the searchlight area to present the spatial distribution of the beta
493 coefficients. We calculated R^2 for each sphere and the median R^2 across all spheres. Because the model we
494 used does not include an intercept, negative R^2 values indicate that a model is worse in predicting
495 the dependent variable compared to a model that includes only an intercept. To examine the relationship
496 between the difference between the face and body beta coefficients and the selectivity to face over a body
497 (i.e., the t values of the contrast Face>Body from the independent functional localizer data) we performed a
498 Pearson correlation across subjects. To assess the level of significance of the correlations, the correlation
499 values were transformed to fisher's Z, and a one-sample t-test was used against a null-hypothesis of zero.

500 **Experiment 2:**

501 ***ROI Analysis***

502 Based on the functional localizer data, face- body- and object-selective voxels were defined individually for
503 each subject. Regions of interest (ROI) were defined as clusters (>10 voxels) of category selective voxels ($p < 10^{-4}$)
504 within specific anatomical locations that show preference to a single category relative to all other categories:
505 (1) Fusiform face area (FFA): Face > Body, Object & Scrambled-object within the Fusiform gyrus; (2) Fusiform
506 body area (FBA): Body > Face, Object & Scrambled-object within the Fusiform gyrus; (3) Ventral object area:
507 Object > Face, Body & Scrambled-object within the medial part of the ventral temporal cortex. As in
508 Experiment 1, the 30 most selective voxels from each ROI in the right hemisphere were chosen for model
509 fitting. ROIs with less than 30 voxels were excluded from further ROI analysis.

510 The model fitting described in Experiment 1 was used to separately predict the response to the face+body
511 based on the response to the face and the body (equation 1) and to predict the response to the face+object
512 based on the response to the face and the object (equation 2). Similar to Experiment 1, we calculated the beta
513 coefficients of the model, the mean difference between the coefficients and their mean sum for each model
514 for each subject.

515 To examine whether the pattern of response to face+body and face+object is different, we ran a repeated
516 measure ANOVA with Pair Type (face+body, face+object) and ROI (face-selective, body/object selective) as
517 within-subject factors and the difference between the coefficients as a dependent variable. We excluded from
518 this analysis subjects that did not had 30 voxels for all three ROIs (3 subjects excluded).

519 ***Searchlight analysis***

520 For the searchlight analysis, we defined a category-selective region based on the localizer data by the contrast
521 $[(\text{Face} + \text{Body} + \text{Object})/3 > \text{Scrambled Object}]$ ($p < 10^{-4}$) within the Ventral-Temporal cortex and Lateral
522 Occipital-Temporal cortex. A similar analysis that was performed in Experiment 1 was performed separately
523 to the face and body runs and the face and object runs.

524 As a control area we also defined early visual cortex (EVC). EVC was extracted by performing an inverse
525 normalization from an MNI space Brodmann area 17 mask to each subject's native space. We matched the
526 number of voxels in EVC to the number of voxels within the category-selective region by randomly choosing

527 voxels from EVC. To further examine the correspondence between category selectivity and the contribution
528 of each stimulus to the representation of the combined stimulus, correlations were computed between the
529 selectivity to each of the three stimuli (face, body, object), measured by the t-value (each single category
530 against all other categories), and the coefficients for each of the stimuli. The correlation values of all subjects
531 were transformed to fisher's Z, to examine the level of statistical significance against a null-hypothesis of a
532 zero correlation. This analysis was performed in object-selective cortex and in early visual cortex (EVC).

533 **Data Availability**

534 The code that was used for data analysis is available at
535 https://github.com/LibiKI/multiple_objects_fmri_analysis. Data that was collected in this study will be
536 available after publication in a shared repository (<https://openneuro.org/>).

537 **Acknowledgments**

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

541 **Author Contributions**

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

544 **References**

- 545 Baeck, A., Wagemans, J., & de Beeck, H. P. (2013). The distributed representation of random and meaningful
546 object pairs in human occipitotemporal cortex: The weighted average as a general rule. *NEUROIMAGE*,
547 *70*, 37–47. <https://doi.org/10.1016/j.neuroimage.2012.12.023>
- 548 Baldassano, C., Beck, D. M., & Fei-Fei, L. (2016). Human-object interactions are more than the sum of their

- 549 parts. *Cerebral Cortex*, 1–13. <https://doi.org/10.1093/cercor/bhw077>
- 550 Bao, P., & Tsao, D. Y. (2018). Representation of multiple objects in macaque category-selective areas. *Nature*
551 *Communications*, 9(1), 1–16. <https://doi.org/10.1038/s41467-018-04126-7>
- 552 Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*. <https://doi.org/10.1163/156856897X00357>
- 553 Carandini, M., & Heeger, D. (2012). Normalization as a canonical neural computation. *Nature Reviews*
554 *Neuroscience*, (November), 1–12. <https://doi.org/10.1038/nrn3136>
- 555 Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A
556 probabilistic programming language. *Journal of Statistical Software*.
557 <https://doi.org/10.18637/jss.v076.i01>
- 558 Cohen, M. A., Konkle, T., Rhee, J. Y., Nakayama, K., & Alvarez, G. A. (2014). Processing multiple visual objects
559 is limited by overlap in neural channels. *Proceedings of the National Academy of Sciences of the United*
560 *States of America*, 111(24), 8955–8960. <https://doi.org/10.1073/pnas.1317860111>
- 561 Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage*, 9(2), 179–194.
562 <https://doi.org/10.1006/nimg.1998.0395>
- 563 Desimone, R., & Duncan, J. (1995). Neural mechanism of selective visual attention. *Annu Rev Neurosci*, 18,
564 193–222.
- 565 Downing, P. E., Chan, A. W. Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual
566 cortex. *Cerebral Cortex*. <https://doi.org/10.1093/cercor/bhj086>
- 567 Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A Cortical Area Selective for Visual Processing
568 of the Human Body. *Science*, 293(5539), 2470–2473. <https://doi.org/10.1126/science.1063414>
- 569 Fisher, C., & Freiwald, W. A. (2015). Whole-agent selectivity within the macaque face-processing system.
570 *Proceedings of the National Academy of Sciences*, 112(47), 201512378.

- 571 <https://doi.org/10.1073/pnas.1512378112>
- 572 Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role
573 in categorization. *Nature Reviews. Neuroscience*, 15(8), 536–548. <https://doi.org/10.1038/nrn3747>
- 574 Kaiser, D., & Peelen, M. V. (2018). Transformation from independent to integrative coding of multi-object
575 arrangements in human visual cortex. *NeuroImage*. <https://doi.org/10.1016/j.neuroimage.2017.12.065>
- 576 Kaiser, D., & Peelen, M. V. (2017). Transformation From Independent To Integrative Coding Of Multi-Object
577 Arrangements In Human Visual Cortex.
- 578 Kaiser, D., Strnad, L., Seidl, K. N., Kastner, S., & Peelen, M. V. (2014). Whole person-evoked fMRI activity
579 patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-
580 evoked activity patterns. *Journal of Neurophysiology*, 111(1), 82–90.
581 <https://doi.org/10.1152/jn.00371.2013>
- 582 Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate
583 cortex specialized for face perception. *The Journal of Neuroscience*, 17(11), 4302–4311. Retrieved from
584 <http://www.ncbi.nlm.nih.gov/pubmed/9151747>
- 585 Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of
586 faces. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 361(1476),
587 2109–2128. <https://doi.org/10.1098/rstb.2006.1934>
- 588 Kastner, S., & Ungerleider, L. G. (2001). The neural basis of biased competition in human visual cortex.
589 *Neuropsychologia*. [https://doi.org/10.1016/S0028-3932\(01\)00116-6](https://doi.org/10.1016/S0028-3932(01)00116-6)
- 590 Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in
591 Psychtoolbox-3? *Perception*. <https://doi.org/10.1068/v070821>
- 592 Macevoy, S. P., & Epstein, R. a. (2009). Decoding the representation of multiple simultaneous objects in human
593 occipitotemporal cortex. *Current Biology : CB*, 19(11), 943–947.

- 594 <https://doi.org/10.1016/j.cub.2009.04.020>
- 595 MacEvoy, S. P., & Epstein, R. a. (2011). Constructing scenes from objects in human occipitotemporal cortex.
596 *Nature Neuroscience*, 14(10), 1323–1329. <https://doi.org/10.1038/nn.2903>
- 597 Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., ... Tootell, R. B. (1995). Object-
598 related activity revealed by functional magnetic resonance imaging in human occipital cortex.
599 *Proceedings of the National Academy of Sciences*, 92(18), 8135 LP – 8139.
600 <https://doi.org/10.1073/pnas.92.18.8135>
- 601 Op de Beeck, H. P., Brants, M., Baeck, A., & Wagemans, J. (2010). Distributed subordinate specificity for bodies,
602 faces, and buildings in human ventral visual cortex. *NeuroImage*.
603 <https://doi.org/10.1016/j.neuroimage.2009.11.022>
- 604 Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of*
605 *Neurophysiology*, 93(1), 603–608. <https://doi.org/10.1152/jn.00513.2004>
- 606 Peelen, M. V., & Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends*
607 *in Cognitive Sciences*, 18(5), 242–250. <https://doi.org/10.1016/j.tics.2014.02.004>
- 608 Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., & Kastner, S. (2009). Neural
609 Representations of Faces and Body Parts in Macaque and Human Cortex: A Comparative fMRI Study.
610 *Journal of Neurophysiology*. <https://doi.org/10.1152/jn.91198.2008>
- 611 Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic
612 correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59(3),
613 2142–2154. <https://doi.org/10.1016/J.NEUROIMAGE.2011.10.018>
- 614 Premereur, E., Taubert, J., Janssen, P., Vogels, R., & Vanduffel, W. (2016). Effective Connectivity Reveals
615 Largely Independent Parallel Networks of Face and Body Patches. *Current Biology*.
616 <https://doi.org/10.1016/j.cub.2016.09.059>

- 617 R Development Core Team, R. (2011). *R: A Language and Environment for Statistical Computing*. R Foundation
618 *for Statistical Computing*. <https://doi.org/10.1007/978-3-540-74686-7>
- 619 Reddy, L., Kanwisher, N. G., & Vanrullen, R. (2009). Attention and biased competition in multi-voxel object
620 representations. *Proceedings of the National Academy of Sciences of the United States of America*,
621 *106*(50), 21447–21452.
- 622 Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque
623 areas V2 and V4. *Journal of Neuroscience*.
- 624 Reynolds, J. H., & Heeger, D. J. (2009). The Normalization Model of Attention. *Neuron*, *61*(2), 168–185.
625 <https://doi.org/10.1016/j.neuron.2009.01.002>
- 626 Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate Face and Body Selectivity on the Fusiform
627 Gyrus. *Journal of Neuroscience*, *25*(47), 11055–11059. [https://doi.org/10.1523/JNEUROSCI.2621-](https://doi.org/10.1523/JNEUROSCI.2621-05.2005)
628 [05.2005](https://doi.org/10.1523/JNEUROSCI.2621-05.2005)
- 629 Song, Y., Luo, Y. L. L., Li, X., Xu, M., & Liu, J. (2013). Representation of Contextually Related Multiple Objects in
630 the Human Ventral Visual Pathway. *Journal of Cognitive Neuroscience*, *25*(8), 1261–1269.
631 https://doi.org/10.1162/jocn_a_00406
- 632 Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A Cortical Region Consisting Entirely
633 of Face-Selective Cells. *Science*, *311*(5761), 670 LP – 674. <https://doi.org/10.1126/science.1119983>
- 634 Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans.
635 *Proceedings of the National Academy of Sciences*, *105*(49), 19514–19519.
636 <https://doi.org/10.1073/PNAS.0809662105>
- 637 Weiner, K. S., & Grill-Spector, K. (2013). Neural representations of faces and limbs neighbor in human high-
638 level visual cortex: evidence for a new organization principle. *Psychological Research*, *77*(1), 74–97.
639 <https://doi.org/10.1007/s00426-011-0392-x>

640 Yovel, G., & Freiwald, W. A. (2013). Face recognition systems in monkey and human: Are they the same thing?

641 *F1000Prime Reports*. <https://doi.org/10.12703/P5-10>

642 Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron*.

643 <https://doi.org/10.1016/j.neuron.2004.11.018>

644 Zoccolan, D., Cox, D. D., & DiCarlo, J. J. (2005). Multiple object response normalization in monkey

645 inferotemporal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*,

646 25(36), 8150–8164. <https://doi.org/10.1523/JNEUROSCI.2058-05.2005>

647

648 **List of supplementary material**

649 **Supplementary figures**

650 Supplementary Figure 1: Derivations of the normalization equation.

651 Figure 3–figure supplement 1: ROI analysis with different number of voxels.

652 Figure 3–figure supplement 2: Results for lateral face and body areas OFA and EBA.

653 Figure 4–figure supplement 1: Searchlight results for left hemisphere.

654 Figure 5-figure supplement 1: Eye tracker data distribution.

655 **Supplementary tables**

656 Figure 3–table supplement 1: Experiment 1.

657 Figure 5–table supplement 1: Experiment 2 - Beta coefficients of face and body, ROI analysis.

658 Figure 5–table supplement 2: Experiment 2 - Beta coefficients of face and object, ROI analysis.