Modeling neural tuning in human ventral temporal cortex predicts the perception of visual ambiguity

Rosenke, Mona[1], Davidenko, Nicolas[2], Grill-Spector, Kalanit[1,3], Weiner, Kevin S.[4,5]

[1] Psychology Department, Stanford University;
[2] University of California, Santa Cruz, Department of Psychology
[3] Neuroscience Institute, Stanford University
[4] Psychology Department, University of California, Berkeley
[5] Helen Wills Neuroscience Institute, University of California, Berkeley

Corresponding Author:
Mona Rosenke, Department of Psychology, Stanford University, Stanford CA,
rosenke@stanford.edu

ABSTRACT

We have an amazing ability to categorize objects in the world around us with millisecond precision. Nevertheless, how cortical regions in human ventral temporal cortex (VTC), which is critical for categorization, support this behavioral ability, is presently unknown. Here, we used a model-based approach to examine the relationship between neural tuning and behavioral tuning using morphed silhouettes of faces and hands, which are animate categories processed in cortically adjacent regions in VTC of humans and non-human primates. Surprisingly, our results reveal that the combination of neural tuning from regions selective for different domains more accurately explains behavioral categorization than neural tuning from just one region selective for either domain. Furthermore, we built a generalizable model that predicts a person's behavioral tuning after being trained on neural tuning from other people. These findings support a new theoretical idea: cortical regions from different domain-specific networks work together to achieve accurate perceptual categorization.

We categorize objects within our visual world in a split second[1,2]. Despite this remarkable ability, it is still presently unknown how regions and networks within the human brain support efficient and accurate visual categorization across multiple domains, especially when faced with visual ambiguity[3–5]. Prior research has shown that human ventral temporal cortex (VTC) is critical for visual categorization[6,7] and also, that shape is a key feature for neural categorization[8–10]. Additionally, it has been hypothesized that multiple implementational features contribute to - and are integral for - efficient categorization[7]. One such feature is the organization of functional clusters, or regions, into different networks that are specialized for processing different domains[6,11–16]. Indeed, recent findings in both humans and non-human primates have revealed causal relationships between neural responses and perception within specific regions of domain-specific networks – for example, separate regions are critical for face[17–19] and body[20] perception. Additionally, several theories have been proposed to explain how these regions and networks form in the human brain, as well as the many functional and behavioral purposes they may serve[11,21–24]. However, it is presently unknown whether or how brain regions from different domain-specific networks may work together to achieve a behavioral goal such as perceptual categorization.

This gap in knowledge persists because previous research has largely focused on understanding how either (a) neural responses from one functional region contribute to perception[3–5,25] or (b) neural responses are computationally transformed from regions that are positioned in early cortical stages of a network compared to later stages[26,27]. Building on this solid empirical foundation, a recent trend in neuroimaging studies has begun to unveil how information regarding different domains is combined within functional regions. For example, several studies have shown that separate information about faces and bodies can be integrated within a functional region to generate new information that represents a person[28,29]. While this approach is critical in that it sheds light on how information from two separate stimuli and two different domains is integrated and read out by a single region, it remains untested whether (and how) information is integrated *across* multiple regions when presented with a stimulus that shares features from two categorical domains (**Fig. 1**). For instance, in the case of the previous findings, whether images of people are better coded from neural responses of *both* face- and body-selective regions together rather than separately.

A rigorous way to begin to fill this gap in knowledge is with images that are generated on a known continuum of shape space in which (a) the ends of the continuum represent shapes from

two different categories and (b) there is a clear categorical transition such that the center shape of the continuum is composed of equal parts of both categories (**Fig. 1**). Such a space enables a clear quantification of how neural and behavioral responses change when exposed to stimuli that cross category boundaries. Here, we generated a shape space containing a morphed continuum between faces and hands using the silhouette methodology[8,30]. Using Mechanical Turk, we then validated that individuals perceived the center stimulus as equally face- and hand-like. In a separate group of participants, we implemented a functional magnetic resonance imaging (fMRI) experiment during which we measured brain activity in face- and body-selective regions within human VTC in response to different blocks of face/hand morphs. We then conducted a behavioral experiment in the same participants outside the scanner to examine the relationship between neural and behavioral tuning.

For the first time, our results show that the neural tuning of functional regions that are selective for different domains together predict behavioral categorization of ambiguous images better than any one region by itself. Specifically, the opposite weights of neural tuning from two regions that are selective for different domains more accurately explains human categorical judgements than neural responses from a single domain-specific region. Moreover, we show that a model trained on neural tuning from a group of subjects can predict a new person's behavioral tuning. These findings support a new theoretical idea in which cortical regions from different domain-specific networks work together to achieve accurate perceptual categorization.

RESULTS

We first generated face-hand morphs using the silhouette methodology[30]. This approach generated a large set of carefully controlled images consisting of continuous morphs between faces and hands (**Fig. 1A**; Materials and Methods). Then, we calibrated the stimuli on Mechanical Turk (www.mturk.com). Calibration results indicated that (a) the morph continuum was centered, which means that the middle morph level was ambiguous and equally likely to be perceived as either a face or a hand by individual observers (**Fig. 1B**), and (b) the perceptual variability across exemplars within each morph level was similar (e.g., level 1 exemplars were perceived as being as variable as level 4 exemplars, **Fig. 1C**).

Using these stimuli, we conducted two types of experiments in 14 adults (8 females, ages 22 - 44): (1) a functional magnetic resonance imaging (fMRI) block-design experiment during which we measured mean neural tuning to face-hand morphs in face- and body-selective regions in VTC (**Fig. 1D**) and (2) a behavioral categorization experiment outside the scanner during which we measured behavioral tuning to face-hand morphs (**Fig. 1D**).

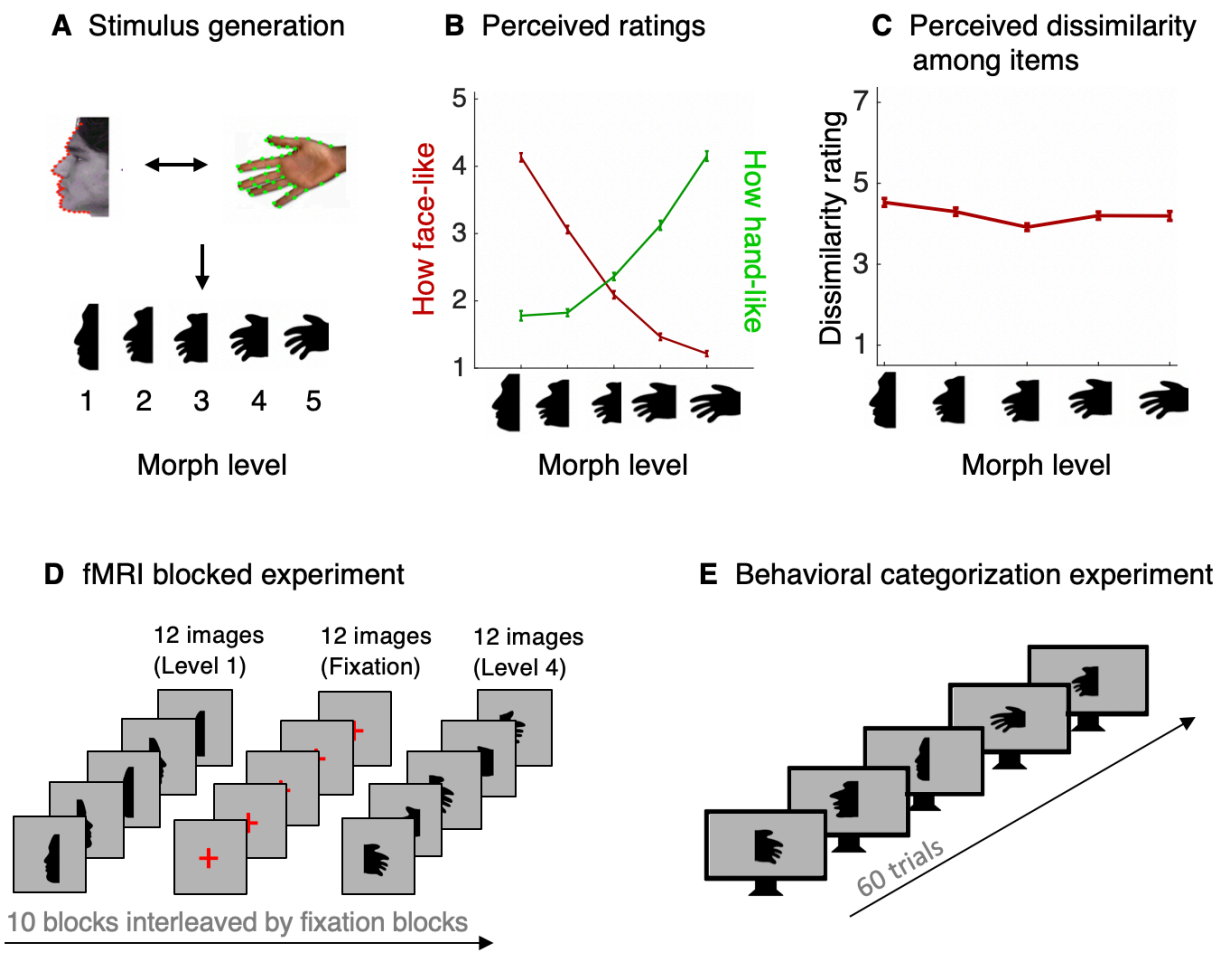

Figure 1. Stimulus calibration and experimental setup. A. Construction of face-hand silhouettes. *Top:* The contours of a profile face and open hand photograph are parameterized with 36 points in which points of maximum and minimum curvature are matched. *Bottom:* Five morph levels between faces and hands are created by interpolating the key point coordinates of parameterized faces and hands. The silhouettes are rendered by applying bi-cubic splines to the key points and shading the resulting figure in black. **B.** Morphed stimuli were calibrated on Mechanical Turk to produce intermediate stimuli (level 3) that were as face-like as they were hand-like (see Methods). *X-axis*: morph levels ranging from face (level-1, left) to hand (level-5, right). *Y-axis:* face-likeness ratings (red) and hand-likeness ratings (green) averaged across MTurk participants. *Errorbars:* Standard error across the 60 exemplars at each morph level. **C.** Matching perceptual variability. Each morph level consisted of multiple (12) exemplars. For each level, we

calibrated the perceptual variability across those exemplars based on pairwise dissimilarity ratings to ensure equal variability of stimuli across all morph levels. *X-axis:* Morph level; *Y-axis*: Dissimilarity scores (higher scores indicate greater perceptual variability across the exemplars) averaged across MTurk participants. *Errorbars:* Standard error across 12 pairs of exemplars at each morph level. **D.** Each participant underwent a block-design fMRI experiment during which they viewed blocks containing images of carefully controlled silhouettes. 12 silhouettes from the same morph level were shown within each block. Subjects pressed a button when the same image appeared twice in a row (1-back task). **E.** Following the fMRI blocked experiment, each subject underwent a separate behavioral categorization experiment outside the scanner. Participants viewed silhouettes (N=60) and performed a forced-choice task to categorize each silhouette as a face or a hand. Presentation order was randomized across morph levels and participants.

Face- and body-selective regions in ventral temporal cortex (VTC) show differential tuning to morphed face-hand silhouettes

To examine how neural responses in face- and body-selective regions in VTC respond to continuous morphs between faces and hands (referred to as *neural tuning*), we used fMRI to first localize three face- and body-selective functional regions of interest (ROIs): (1) pFus-faces/FFA-1, which is a face-selective region located in the posterior-lateral portion of the fusiform gyrus (FG) extending into the occipito-temporal sulcus (OTS), (2) OTS-bodies, which is a body-selective region located within the OTS extending into the lateral FG[31–33], and (3) mFus-faces/FFA-2, which is a face-selective region located within the FG extending laterally from the anterior tip of the mid-fusiform sulcus[33–35] (**Fig. 2**). Then, we extracted neural responses (z-scored beta values from general linear model fits; see Methods) from these three regions during a second, independent experiment during which participants viewed blocks of perceptually ambiguous stimuli that consisted of morphs between face and hand silhouettes (**Fig. 1D**).

FMRI measurements during the morph experiment revealed that face- and body-selective ROIs in human VTC: (1) show graded responses to morphs between face and hand silhouettes, and (2) exhibit differential neural tuning to carefully controlled visual stimuli. Specifically, neural responses in face-selective regions were highest to silhouettes that were most face-like, and decreased the more the silhouette looked like a hand (right hemisphere: **Fig. 2B,D**; left hemisphere: **Supplementary Fig. 1A,C**; significant main effect of morph level in a 1-way ANOVA in each ROI: mFus-faces: $F(4,120) = 5.08$, $p<0.001$, pFus-faces: $F(4,110) = 4.52$, $p<0.005$). In contrast, neural tuning curves in OTS-bodies showed the opposite graded response pattern: the highest neural responses were to hand silhouettes and the lowest neural responses were to face silhouettes

(F(4,80) = 5.93, p<0.001). These differences in neural tuning among face- and body-selective regions in VTC are statistically supported by a 3-way repeated measures ANOVA with factors of ROI, morph level, and hemisphere that showed a significant morph level x ROI interaction (F(8,305)=16.45, p <0.001; no significant difference between hemispheres (F(1,168) = 5.6458e-29, p = 1.0). While neural tuning curves are differentiable between face- and body-selective regions, an interesting question then becomes: *Do neural responses in face- and body-selective regions relate to behavioral categorization of face and hand morphs?*

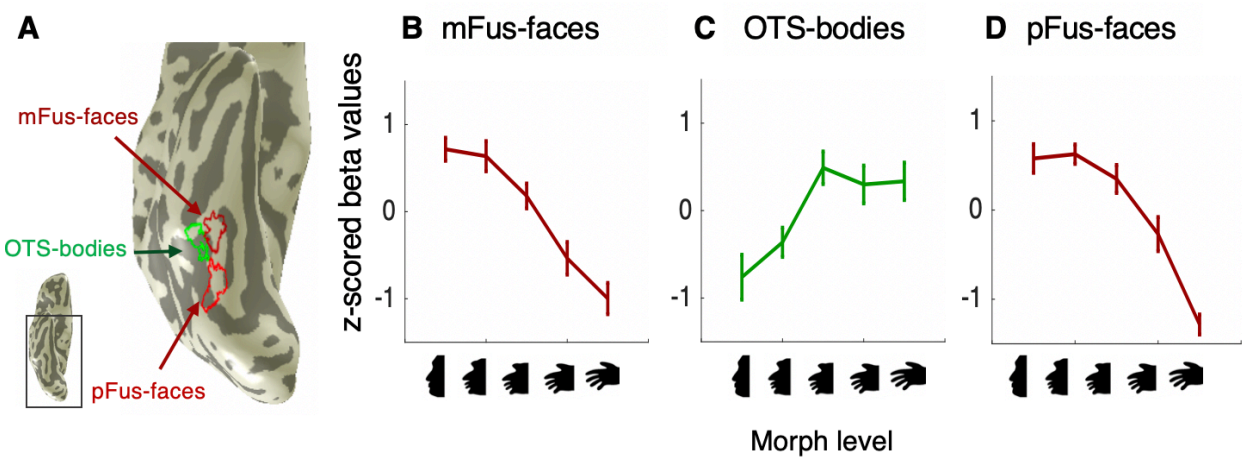

Figure 2. Face- and limb-selective regions in ventral temporal cortex (VTC) display differential neural tuning. A. Zoomed image showing the inflated cortical surface in ventral view of an example participant illustrating three regions of interest (fROIs). From posterior to anterior: pFus-faces (red), OTS-bodies (green), and mFus-faces (dark red). *Dark gray:* sulci. *Light gray*: gyri. *Inset:* Ventral view of the inflated cortical surface showing the location of the zoomed region. **B-D.** The mean neural tuning across participants for each fROI of the right hemisphere. Normalized responses were derived by first computing the beta coefficients to each one of the 5 morph levels of face-hand silhouettes and then transforming the beta coefficients into z-scores. *X-axis*: face-hand silhouette morph level. *Y-axis:* Z-scored beta values. *Errorbars:* Standard error across participants. See **Supplementary Fig. 3** for exact number of participants per ROI (N=11-14). See **Supplementary Fig. 1** for left hemisphere.

Categorization behavior of face-hand morphs is best explained by multiple domain-specific regions from different networks

To examine and model the relationship between neural and behavioral tuning, we first used the same face-hand silhouettes that participants previously viewed inside the scanner to measure

how participants perceptually categorized the same stimuli outside the scanner. Specifically, participants viewed 60 silhouettes randomly drawn from the 5 morph levels (12 exemplars of each morphing level) and were instructed to perceptually classify each image as either a face or a hand, in a self-paced manner (**Fig. 1E**). We refer to the profiles relating these forced-choice behavioral responses to the morph levels as *behavioral tuning* throughout the paper (see Methods).

Behavioral results showed that participants perceptually differentiated face from hand silhouettes. That is, participants classified the two more face-like morph-levels as faces in the majority of trials (**Fig. 3**, proportion face responses (mean±standard deviation (SD)): .97±.05 and .99±.02, respectively), while they classified the two more hand-like morph levels as hands (proportion face response: .06±.08 and .01±.03, respectively). Categorization of the middle morph level was more variable across participants. Despite ambiguity, images in this middle, or intermediate, morph level 3 were classified slightly more as faces (proportion face response: .64±.26). Interestingly, the neural tuning in each face-and body-selective region was non-linear and more broadly tuned than the perceptual ratings of face-likeness or hand-likeness. For example, level-3 stimuli were rated relatively low on both perceptual scales (**Figs. 1 and 3**), but elicited relatively strong response in both face- and body-selective regions (**Fig. 2**). This indicates that the neural tuning of these regions does not simply reflect how 'face-like' or 'hand-like' the stimuli appear.

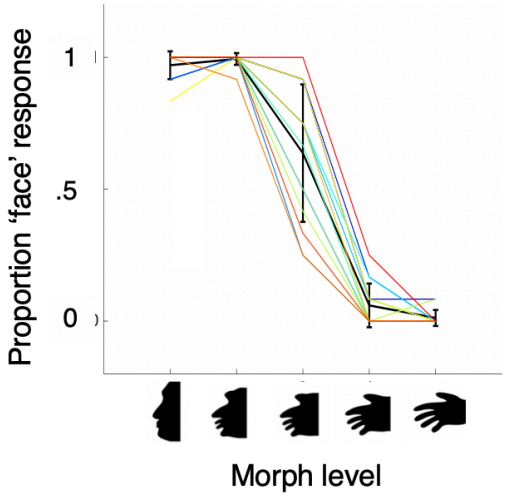

Figure 3. Behaviorally, participants effectively categorize the face-hand morphs. Behavioral tuning for each subject (colored lines) as well as the mean tuning across subjects (black line). Behavioral responses at each morph level indicate the participant's average performance across 12 trials. Proportion 'face' responses decline with morph

level showing a categorical transition. The highest variability in behavioral responses across participants occurred for the middle, ambiguous stimulus. *X-axis:* morph level of silhouette (1 = face, 5 = hand). *Y-axis:* proportion face response. *Errorbars:* Standard deviation across the 14 participants.

To test if there is a relationship between neural tuning and behavioral tuning, we conducted a linear regression analysis to examine if and how neural tuning in VTC is associated with behavioral tuning. In this analysis, the behavioral tuning of each subject was the dependent variable and the neural tuning of each subject was used as the independent variable (**Fig. 4A**). We considered two models: (1) a linear regression model examining if neural responses from a single face- or body-selective region could predict behavioral categorization (**Fig. 4B, 1-ROI model**), and (2) a separate linear regression model examining if neural responses from multiple face- and body-selective regions could predict behavioral categorization (**Fig. 4B, Multi-ROI model**).

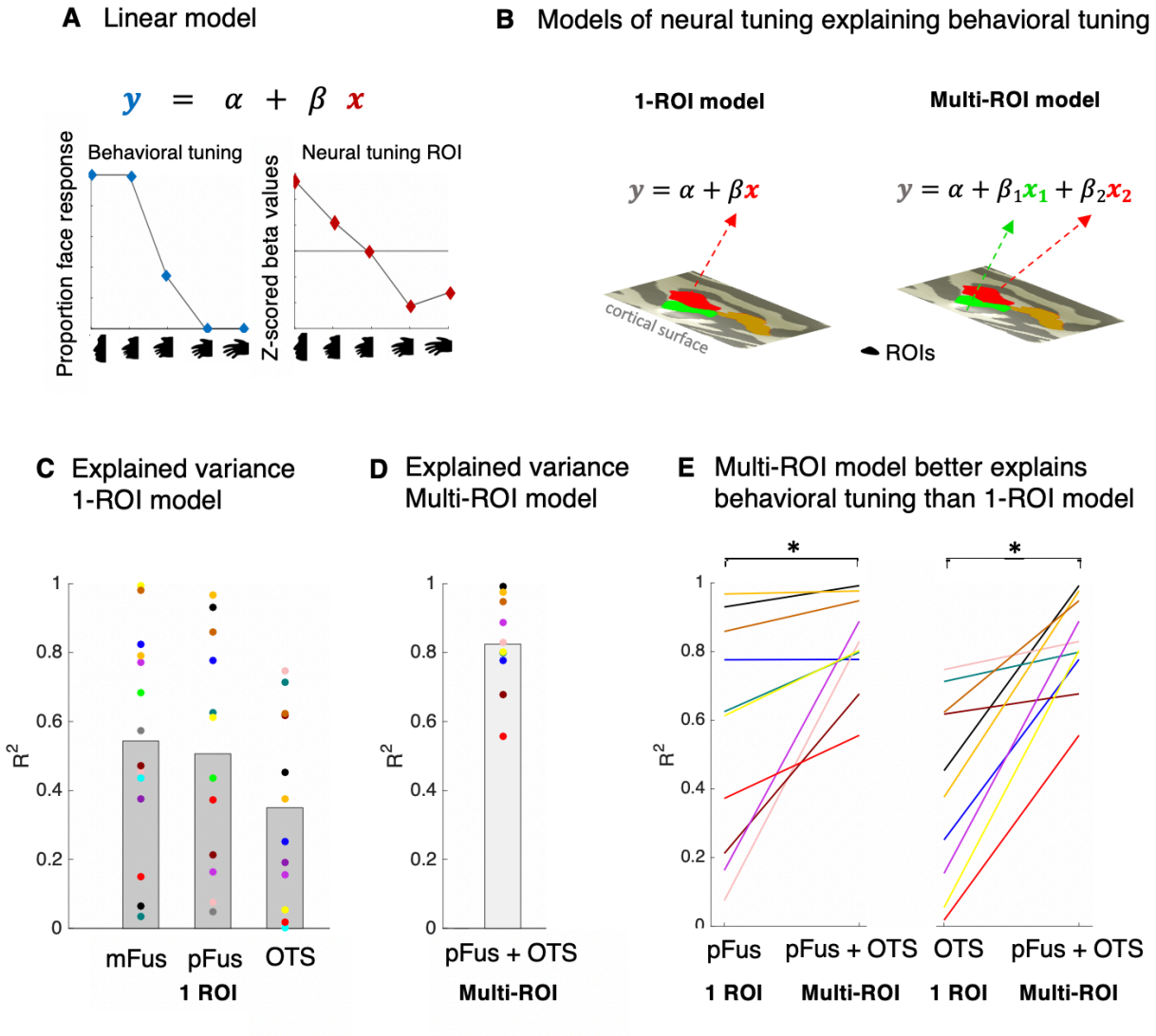

Figure 4. A stepwise linear regression model reveals that behavioral tuning can be predicted by a linear combination of the neural tuning from pFus-faces and OTS-bodies. A. *Top:* A linear regression model was used to relate behavioral and neural turning. α = model intercept, β = beta coefficient for predictor, x =neural predictor, y = behavioral tuning. *Left:* Behavioral tuning. *Right:* Neural tuning. **B.** We tested two different types of linear models: (1) a 1-ROI model, which used neural tuning from one ROI as a predictor for behavioral tuning, and (2) a multi-ROI model (MRM), which used neural tuning from multiple ROIs as separate predictors based on results from a stepwise linear regression. **C.** Explained variance (R^2) from the 1-ROI model in the right hemisphere. Neural tuning from each ROI was a significant predictor of behavioral tuning. *X-axis:* ROIs used as predictors for the linear regression model. *Y-axis:* Variance explained of the behavioral data by three 1-ROI models. Bars indicate the mean across participants, and each dot represents data from one participant. The color of each participant is matched across models and data in Fig. 3. **D.** Variance explained of the behavioral data by the MRM including right pFus-faces and right OTS-bodies. Conventions same as C. **E.** Improvement in variance explained by the MRM compared to the 1-ROI model. Each line is a participant. Line colors correspond to (C). *Asterisk*: significant improvement, $p<.05$, permutation testing.

Results of the 1-ROI model showed that neural tuning and behavioral tuning are related to one another. Specifically, on average, behavioral categorization performance from individual participants was linearly related to their neural responses from either face- or body-selective regions. That is, neural responses from face (pFus-faces, mFus-faces) or body (OTS-bodies) ROIs predicted a substantial amount of variance for each participant's behavioral responses (R^2± standard error (SE), left hemisphere (LH) and right hemisphere (RH), respectively: mFus-faces = .43±.10 and .54±.09; pFus-faces = 47±.09 and .51±.09; OTS-bodies = .41±.06 and .35±.07; RH: **Fig. 4C;** LH: **Supplementary Fig. 2A**). However, there were extensive individual differences in model fits across participants: the variance explained was close to one for some participants, but it was close to zero for others (R^2 min-max: 0 - .99; **Fig. 4C**).

In order to test if a model using neural responses from multiple ROIs better explains the behavioral data across participants, we implemented a stepwise linear regression model to test which ROIs significantly explained behavioral tuning. The stepwise regression was applied to data from all subjects and neural responses from all three regions (**Fig. 4B**). Interestingly, this model revealed that two neural predictors, pFus-faces (LH and RH, *p*s<0.001) and OTS-bodies (LH and RH, *p*s<0.001), significantly explained behavioral categorization. In contrast, mFus-faces did not significantly explain additional variance (LH and RH: *p*s = 0.18). Consequently, the remaining analyses only consider pFus-faces and OTS-bodies as predictors, which we refer to as the Multi-ROI model (MRM).

In the MRM, a linear combination of the neural tunings from pFus-faces and OTS-bodies is related to behavioral tuning in each individual participant. Notably, across participants, the MRM model explained 0.71±.03 and 0.83±.04 (R^2±SE) of the variance explaining behavioral and neural categorization in the left and right hemispheres, respectively (**Fig. 4D, Supplementary Fig. 2B**). The variance explained by the MRM model was significantly higher than the variance explained by the 1-ROI model, which we validated in two ways: (1) Akaike's Information Criterion (AIC) and (2) paired permutation tests between each of the 1-ROI models and the MRM. In terms of the former, lower is better and the AIC (mean across subjects ± SE for LH, RH) was as follows: pFus-faces = 5.40±1.24, 4.55±1.46; OTS-bodies = 6.96±0.57, 7.27±0.63; pFus-faces + OTS-bodies = 5.19±0.66, 0.88±1.69. In terms of the latter, paired permutation tests between each of the 1-ROI models and the MRM showed that significantly more variance was explained

with the MRM (all ps<0.05, **Fig. 4E**). Differences in model predictions across hemisphere are not significant (paired permutation tests between left and right hemisphere model: *p*s>0.05). All permutation tests were Bonferroni corrected (see Methods for details).

While these results are intriguing, it may just be the case that neural signals from multiple ROIs – irrespective of their selectivity – are sufficient to explain behavioral categorization better than a single ROI alone. To test this possibility, we implemented an identical MRM except that we used neural tunings from two other functional regions as predictors that neither (1) illustrated a specific categorical preference (retinotopic area VO-2[36]), nor (2) were selective for faces or bodies/limbs (Parahippocampal Place Area (PPA)/CoS-places[37–39]). This model had a lower average explained variance (R^2±SE: .52±.06 (LH) and .58±.08 (RH), **Supplementary Fig. 2C**) and a higher AIC criterion (7.64±0.79 (LH) and 5.48±1.55 (RH)) compared to the MRM model (permutation test, $p < .05$). These analyses rule out the possibility that neural responses from any two regions are sufficient to explain categorization behavior during this particular task.

A closer look at the relationship that was derived by the MRM model revealed that the model-estimated behavioral tuning closely matched the actual behavioral tuning of a given participant (**Fig. 5A,** compare to 1-ROI model estimations in **Supplementary Fig. 3**). Perhaps even more impressively, the model consistently assigned opposing weights to pFus-faces (LH: β_{pFus} = 0.22±.06, RH: β_{pFus} = 0.39±.06) and OTS-bodies (LH: β_{OTS} = -0.22±.05, RH: β_{OTS} = -0.22±.07, **Fig. 5B**). This indicates that the differential neural tuning between these two regions is likely tightly linked to the behavioral tuning responsible for the individual's categorization of these ambiguous morphs. Despite this intriguing possibility, there is still inter-subject variability as to how well neural tuning relates to behavioral tuning. This opens the following question: *Can a model built from neural tuning in one group of participants predict the behavioral tuning in a new participant*?

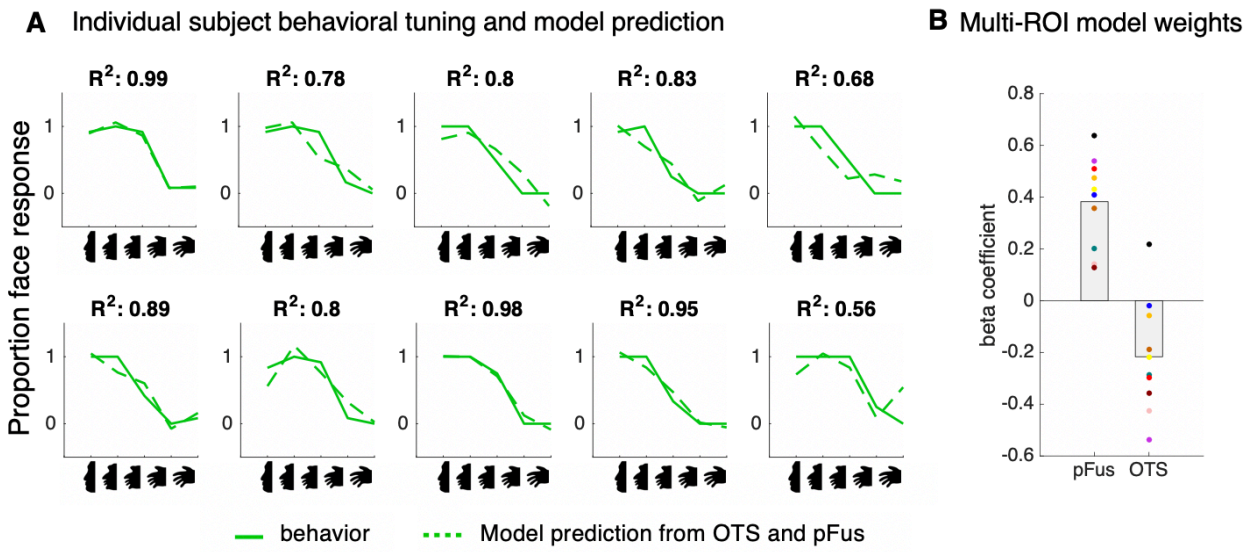

Figure 5. Individual differences in the predictions of the multi-ROI model. A. Each panel shows the behavioral tuning (solid) for a participant and the MRM prediction using the neural tuning from both pFus-faces and OTS-bodies (pFus+OTS) in the right hemisphere (dashed). *X-axis:* morph level; *Y-axis:* proportion 'face' responses. Explained variance for the respective participant is indicated at the top of each subplot. **B.** Beta coefficients of the right pFus+OTS MRM. The model assigned positive weights to pFus-faces and negative weights to OTS-bodies for most participants. *X-axis:* predictors; *Y-axis:* beta coefficients. Bars indicate the mean coefficient, and each dot indicates a participant. Colors are the same as in prior figures.

A model that includes neural tuning from different domain-specific regions in one group of participants accurately predicts behavioral tuning in new participants

If the relationship between neural and behavioral tuning reflects a general relationship between brain and behavior that is similar across individuals, then a model built from neural responses in one group of participants should be capable of predicting behavioral tuning in a new person. To test this hypothesis, we built an MRM model relating neural responses from pFus-faces and OTS-bodies from N-1 participants to their behavioral tuning. This model was implemented separately for each hemisphere and estimated the coefficients for pFus-faces and OTS-bodies (see Methods for details). We then used the coefficients to predict the behavioral tuning in a new subject by multiplying that subject's neural tuning of pFus-faces and OTS-bodies with the derived model coefficients (**Fig. 6A**). Finally, this approach was repeated using a 10-fold cross-validation procedure, using every subject once as a left-out test subject.

Impressively, this cross-validated MRM using right hemisphere data accurately predicted behavioral tuning of the left-out subject (mean $R^2 \pm$ SE: 0.70 ± 0.05, **Supplementary Fig. 4**). In

contrast, predicting behavior from an MRM based on left hemisphere data was significantly lower (mean $R^2 \pm$ SE: 0.50 ± 0.09; $p < 0.01$, paired permutation testing). In addition, the cross-validated MRM that used neural tuning from regions selective for faces and bodies in the right hemisphere out-performed a cross-validated MRM that used neural tuning from regions that were not selective for faces and bodies (cross-validated MRM based on neural responses from VO-2 and CoS-places/PPA: LH and RH mean $R^2 \pm$ SE: 0.26 ± 0.06 and 0.23 ± 0.08; permutation tests: $p<0.01$), while the left hemisphere cross-validated MRM-model was not significantly different from the control model in its prediction accuracy of behavioral tuning in left-out-subjects (permutation tests: *p*s > 0.2).

Investigating the right hemisphere cross-validated MRM more closely shows that the model captures the trend of each subject's behavioral tuning even though the estimated beta coefficients were generated using neural tuning from an independent group of participants (**Fig. 6B**). The cross-validated MRM displayed differential weights for pFus-faces and OTS-bodies, which was consistent with the within-subject MRM in that positive coefficients were derived for pFus-faces ($\beta = 0.31 \pm 0.003$) and negative coefficients for OTS-bodies ($\beta = -0.22 \pm .004$). While the cross-validated MRM predicted behavior well over-all, it is worth emphasizing that neural predictions deviated most from behavioral tuning at the intermediate morph (level 3), which is the stimulus associated with the largest within-subject variability (**Fig. 3**). Interestingly, it is also worth emphasizing that participants who showed a lower explained variance in the within-subject MRM (**Fig. 5B**), also showed a lower explained variance in the cross-validated MRM (**Fig. 6B**).

Altogether, these analyses reveal that (1) opposed neural tuning from functional regions in human VTC predict behavioral tuning and (2) this relationship is participant-general. That is, behavioral tuning in one participant can be predicted from the differential neural tuning across regions with predictor weights derived from a separate group of participants.

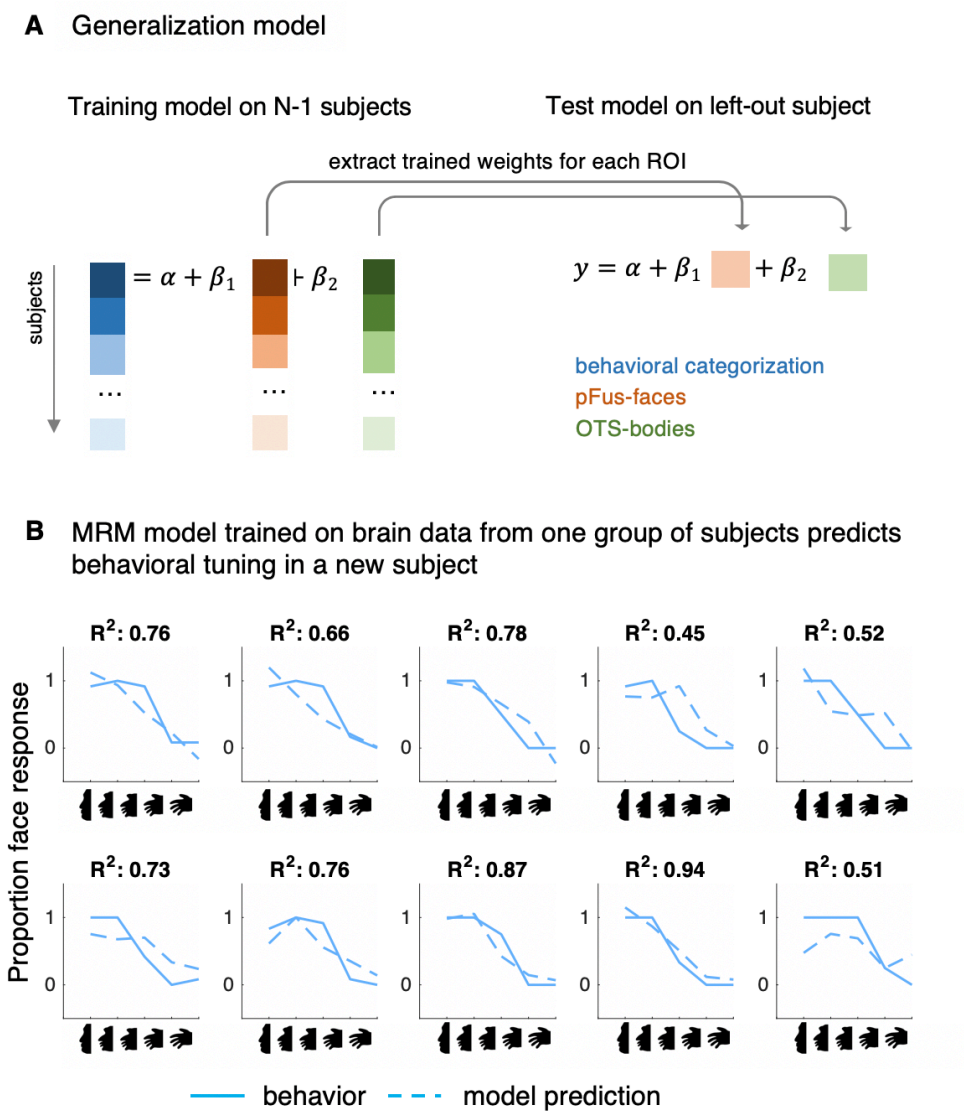

Figure 6. A cross-validated, multi-ROI model that is trained on brain data from one group of subjects predicts behavioral tuning in a new subject. A. Schematic illustration of the cross-validation procedure. *Left:* N-1 subjects were used to train the model using the neural tuning from pFus-faces and OTS-bodies in the right hemisphere (see text for details). *Right:* The resulting beta coefficients were then multiplied with the fMRI data of the left-out-subject and used to predict their behavioral tuning for each morph level. **B.** In separate subplots, behavioral tuning (solid) for a participant and the model's prediction (dashed) are shown. *X-axis:* morph level; *Y-axis:* proportion face response. *Solid blue line:* Behavioral tuning from the left-out subject. *Dashed blue line:* Model. Explained variance for each participant is indicated at the top of each subplot.

DISCUSSION

In the present study, we measured neural tuning from face- and body-selective regions in human ventral temporal cortex (VTC) while participants viewed carefully controlled stimuli that were morphed in a continuous shape space in which one side was a face silhouette and the other, a hand. The center shape of the continuum was perceived as being approximately equal parts of both categories. We then conducted a behavioral experiment in the same participants outside the scanner to examine the relationship between neural and behavioral tuning. Our results revealed the surprising finding that neural tuning from face- and body-selective regions together better predicted behavioral tuning than any one region by itself. Furthermore, despite individual differences among participants, the relationship between neural tuning and behavioral tuning was consistent such that a model trained on neural tuning from one group of participants accurately predicted a new participant's behavioral tuning, but only if the neural tuning was from the right hemisphere.

We consider these findings in the context of recent theories proposing that large-scale cortical networks are composed of smaller, recurrent networks. As face- and body-selective regions are adjacent to one another in human VTC, we first consider computational principles of cortical adjacency that could explain why there is a stronger relationship between neural tuning from both regions and behavior compared to neural tuning from just one region. Specifically, we propose that the activity of smaller units (e.g. face- or body-selective regions by themselves) of these recurrent networks may form a larger unit (e.g. face- and body-selective regions together) that are hemisphere-specific under certain task demands. Finally, we discuss how the model-based approach implemented here is generalizable to future studies examining the relationship between neural and behavioral categorization both within and outside visual domains.

Our data show that opposed neural tuning from both a face- and body-selective region together more accurately explains behavioral categorization of ambiguous stimuli than neural tuning from (a) one face- or body-selective region alone or (b) two VTC regions with either no tuning or the same tuning. Indeed, two features of our results indicate that differential neural tuning between regions is necessary to accurately predict categorization behavior. First, the combination of neural tuning from two regions (CoS-places and VO-2; **Supplementary Fig. 4**) that are not tuned to either category used in the present experiment (faces and hands) could not predict behavioral tuning. Second, a stepwise linear regression model revealed that two regions with similar tuning (e.g. mFus-faces and pFus-faces) did not add a significant amount of explained

variance to explain categorization behavior compared to the neural tuning from either region by itself. Instead, the model revealed that the most variance was explained from two regions with differential neural tuning (pFus-faces and OTS-bodies; **Fig 4D,E**).

Interestingly, we also found that certain combinations of face- and body-selective regions more accurately predicted behavior than other combinations. Specifically, the combination of neural tuning from pFus-faces and OTS-bodies better predicted behavioral tuning compared to the combination of neural tuning from mFus-faces and OTS-bodies – especially in the right hemisphere. We hypothesize that a) combinatorial tuning between pFus-faces and OTS-bodies will not always out-perform combinatorial tuning between mFus-faces and OTS-bodies and b) combinations of neural tuning in the right hemisphere will not always be more closely associated with behavioral categorization than combinations in the left hemisphere. Specifically, we propose that the "best" combination of neural tuning that predicts behavioral categorization will be dependent on the stimulus and task. For example, mFus-faces is situated at a higher position in the visual hierarchy compared to pFus-faces[14,40]. As such, we hypothesize that the present stimuli, which were simple, two-dimensional shapes, were more ideal for pFus-faces. Future research can assess whether conducting an experiment with more complex stimuli would result in the combination of neural tuning from mFus-faces and OTS-bodies out-performing the combination of pFus-faces and OTS-bodies. Additionally, future research can examine how combinations of neural tuning affect the lateralization effects identified here. For example, recent proposals hypothesize that the lateralization for word processing in the left hemisphere and face processing in the right hemisphere develops from the competition of foveal resources during development in combination with the lateralization of the language system in the left hemisphere[41,42]. As face- and word-selective regions form a cortical cluster in left posterior VTC, while face-, body-, and word-selective regions form a cluster in left mid-VTC, future studies examining the relationship between neural tuning and behavioral tuning during categorization tasks that involve faces, bodies, and words may find that combinatorial tuning from all three regions in the left, but not the right, hemisphere would out-perform other models. As such, despite differences in selectivity, the cortical adjacency of functional clusters in VTC would enable flexible combinations of neural tuning that can accommodate different types of stimuli and task demands.

From a network perspective, the finding that neural tuning from both face- and body-selective regions best predicts behavioral categorization supports a new theoretical idea in which

cortically adjacent regions selective for different domains work together to achieve perceptual categorization. We consider these findings in the context of recent[7,43] and classic[44] theories proposing that large-scale cortical networks are composed of smaller, recurrent networks that traverse nested spatial scales. Specifically, we propose that the activity of smaller units (e.g. face- or body-selective regions by themselves) of these recurrent networks may form a larger unit (e.g. face- and body-selective regions together) under certain task conditions. One possibility is that joint processing by cortically adjacent regions is a general functional feature of the brain, as it would allow cortically proximal regions to form larger processing units depending on the task and the stimulus, therewith increasing the flexibility of brain computations. Additionally - and as mentioned in the previous paragraph – the number of regions, as well as the hemisphere in which those regions are located that are most predictive of behavior, is likely dependent on the stimuli used and the task being implemented. Future studies can examine this possibility by a) changing tasks in the scanner and b) acquiring trial by trial behavioral categorization judgements in the scanner, which was an approach not implemented in the present study. Such findings may further support the flexibility of cortical adjacency that we are proposing – neural tuning can be read-out from multiple combinations of cortically adjacent regions depending on the complexity of the stimuli and task demands.

Consistent with our results and the interpretation of our findings, recent anatomical and functional connectivity studies in macaques also propose the integration of neural signals between domain-specific networks selective for faces and bodies. Indeed, while recent research in macaques identifies distinct cortical networks that separately process face and body information[45,46], these studies also identify connections that are shared between these two networks[46,47]. These studies propose that the subset of these networks that overlap with one another may be integral for the integration of face and body information both anatomically[46] and functionally[47]. Here, we show that cortical adjacency may be enough to enable the integration of signals between regions – perhaps through max pooling operations resulting from short-range, local connectivity between cortically adjacent patches that are selective for different domains. While we were unable to test this hypothesis at the neuronal level in the present study, this can be achieved in future studies as face- and body-selective regions are cortically adjacent in other species[47–49].

In the present study, we have operationalized visual categorization as the discrimination between two animate categories (faces and hands) and have focused on regions within human VTC[7] not only to investigate the relationship between neural and behavioral categorization, but also to build a generalizable model explaining this brain-behavior relationship in human participants. Of course, there are (a) several ways to operationalize visual categorization, (b) many cortical regions to explore, (c) many methodologies to use, and (d) other modalities to measure. The model-based approach we have implemented in the present study is applicable to these additional factors. For example, previous empirical studies have operationalized visual categorization in the context of other stimuli such as shapes and objects[50] and additional stimulus features such as color and motion[51]. Neurally, previous studies have also examined visual categorization in brain areas outside of VTC[50–52], as well as have considered how other brain areas (for example, within prefrontal cortex[53]) may act as gatekeepers that guide the categorical decision process within cortical regions in VTC. Of course, categorization is also not limited to the visual domain. For instance, empirical studies examine the categorization of (a) movements in motor cortex[54], (b) odors in piriform cortex[55], and (c) sounds in the visual cortex of blind individuals[56]. The approach we have implemented here can be applied to these examples, as long as two requirements are fulfilled. First, a boundary between (at least) two categories must be carefully controlled for in the stimulus domain. Second, there must be a known or hypothesized relationship between behavioral tuning and neural tuning from the functional regions of interest. As long as these two requirements are met, our model-based approach can be applied to additional types of categorization, as well as additional functional regions in future studies. Further, we also emphasize that while we establish this approach with neural tuning that is read-out from entire functional regions, the spatial scale of this approach is presently unknown. For instance, distinct cortical columns in different parts of the brain represent (1) different viewpoints of faces[57], (2) distinct motion directions[58], or (3) specific parts of the body surface[44]. Future studies can test if the combination of neural tuning signals from multiple cortical columns better predicts behavioral categorization compared to neural tuning signals from single cortical columns.

In summary, we find that neural categorization tuning from multiple domain-specific regions in human ventral temporal cortex accurately models behavioral categorization tuning. Furthermore, despite individual differences in neural and behavioral tuning, this model trained on the neural tuning from one group of participants nicely aligns with behavioral tuning in new

participants with a right hemisphere dominance. Together, these data suggest that adjacent cortical regions with opposed neural tuning may serve as recurrent networks that can accommodate efficient and flexible categorization within and outside visual domains, and likely for additional cognitive and perceptual tasks.

METHODS

I. Generating and calibrating carefully-controlled face-hand morphs

We generated a novel set of parameterized silhouette stimuli that spanned a continuous morph space between faces and hands while controlling for many low-level image properties, including brightness, contrast, and total silhouette area. We generated 60 silhouette stimuli at each of 5 morph levels (ranging from fully face-like (level-1), ambiguously face- or hand-like (level-3), to fully hand-like (level-5)), which resulted in 300 stimuli. These stimuli were created by parameterizing the contours of a large set of photographs of profile faces and open hands using the same number of key points for images from both categories (**Fig. 1A**). Profile face photographs were obtained from the FERET database[59,60] and were parameterized as outlined in Davidenko (2007), except that 36 rather than 18 key points were used to define each face contour. The increased number of key points compared to our prior work enabled a one-to-one mapping between the profile face photographs with the more detailed hand contours. Open hand photographs were obtained from 12 volunteers who each provided between 10 and 20 open hand poses, for a total of 148 unique hand images. As with the face images, 36 key points were defined along the contours of each hand, such that points of maximum and minimum curvature matched between faces and hands (e.g. point 7 corresponds to the tip of the brow in faces and the tip of the thumb in hands; see **Fig. 1A**).

Stimulus Calibration

The silhouette stimuli were calibrated in an iterative process by (1) obtaining perceptual ratings from Mechanical Turk (mTurk) workers (https://www.mturk.com/), (2) adjusting the

parameterized stimuli, and (3) repeating the process. This calibration procedure had two goals: first, to obtain intermediate (level-3) stimuli that appeared equally face- and hand-like, and second, to generate multiple exemplars at each morph level that were matched for perceptual variability across the five morph levels. To accomplish the first goal, mTurk workers rated between 1 and 10 stimuli randomly selected from among the 300 exemplars (60 at each morph level) on either how face-like (1 to 5 scale) or how hand-like (1 to 5 scale) they appeared, until each exemplar had been rated by at least 5 workers. Following this first round, intermediate stimuli were found to be too hand-like, so these were adjusted in the face-hand morphing trajectory toward faces. The second round produced improved results, and the third round produced stimuli that were well-balanced across the five morph levels (**see Fig. 1B**). To control for perceptual variability, a similar iteration of mTurk studies was conducted in which workers rated between 1 and 10 pairs of stimuli from within each of the 5 morph levels, according to a 7-point dissimilarity metric (1=identical and 7=maximally dissimilar). After the first round, level-1 (face) stimuli were found to be more perceptually variable than the other 4 levels, and so their variability was reduced for the second round by morphing each face stimulus 40% toward the average face. The second round of mTurk ratings produced well-matched perceptual variability across the five morph levels (**Fig. 1C**).

II. Testing categorization of face-hand morphs within and outside the scanner:

Participants

14 participants (8 females, ages 22 - 44) were scanned at the Center for Neurobiological Imaging (CNI) at Stanford University using a 3 T GE Discovery MR750 scanner. Each subject underwent (1) an anatomical scan, (2) a functional localizer experiment, (3) a face/hand morphing experiment within the scanner (**Fig. 1D**), and afterwards, (4) a behavioral experiment outside the scanner that measured their categorization of face/hand morphs as either faces or hands (**Fig. 1E**). All participants gave their written informed consent. Procedures were approved by the Stanford Internal Review Board on human participant research.

1. Anatomical imaging and cortical surface reconstruction

Data acquisition. A standard high-resolution anatomical volume of the whole brain was acquired using a T1-weighted BRAVO pulse sequence (TR=450 ms, flip angle=12°, 1 NEX, FOV=240 mm, resolution: 1.0mm isotropic).

Cortical surface reconstruction. All anatomical volumes were aligned to the AC-PC plane. Using automated (FreeSurfer: http://surfer.nmr.mgh.harvard.edu) and manual segmentation tools (ITK-SNAP: http://www.itksnap.org/pmwiki/pmwiki.php), each anatomical volume was segmented to separate gray from white matter, and the resulting boundary was used to reconstruct the cortical surface for each subject[61].

2. Functional Imaging

Data acquisition. Using a custom-built phase-array, 32-channel head coil, we acquired 34 oblique coronal slices covering occipito-temporal cortex (resolution: 2.4 mm x 2.4 mm x 2.4 mm; one-shot T2*-sensitive gradient echo acquisition sequence: FOV = 192 mm, TE = 30 ms, TR = 2000 ms, flip angle = 77°).

Localizer Experiment. All subjects participated in 2 runs of an fMRI functional localizer experiment (150 volumes each). Each run consisted of 12 s blocks, which contained different images from the same category presented at a rate of 1 Hz. Images within each block were from 1 of 7 different categories: faces (frontal / profile view), body parts (hands / lower limbs), objects (cars / tools), and scrambled versions of these images. Across the two experimental runs, there were 8 repetitions of each block type. Participants maintained fixation and responded by button press when an image repeated (1-back task).

Morphing experiment. Using the carefully calibrated face and hand morph stimuli, all subjects participated in 4 runs of a block design experiment during which subjects viewed silhouettes alternating with a gray screen. A fixation cross was on the screen at all times. (**Fig. 1D**). Each run consisted of 10, 12-second blocks and each consisted of 12 silhouettes displayed at a frequency of 1 Hertz. Blocks of silhouettes (a) always contained images from the same morphing level and (b) appeared in a pseudo-random order, in which there was a difference of two morphing levels (on average) between consecutive blocks (e.g. level 3 followed by level 1, or level 2 followed by level 4). In order to assure that any measurable effects in high-level visual regions were not driven by

consistent placement of silhouette edges in a particular location in the visual field[8], (1) images were placed centrally in one of 8 randomly-chosen locations around the fixation point (max visual degree: 3), and (2) silhouettes faced either left or right. Participants maintained fixation and responded by button press when an image repeated (1-back task), which happened 0-2 times per stimulus block. To keep task and cognitive processes as similar as possible between stimulus and rest blocks, fixation crosses also changed with a 1 Hertz frequency displaying different lengths of lines, and participants performed the 1-back task when the same fixation cross appeared twice in a row.

3. Behavioral testing of face/hand morph categorization

Each of the 14 subjects also participated in a behavioral testing experiment outside the scanner after the fMRI experiment. The experiment was self-paced, as well as consisted of 60 trials (12 images from each of the 5 morph levels, presented in random order). In each trial, a silhouette appeared at the center of the screen (**Fig. 1D**). Subjects then performed a forced-choice task, deciding whether the silhouette was a face (rating: 1) or a hand (rating: 0). The stimulus remained on the screen until the subject chose one of the two options. For each of the subjects, we recorded the mean response for each morphing level, resulting in 5 values between 0 (all hand) and 1 (all face) for each subject.

III. Data analyses

fMRI data were analyzed using mrVista (http://github.com/vistalab) and custom software written in and implemented by MATLAB (www.mathworks.com).

1. Timeseries preprocessing

Functional data were motion corrected using an affine transformation[62]. Time series data were processed with a temporal high-pass filter (1/20 Hz cutoff) and then converted to percentage signal change by dividing the time series of each voxel by its mean intensity. We estimated the blood oxygen level dependent (BOLD) response amplitudes for each condition using a general linear model (GLM) applied to the time series of each voxel. Predictors of the GLM were the experimental conditions convolved with the hemodynamic impulse response function used in SPM (http://www.fil.ion.ucl.ac.uk/spm/). Data were not spatially smoothed.

2. Functional regions of interest (ROIs)

Our analyses focused on functional regions of interest (ROIs) that are selective for faces and limbs within ventral temporal cortex (VTC) as VTC is critical for categorization[7] and participants performed a categorization task during the behavioral experiment.

Face-selective ROIs: Face-selective voxels were identified as those voxels in VTC that illustrated higher BOLD responses to grayscale photographs of faces (frontal or profile view) compared to body parts (hands and lower limbs) and objects (cars and tools) in the functional localizer experiment (**Fig. 2A**). As in our prior publications, two ROIs were defined within each subject's native brain anatomy using a common threshold ($t > 3$, voxel-level) and were positioned laterally to the mid-fusiform sulcus (MFS; Weiner et al., 2014). mFus-faces (N=14) is adjacent to the anterior tip of the MFS, while pFus-faces (N=12) is located 1-1.5cm more posteriorly on the lateral fusiform gyrus extending into the posterior occipitotemporal sulcus (OTS). In some studies mFus-faces is also referred to as FFA-2, while pFus-faces is also referred to as FFA-1, in which the FFA stands for Fusiform Face Area[34].

Body-selective ROI: Body-selective voxels were identified as those voxels in VTC that illustrated higher BOLD responses to grayscale photographs of limbs (hands and lower limbs) compared to faces (frontal or profile view) and objects (cars and tools) in the functional localizer experiment. As in our prior publications, one ROI (N=12) was defined within each subject's native brain anatomy using a common threshold ($t > 3$, voxel-level) and was located within the OTS, positioned between mFus-faces and pFus-faces. We refer to this region as OTS-bodies, but it is also referred to as the Fusiform Body Area[63,64].

3. Mean tuning analyses

Neural tuning: Within each individual subject, a GLM was conducted on the mean fMRI timeseries of each ROI (mFus-faces, pFus-faces, OTS-bodies), separately for the left and right hemisphere. Response amplitudes (betas) and residual variance of each ROI were estimated from the GLM, which yielded 5 beta values per ROI (one for each morphing level; **Fig. 2**). Beta values were z-scored (mean: 0, SD: 1) to allow for comparison across subjects and ROIs. For each ROI,

main effects of morph level were separately evaluated using a 1-way repeated measures analysis of variance (ANOVA) with morph level as a factor. The interaction between morph level, ROI, and hemisphere was also evaluated using a repeated measures ANOVA using morph level, ROI, and hemisphere as factors. We refer to the relationship among these five beta values as *neural tuning* to these face-hand stimuli.

Behavioral Tuning: For each subject, we measured the average categorization responses of each morph level across the 12 trials of that level. This behavioral response indicates the proportion of face responses for each morphing level. We refer to this function as *behavioral tuning* to face-hand stimuli. Like neural responses, we visualize behavioral tuning by fitting the categorization responses to the five morphing levels with a line (**Fig. 3**).

4. Relating neural tuning to behavioral tuning:

In order to quantify the relationship between neural tuning and behavioral tuning, we used a linear regression model that relates each subject's behavioral tuning to their neural tuning. Each ROI's tuning (z-scored beta values) to the five morph levels was used as a predictor for a linear regression model (the independent variable) relating it to the behavioral data (the dependent variable, **Fig. 4A**). This analysis was performed on a subject-by-subject basis and separately for neural data from each hemisphere. Goodness-of-fit of the model was evaluated by computing the explained variance (R^2). R^2 was calculated as the square of the correlation between the model's estimation and the behavioral data. We used paired permutation testing (10,000 iterations) to assess whether the explained variances differed significantly across the VTC ROIs. Permutation tests were used as we did not want to make any assumptions about the distribution of the R^2 values. Results were corrected for multiple comparison using the adjusted Bonferroni method (dividing the p-value by the number of comparisons minus one). Permutation testing was performed using participants in which all ROIs were identified.

Linear regression model for individual ROI data (1-ROI model): We implemented 6, 1-RO1 linear regression models predicting the behavioral tuning curve from each of the ROIs (mFus-faces, pFus-faces, OTS-bodies; separately for LH and RH, **Fig. 4B**).

Linear regression model using data from multiple ROIs (Multi-ROI model, MRM): We ran a stepwise linear regression model using all subjects and the three ROIs (mFus-faces, pFus-faces, OTS-bodies; separately for each hemisphere) to establish which of the ROIs, or which combination of ROIs, best explained behavioral tuning. After identifying which regions were significant predictors, we applied a linear regression model that used data from pFus-faces and OTS-bodies as separate predictors for behavioral tuning as these were the ROIs that were identified as significant predictors during the stepwise linear regression (**Fig. 4B**).

To test the improvement by the MRM compared to the 1-ROI model, we used the model selection criterion AIC (Akaike's Information Criterion) to compare model performances, in which lower values indicate a better model fit. In addition, we used paired permutation testing (10,000 iterations) to examine whether the Multi-ROI model outperformed the 1-ROI models. Finally, we compared if there were any hemispheric differences in model performance using paired permutation testing (using multiple comparison correction as described above).

Linear regression Multi-ROI control model: To test if any 2-ROI model can predict behavioral tuning to face/hand morphs, we ran a control model using 2 ROIs that are cortically proximal to face- and limb-selective regions, but are not selective for either faces or limbs. These ROIs were a retinotopically defined visual area (VO-2 from Wang et al., 2014) and a place-selective region (CoS-places from Weiner et al., 2018). Analyses were performed in the same fashion as for the Multi-ROI model described above.

5. Cross-Validated MRM model across participants

We tested if a model built from the neural responses in one group of participants could predict behavioral tuning in a new, independent participant. For these analyses, we used neural tuning from pFus-faces and OTS-bodies because our previous linear regression analyses indicated that neural tuning from these regions were the best predictors of behavioral tuning.

We first estimated regression weights using a linear regression analyses based on data from N-1 subjects and then predicted responses of the left-out subjects. This analysis was done using 10 subjects in which we could define both of the ROIs. To perform the analysis, we concatenated the z-scored neural responses to the 5 morph levels in each ROI across subjects, which resulted in a

1x45 (e.g. 9 subjects x 5 morph levels) vector per ROI. Likewise for behavioral responses we concatenated the behavioral responses to the 5 morph levels across subjects (1x45 vector). The two neural vectors served as the predictors for the behavioral vector. The regression analysis derived the weights (model coefficients) for each of the two neural predictors. To predict responses of the left out subjects, we multiplied the model coefficients by the measured neural responses of the left-out-subject to generate the predicted behavioral tuning for the left-out-subject. We then calculated the goodness-of-fit between the predicted and actual behavioral tuning of the left-out-subject using R^2 as described previously. This procedure was cross-validated using all 10 subjects, resulting in a 10-fold cross-validation. This cross-validated Multi-ROI model was compared to a cross-validated Control Model, which used neural data from VO-2 and CoS-places as predictors. The models were compared using paired permutation testing as well as multiple comparison correction as described in previous sections.

Code availability: All code relevant to data analysis for the main findings (Figs. 1–6) is available on github.com/VPNL/faceHands/. The majority of the code used in this study was based on scripts and functions available with the open-source vistasoft code library: https://github.com/vistalab/vistasoft

Data availability: Any source data relevant for the main findings (Figs. 1-6) is available on github.com/VPNL/faceHands/ in the form of MATLAB files containing voxel data.

Acknowledgements: We thank Kendrick Kay for useful discussions regarding statistical analyses. This research was supported by UC Berkeley start-up funds (to KSW) and NEI grant: 1R01EY02391501A1 (to KGS).

Author contributions: KSW, ND, and KGS designed research. KSW and ND performed research; MR, KSW, and ND analyzed the data; MR, KGS, and KSW wrote the manuscript.

References

1. Thorpe, S., Fize, D. & Marlot, C. Speed of processing in the human visual system. *Nature* **381,** 520–522 (1996).
2. Grill-Spector, K. & Kanwisher, N. Visual Recognition. *Psychol. Sci.* **16,** 152–160 (2005).
3. Moutoussis, K. & Zeki, S. The relationship between cortical activation and perception investigated with invisible stimuli. *Proc. Natl. Acad. Sci.* **99,** 9527–9532 (2002).
4. Andrews, T. J., Schluppeck, D., Homfray, D., Matthews, P. & Blakemore, C. Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion. *Neuroimage* **17,** 890–901 (2002).
5. Hasson, U., Hendler, T., Bashat, D. Ben & Malach, R. Vase or Face? A Neural Correlate of Shape-Selective Grouping Processes in the Human Brain. *J. Cogn. Neurosci.* **13,** 744–753 (2001).
6. Haxby, J. V., Hoffman, E. A. & Gobbini, M. I. The distributed human neural system for face perception. *Trends Cogn. Sci.* **4,** 223–233 (2000).
7. Grill-Spector, K. & Weiner, K. S. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* **15,** 536–548 (2014).
8. Davidenko, N., Remus, D. A. & Grill-Spector, K. Face-likeness and image variability drive responses in human face-selective ventral regions. *Hum. Brain Mapp.* **33,** 2334–2349 (2012).
9. Bracci, S. & Op de Beeck, H. Dissociations and Associations between Shape and Category Representations in the Two Visual Pathways. *J. Neurosci.* **36,** 432–444 (2016).
10. Proklova, D., Kaiser, D. & Peelen, M. V. Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate–Inanimate Distinction. *J. Cogn. Neurosci.* **28,** 680–692 (2016).
11. Kanwisher, N. Functional specificity in the human brain: A window into the functional architecture of the mind. *Proc. Natl. Acad. Sci.* **107,** 11163–11170 (2010).
12. Peelen, M. V & Downing, P. E. The neural basis of visual body perception. **8,** 636–648 (2007).
13. Op de Beeck, H. P. *et al.* Fine-scale spatial organization of face and object selectivity in the temporal lobe: do functional magnetic resonance imaging, optical imaging, and electrophysiology agree? *J. Neurosci.* **28,** 11796–11801 (2008).
14. Grill-Spector, K., Weiner, K. S., Gomez, J., Stigliani, A. & Natu, V. S. The functional

neuroanatomy of face perception: From brain measurements to deep neural networks. *Interface Focus* **8,** (2018).

15. Wandell, B. A., Rauschecker, A. M. & Yeatman, J. D. Learning to See Words. *Annu. Rev. Psychol.* 31–53 (2012). doi:10.1146/annurev-psych-120710-100434.Learning
16. Tsao, D. Y. & Livingstone, M. S. The Interface Theory of Perception. *Ann Rev Neurosci* **24,** 1–1 (2008).
17. Parvizi, J. *et al.* Electrical Stimulation of Human Fusiform Face-Selective Regions Distorts Face Perception. *J. Neurosci.* **32,** 14915–14920 (2012).
18. Jonas, J. *et al.* Focal electrical intracerebral stimulation of a face-sensitive area causes transient prosopagnosia. *Neuroscience* **222,** 281–288 (2012).
19. Moeller, S., Crapse, T., Chang, L. & Y Tsao, D. *The effect of face patch microstimulation on perception of faces and objects*. *Nature Neuroscience* **20,** (2017).
20. Downing, P. E. & Peelen, M. V. Body selectivity in occipitotemporal cortex: Causal evidence. *Neuropsychologia* **83,** 138–148 (2016).
21. Haxby, J. V. Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science (80-.).* **293,** 2425–2430 (2001).
22. Martin, A. *The Representation of Object Concepts in the Brain*. *Annual review of psychology* **58,** (2007).
23. Mahon, B. Z. & Caramazza, A. What drives the organization of object knowledge in the brain? *Trends Cogn. Sci.* **15,** 97–103 (2011).
24. Gauthier, I. & Tarr, M. J. Visual Object Recognition: Do We (Finally) Know More Now Than We Did? *Annu. Rev. Vis. Sci.* **2,** 377–396 (2016).
25. Grill-Spector, K., Knouf, N. & Kanwisher, N. The fusiform face area subserves face perception, not generic within-category identification. *Nat. Neurosci.* **7,** 555–562 (2004).
26. Serre, T., Olivia, A. & Poggio, T. A feedforward architecture accounts for rapid categorization. *PNAS* **104,** 6424–6429 (2007).
27. Yamins, D. L. *et al.* Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci U S A* **111,** 8619–8624 (2014).
28. Bernstein, M., Oron, J., Sadeh, B. & Yovel, G. An Integrated Face–Body Representation in the Fusiform Gyrus but Not the Lateral Occipital Cortex. *J. Cogn. Neurosci.* **26,** 2469–2478 (2014).

29. Kaiser, D., Strnad, L., Seidl, K. N., Kastner, S. & Peelen, M. V. Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns. *J. Neurophysiol.* **111,** 82–90 (2014).
30. Davidenko, N. Silhouetted face profiles: A new methodology for face perception research. *J. Vis.* **7,** 1–17 (2007).
31. Peelen, M. V & Downing, P. E. Selectivity for the human body in the fusiform gyrus. *J. Neurophysiol.* **93,** 603–8 (2005).
32. Schwarzlose, R. F., Baker, C. I. & Kanwisher, N. Separate Face and Body Selectivity on the Fusiform Gyrus. *J. Neurosci.* **25,** 11055–11059 (2005).
33. Weiner, K. S. & Grill-Spector, K. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage* **52,** 1559–1573 (2010).
34. Kanwisher, N. G., McDermott, J. & Chun, M. M. The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *J. Neurosci.* **17,** 4302–4311 (1997).
35. Weiner, K. S. *et al.* The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *Neuroimage* **84,** 453–465 (2014).
36. Fan, L. *et al.* Connectivity-based parcellation of the human temporal pole using diffusion tensor imaging. *Cereb. Cortex* **24,** 3365–3378 (2014).
37. Aguirre, G. K., Zarahn, E. & D'Esposito, M. An Area within Human Ventral Cortex Sensitive to "Building" Stimuli. *Neuron* **21,** 373–383 (1998).
38. Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. *Nature* **392,** 598–601 (1998).
39. Weiner, K. S. *et al.* Defining the most probable location of the parahippocampal place area using cortex-based alignment and cross-validation. *Neuroimage* **170,** 373–384 (2018).
40. Grill-spector, K., Kay, K. & Weiner, K. S. The Functional Neuroanatomy of Face Processing: Insights from Neuroimaging and Implications for Deep Learning. 3–31 (2017). doi:10.1007/978-3-319-61657-5
41. Behrmann, M. & Plaut, D. C. A vision of graded hemispheric specialization. *Ann. N. Y.*

Acad. Sci. **1359,** 30–46 (2015).

42. Dehaene, S., Cohen, L., Morais, J. & Kolinsky, R. Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nat. Rev. Neurosci.* **16,** 234–244 (2015).

43. Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G. & Mishkin, M. The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* **17,** 26–49 (2013).

44. Mountcastle, V. The columnar organization of the neocortex. *Brain* (1997).

45. Moeller, S., Freiwald, W. A. & Tsao, D. Y. Patches with Links : A Unified System for Processing Faces in the Macaque Temporal Lobe. *Science (80-.).* **320,** 1355–1360 (2008).

46. Borra, E. *et al.* Cortical connections of the macaque anterior intraparietal (AIP) area. *Cereb. Cortex* **18,** 1094–1111 (2008).

47. Premereur, E., Taubert, J., Janssen, P., Vogels, R. & Vanduffel, W. Effective Connectivity Reveals Largely Independent Parallel Networks of Face and Body Patches. *Curr. Biol.* **26,** 3269–3279 (2016).

48. Hung, C.-C. *et al.* Functional Mapping of Face-Selective Regions in the Extrastriate Visual Cortex of the Marmoset. *J. Neurosci.* **35,** 1160–1172 (2015).

49. Pinsk, M. A. *et al.* Neural Representations of Faces and Body Parts in Macaque and Human Cortex: A Comparative fMRI Study. *J. Neurophysiol.* **101,** 2581–2600 (2009).

50. Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291,** 312–6 (2001).

51. Ferrera, V. P., Yanike, M. & Cassanello, C. Frontal eye field neurons signal changes in decision criteria. *Nat. Neurosci.* **12,** 1458–1462 (2009).

52. Jiang, X. *et al.* Categorization Training Results in Shape- and Category-Selective Human Neural Plasticity. *Neuron* **53,** 891–903 (2007).

53. Baldauf, D. & Desimone, R. Neural Mechanisms of Object-Based Attention. *Science (80-.).* **1268,** 424–428 (2014).

54. Seger, C. A. & Miller, E. K. Category Learning in the Brain. *Annu. Rev. Neurosci.* **33,** 203–219 (2010).

55. Tantirigama, M. L. S., Huang, H. H.-Y. & Bekkers, J. M. Spontaneous activity in the piriform cortex extends the dynamic range of cortical odor coding. *Proc. Natl. Acad. Sci.* **114,** 2407–2412 (2017).

56. van den Hurk, J., Van Baelen, M. & Op de Beeck, H. P. Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proc. Natl. Acad. Sci.* **114,** E4501–E4510 (2017).

57. Tanaka, K. Inferotemporal Cortex and Object Vision. *Annu. Rev. Neurosci.* **19,** 109–139 (1996).

58. Albright, T. D. Direction and orientation selectivity of neurons in visual area MT of the macaque Direction and Orientation Selectivity of Neurons in Visual Area MT of the Macaque. *J. Neurophysiol.* **52,** 1106–1130 (1984).

59. Phillips, P. J., Wechsler, H., Huang, J. & Rauss, P. J. The FERET database and evaluation procedure for face-recognition algorithms. *Image Vis. Comput.* **16,** 295–306 (1998).

60. Phillips, P. J., Moon, H., Rauss, P. & Rizvi, S. The {FERET} evaluation methodology for face recognition algorithms. *IEEE Proc. Comput. Vis. Pattern Recognit.* **22,** 137–143 (1999).

61. Wandell, B. A., Chial, S. & Backus, B. T. Visualization and measurement of the cortical surface. *J. Cogn. Neurosci.* **12,** 739–752 (2000).

62. Nestares, O. & Heeger, D. J. Robust multiresolution alignment of MRI brain volumes. *Magn. Reson. Med.* **43,** 705–715 (2000).

63. Peelen, M. V & Downing, P. E. Using multi-voxel pattern analysis of fMRI data to interpret overlapping functional activations. **11,** 4–5 (2006).

64. Schwarzlose, R. F. Separate Face and Body Selectivity on the Fusiform Gyrus. *J. Neurosci.* **25,** 11055–11059 (2005).

65. Wang, L., Mruczek, R. E. B., Arcaro, M. J. & Kastner, S. Probabilistic Maps of Visual Topography in Human Cortex. *Cereb. Cortex* 1–21 (2014).

1