Combined neural tuning in human ventral temporal cortex resolves perceptual ambiguity

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

[1] Psychology Department, Stanford University;
[2] Psychology Department, University of California, Santa Cruz
[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. Nevertheless, how cortical regions in human ventral temporal cortex (VTC), which is critical for categorization, support this behavioral ability, is presently unknown. Here, we examined the relationship between neural responses and behavioral performance during the categorization of morphed silhouettes of faces and hands, which are animate categories processed in cortically adjacent regions in VTC. Our results reveal that the combination of neural responses from VTC face- and body-selective regions more accurately explains behavioral categorization than neural responses from either region alone. Furthermore, we built a model that predicts a person's behavioral performance using estimated parameters of brain-behavioral relationships from a different group of people. Thus, while face- and body-selective regions are located within functionally-distinct domain-specific networks, cortically adjacent regions from both networks likely integrate neural responses to resolve competing and perceptually ambiguous information from both categories.

INTRODUCTION

We categorize objects within our visual world countless times each day and within a split second[1,2]. Despite this frequent occurrence and undisputable relevance to everyday life, it is presently unknown how regions and networks within the human brain support efficient and accurate visual categorization, especially for ambiguous stimuli[3–5]. Prior research has shown that human ventral temporal cortex (VTC) is critical for visual categorization[6,7] and that shape is a key feature for neural categorization[8–10]. In addition, there are several segregated functional networks in human VTC that are specialized for processing different types of ecologically-relevant categories – such as faces and bodies[6,11–16]. Indeed, previous findings in both humans and non-human primates have revealed causal relationships between neural responses in face-selective regions and the perception of faces[17–19], as well as between neural responses in body-selective regions and the perception of bodies[20]. While these series of studies examined the relationship between neural responses and perception separately for each domain, it is presently unknown whether or how brain regions from these distinct cortical networks 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[3–5,21] or (b) neural responses are computationally transformed from regions that are positioned in early cortical stages of a network compared to later stages[22,23]. Building on this empirical foundation, a recent trend in neuroimaging studies has begun to unveil how information regarding different categories is combined. For example, several studies have shown that separate information about faces and bodies can be integrated within a functional region in VTC to generate new information that represents a whole person[24,25]. While this approach sheds light on how information from two separate stimuli and two different categories may be integrated in one functional region, it still remains untested (a) how regions selective for different categories in VTC process stimuli that consist of visual features from two different categories, i.e. ambiguous stimuli (**Fig. 1A**), and (b) if the combination of neural signals between regions explains behavioral categorization better than neural signals from either region by itself.

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 such a shape space containing a morphed continuum between faces and hands using the silhouette methodology[8,26] and tested if behavioral categorization of these stimuli is best predicted by either a) separate or b) combined neural responses from face- or body-selective regions in VTC. We reasoned that if the former model best predicts behavior it would support the idea that behavioral categorization of ambiguous stimuli is based on responses from a single region similar to prior results using natural, unambiguous stimuli [3,21,27], whereas if the latter model best predicts behavior, it would suggest that behavioral categorization depends on relative neural responses across multiple regions despite the fact that these neural responses are from functionally distinct domain-specific networks.

RESULTS

We first generated 60 face-hand morphs using the silhouette methodology[26]. Compared to natural images, the silhouette methodology holds the advantage of generating a stimulus space that is a) controlled for low level visual features, b) changes only one visual attribute of the stimulus (its shape), and c) drives neural responses in category-selective regions in VTC[8]. This approach generated a large set of carefully controlled images consisting of continuous morphs between faces and hands (**Fig. 1A**; Materials and Methods).

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**). The latter is important as the lack of variability of stimuli within a morph level may generate fMRI-adaptation[8,28]. Thus, our calibration ensured that the variability of stimuli is matched across morph levels.

Using these stimuli, we conducted two types of experiments in 14 independent adults (8 females, ages 22 - 44): (1) A functional magnetic resonance imaging (fMRI) block-design experiment in which we measured mean neural responses to face-hand morphs along the morphing continuum in face- and body-selective regions in VTC (**Fig. 1D**). We refer to neural responses across morph levels as neural tuning. (2) A behavioral categorization experiment outside the scanner during which we measured the proportion of face-hand morphs in each level that are perceived as faces or hands, which we refer to as behavioral tuning (**Fig. 1D**).

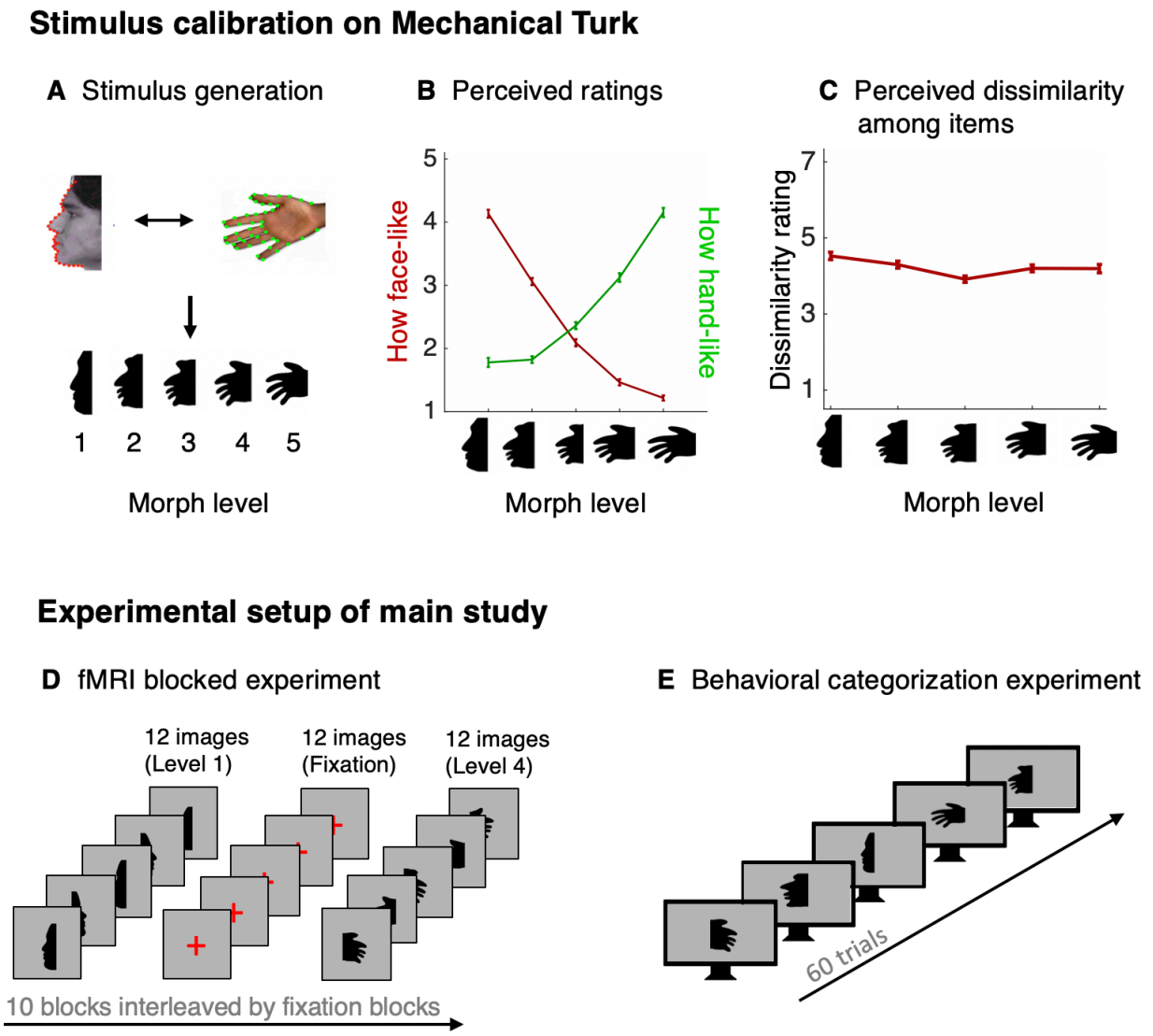

Figure 1. Stimulus calibration and experimental setup. A. Construction of face-hand silhouettes. *Top:* The contours of a profile face and open hand photograph are parameterized with 36 points in which points of maximum and minimum curvature are matched between faces and hands. *Bottom:* Five morph levels 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 (MTurk) 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) are tuned to a continuum of face- and hand-like silhouettes

We first asked how functional regions of interest (ROIs) identified in VTC with natural images respond to ambiguous face-like and hand-like silhouettes, respectively. To do so, we identified face- and body-selective regions within VTC: (1) pFus-faces/FFA-1, which is a face-selective region located in the posterior-lateral portion of the fusiform gyrus (FG[29–31]), (2) OTS-bodies/FBA, which is a body-selective region located within the occipitotemporal sulcus (OTS[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[29–31] (**Fig. 2**). From these ROIs, we then extracted neural responses (z-scored beta values from general linear model fits, see Methods) to the morphed silhouettes that were acquired during a second, independent experiment (**Fig. 1D**).

Qualitatively, both pFus- and mFus-faces showed differential responses, or neural tuning, to face-hand morphs. pFus- and mFus-faces showed the highest neural responses to face-like silhouettes and the lowest responses to hand-like silhouettes. Interestingly, responses were a) similar across the two morph levels that were more face-like and b) only declined past the center morph level. Comparatively, OTS-bodies showed a different tuning profile in which the highest neural responses were to hand-like silhouettes and the lowest responses were to face-like silhouettes. Additionally, responses in OTS-bodies were a) similar across morph levels that were more hand-like, as well as the center morph level, and b) only declined in response for the two morph levels that were more face-like (**Fig. 2B,D** and **Supplementary Fig. 1A,C**). These observations were statistically supported by a 3-way repeated measures ANOVA with ROI, morph level, and hemisphere as factors (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). Follow-up 1-way ANOVAs within each ROI (Bonferroni corrected) further supported an effect of morph-level (mFus-faces: F(4,120) = 5.08, p<0.005; pFus-faces: F(4,110) = 4.52, p<0.01; OTS-bodies: F(4,80) = 5.93, p<0.001). Together, these analyses reveal that a) face- and body-selective regions in VTC display neural tuning to ambiguous face-hand silhouettes and b) this tuning is not linear with the morphing continuum.

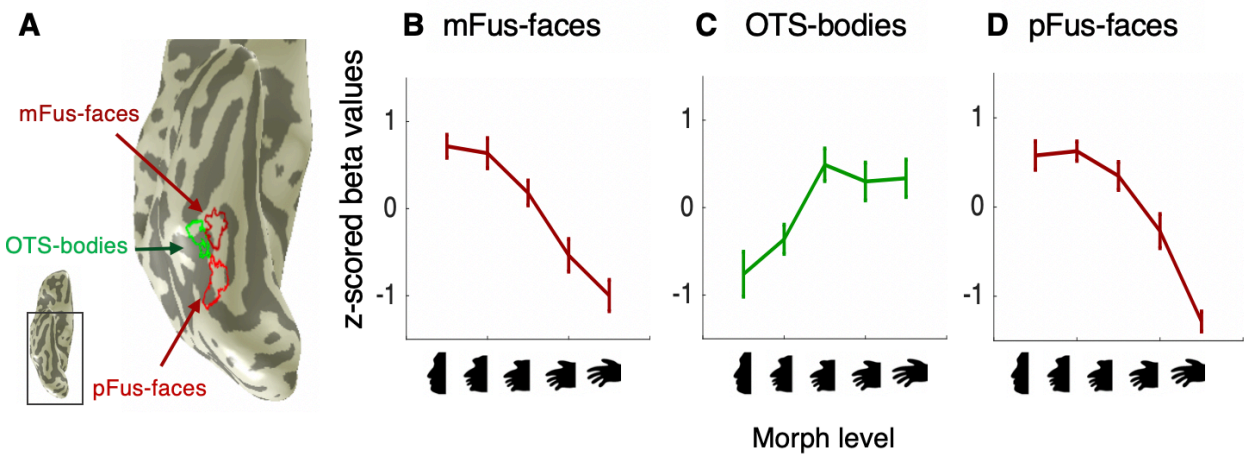

Figure 2. Face- and body-selective regions in ventral temporal cortex (VTC) display differential neural tuning to face-hand morphs. A. Zoomed image showing the inflated cortical surface in ventral view of an example participant illustrating three regions of interest (ROIs). From posterior to anterior: 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 the number of participants per ROI (N=11-14). See **Supplementary Fig. 1** for left hemisphere data.

Categorization behavior of face-hand morphs is best explained by neural tuning from multiple category-specific regions from different domains

Since face- and body-selective regions in VTC display selective tuning to face-hand silhouettes, we asked: *What is the best neural predictor for perceptual categorization of these ambiguous face-hand stimuli?* To answer this question, we first used the same face-hand silhouettes that participants previously viewed inside the scanner to measure how our participants perceptually categorized the same stimuli outside the scanner. In the behavioral experiment, participants viewed 60 silhouettes randomly drawn from the 5 morph levels (12 exemplars from

each morphing level) and were instructed to classify each image as either a face or a hand, in a self-paced manner (**Fig. 1E**). We refer to behavioral responses relating these forced-choice behavioral responses to the morph levels as *behavioral tuning* (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. On average, images in this middle, or intermediate, morph level 3 were classified slightly more as faces (proportion face response: .64±.26).

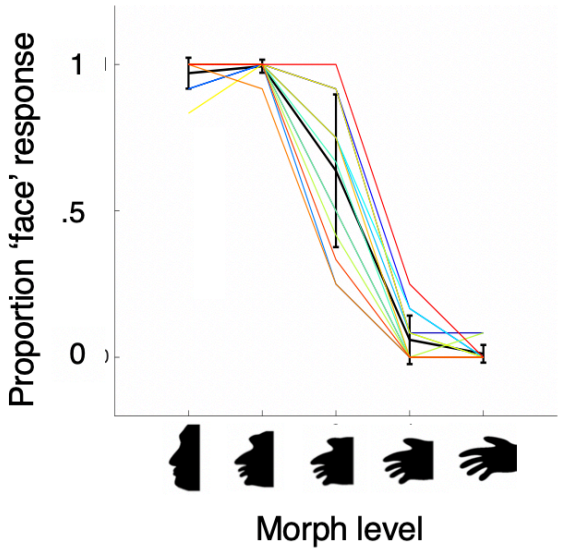

Figure 3. Behaviorally, participants effectively categorize the face-hand morphs. Behavioral tuning for each subject (colored lines) as well as the mean tuning across subjects (black line). Behavioral responses at each morph level indicate the participant's average performance across 12 trials. Proportion 'face' responses decline with morph level showing a categorical transition. The highest variability in behavioral responses across participants occurred for the center morph level. *X-axis:* morph level of silhouette (1 = face, 5 = hand). *Y-axis:* proportion face response. *Errorbars:* Standard deviation across the 14 participants.

We next tested if and how behavioral judgements are linked to neural responses in face- and body-selective regions. To do so, we conducted a linear regression analysis relating the behavioral categorization data to neural responses in face- and body-selective ROIs. In this analysis, the behavioral tuning of each subject is the dependent variable and the neural tuning of that subject is the independent variable (**Fig. 4A**). We considered two models: (1) a linear regression model relating behavioral categorization to neural responses from a single face- or body-selective region (**Fig. 4B, 1-ROI model**) and (2) a linear regression model relating behavioral categorization to neural responses from multiple regions (**Fig. 4B, Multi-ROI model**). For the latter, we implemented three versions of the model that contained neural responses from

a) two face-selective regions, b) a face-selective region and a body-selective region, or c) from two regions that were neither face- nor body-selective.

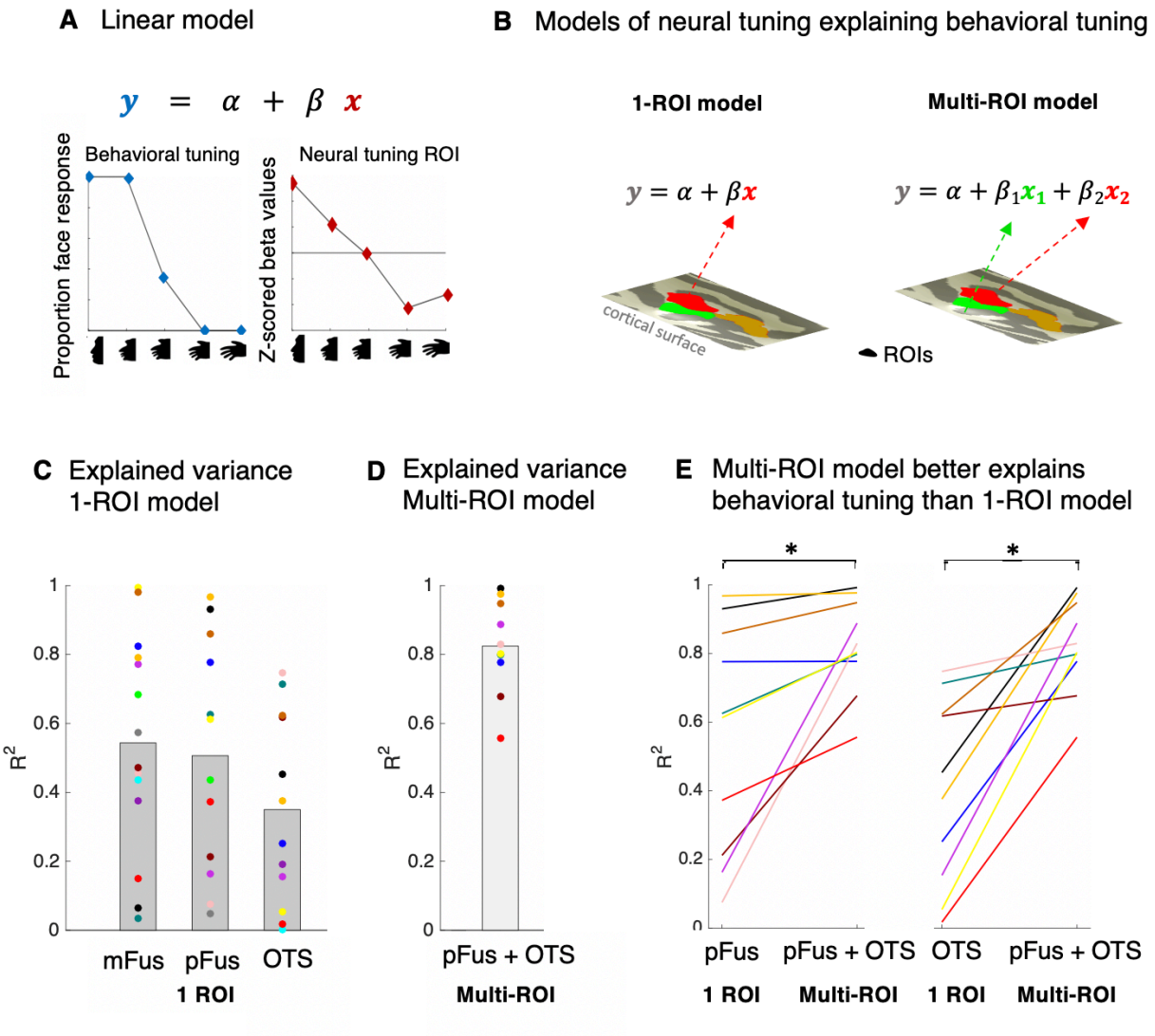

Figure 4. A stepwise linear regression model reveals that behavioral tuning to face-hand morphs can be predicted by a linear combination of the neural tuning from pFus-faces and OTS-bodies. A. *Top:* A linear regression model was used to relate behavioral and neural turning. α = model intercept, β = 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:* Explained variance 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.** Explained variance of the behavioral data by the MRM based on right pFus-faces and right OTS-bodies neural tuning. Conventions same as C. **E.** Improvement in explained variance 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 on average, behavioral categorization performance from individual participants was linearly related to neural responses from either face- or body-selective regions. That is, neural responses from either face (pFus-faces, mFus-faces) or body (OTS-bodies) ROIs explained, on average, a substantial amount of the variance of behavioral responses (R^2± standard error across subjects (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: whereby in some participants the variance explained was close to 1, but in others, it was close to 0 (R^2 min-max: 0 - .99; **Fig. 4C**). This between-subject variability in model fits suggests that the 1-ROI model is not a parsimonious model of behavioral categorization.

We next implemented a multiple ROI model where we first used a stepwise linear regression model to test which combination of ROIs best explains behavioral tuning. The stepwise regression was applied to data from all subjects and neural responses from all three ROIs (**Fig. 4B**). Interestingly, this model revealed that neural tuning from two ROIs: pFus-faces (LH and RH, *p*s<0.001) and OTS-bodies (LH and RH, *p*s<0.001) significantly explained behavioral categorization. Adding mFus-faces to the model did not significantly explain additional variance (LH and RH: *p*s = 0.18) and other combinations (mFus+pFus as well as mFus+OTS) under-performed the pFus+OTS model. Consequently, the remaining analyses only consider a model of behavioral categorization using pFus-faces and OTS-bodies as predictors, which we refer to as the Multi-ROI model (MRM).

Examining the MRM fit for each individual participant showed that the MRM model explained 0.71±.03 and 0.83±.04 (R^2±SE) of the variance of behavioral categorization from neural responses in the left and right hemispheres, respectively (**Fig. 4D, Supplementary Fig. 2B**). Comparing the variance explained by the MRM to the 1-ROI model revealed higher variance explained by the former than the latter in each participant (**Fig. 4E**). We tested the statistical significance of these data in two ways: (1) Akaike's Information Criterion (AIC) and (2) paired permutation tests between each of the 1-ROI models and the MRM. Mean 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. As lower AIC values indicate better model fits, the data suggest that a linear combination of neural responses of the right

pFus+OTS best explains behavioral categorization of face-hand silhouettes. Additionally, paired permutation tests showed that significantly more variance of behavioral categorization was explained by the MRM (all ps<0.05, Bonferroni corrected, **Fig. 4E**) than each of the 1-ROI models. Differences in model predictions across hemispheres are not significant (paired permutation tests between left and right hemisphere model: *p*s>0.05).

While these results are clear, it may just be the case that neural signals from multiple ROIs – irrespective of their selectivity – better explain behavioral categorization compared to responses from a single ROI. To test this possibility, we implemented an additional multiple ROI model that related behavioral categorization to neural tuning from two other functional regions. For this model, we used as predictors neural responses from ROIs that either did not illustrate a specific categorical preference (retinotopic area VO-2[34]), or were not selective to either faces or hands (Parahippocampal Place Area (PPA)/CoS-places[35–37]). This model had a significantly (permutation test, $p < .05$) 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. These analyses rule out the possibility that neural responses from any two regions are sufficient to explain categorization behavior of face-hand silhouettes.

A closer look at the MRM revealed that the model-estimated behavioral tuning closely matched the measured behavioral tuning of a given participant (**Fig. 5A,** compare to 1-ROI model estimations in **Supplementary Fig. 3**). Furthermore, 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**), which indicates that the behavioral categorization of ambiguous face-hand morphs is related to the differential neural responses between a face and a body-selective region.

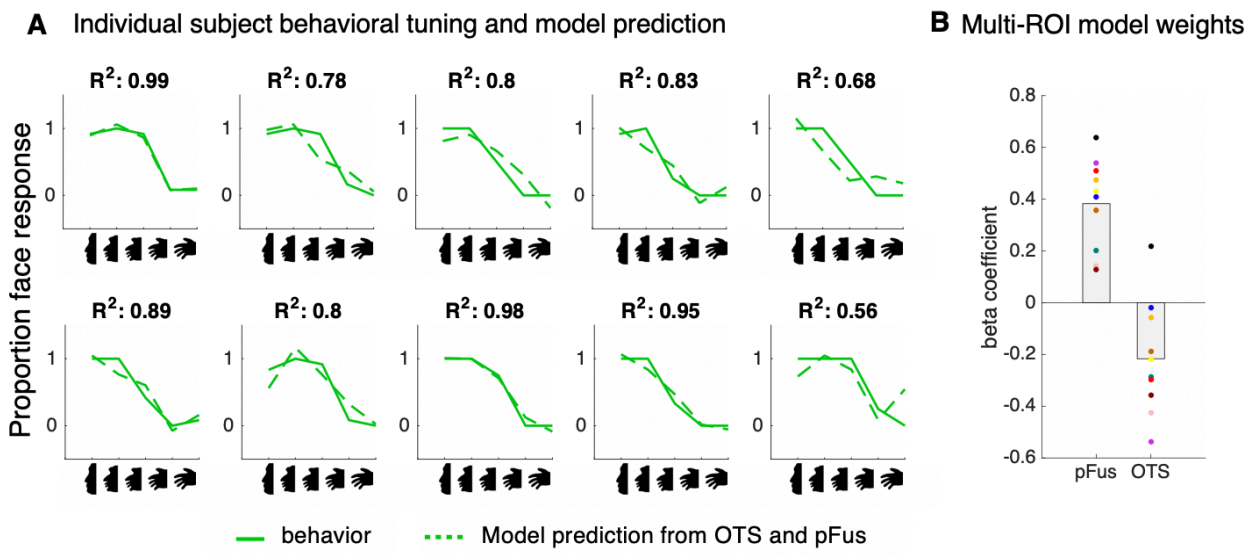

Figure 5. Individual differences in the predictions of the multi-ROI model. A. Each panel shows the behavioral tuning (solid) for a participant and the MRM prediction using the neural tuning from both pFus-faces and OTS-bodies (pFus+OTS) in the right hemisphere (dashed). *X-axis:* morph level; *Y-axis:* proportion 'face' responses. R^2: explained variance for each respective participant, 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.

The MRM built from a group of participants accurately predicts behavioral tuning in a new participant

We reasoned that if the relationship between neural and behavioral tuning revealed by the MRM is a subject-general relationship between brain and behavior, then a model built from neural responses in one group of participants should be capable of predicting behavioral tuning in a new person. However, if the relationship between neural and behavioral tuning is subject-specific, such a model would not predict well the behavioral tuning of a new participant. To investigate this possibility, instead of building an MRM for each subject, we built a single MRM that relates behavioral categorization to neural responses from pFus-faces and OTS-bodies in N-1 participants. The outputs of the model are two coefficients of the linear contribution of pFus-faces and OTS-bodies to behavioral categorization (**Fig 6A**). We then used the model coefficients to predict behavioral tuning in a new subject based on the new subject's neural tuning from pFus-faces and OTS-bodies (**Fig. 6A**). 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: $p<0.01$). The model fits of the latter model were also not significantly different from the left hemisphere cross-validated MRM-model (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**). Consistent with the within-subject MRM, the cross-validated MRM displayed opposing weights for pFus-faces and OTS-bodies, 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 behavioral tuning deviated most from the neural predictions at the middle face-hand morph, which is the morph-level that was associated with the largest within-subject variability (**Fig. 3**). Further, participants with 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 (a) neural tuning from two functional regions in human VTC predict behavioral tuning to face-hand morphs, and (b) this relationship is subject-general. That is, behavioral tuning in one participant can be predicted from the differential neural tuning across two regions with predictor weights derived from a separate group of participants.

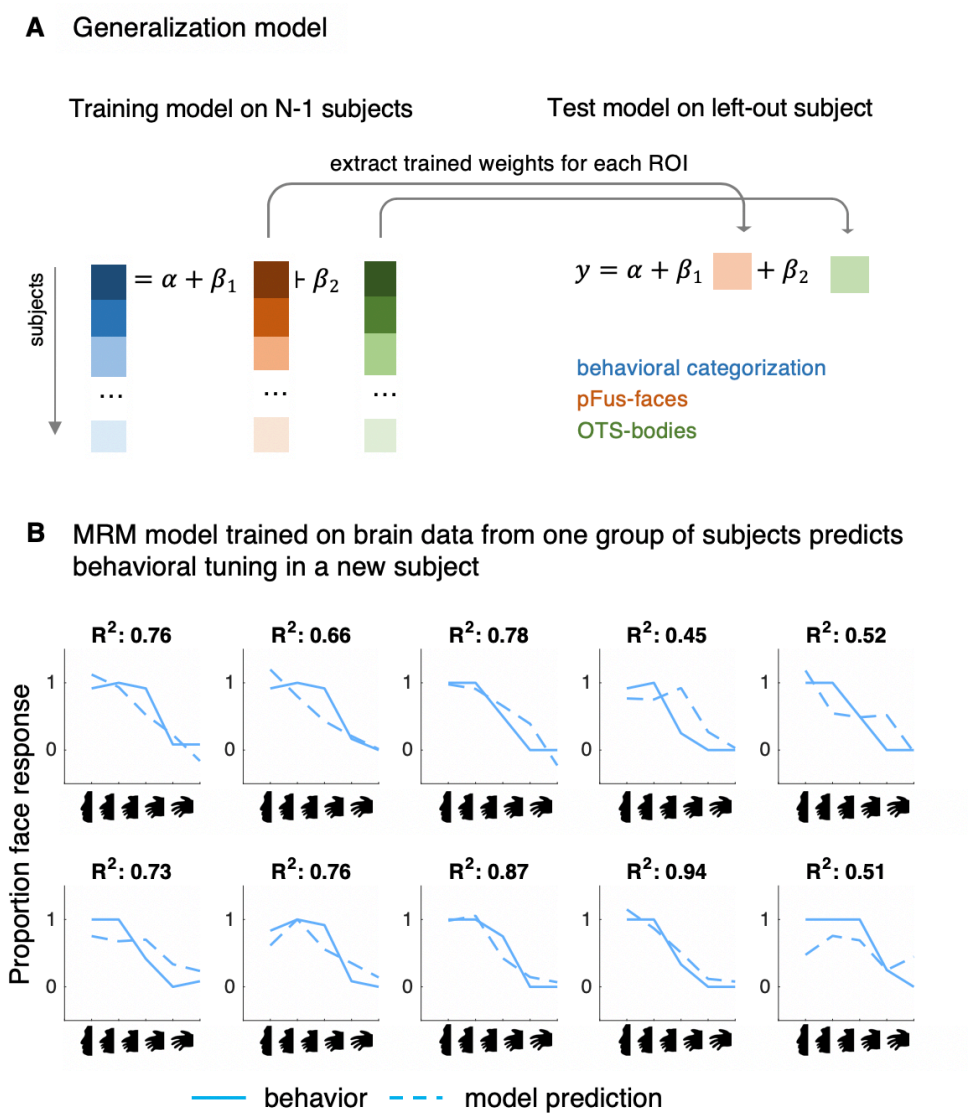

Figure 6. A cross-validated, multi-ROI model that is trained on brain data from one group of subjects predicts behavioral tuning in new subjects. A. Schematic illustration of the cross-validation procedure. *Left:* N-1 subjects were used to estimate model parameters relating behavioral categorization to 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-participant and were 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 participant. *Dashed blue line:* Model. R^2: explained variance for each left-out participant is indicated at the top of each subplot.

DISCUSSION

In the present study, we measured neural tuning to face-hand silhouettes 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 silhouette. 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 that neural tuning from both face- and body-selective regions together better predicted behavioral tuning compared to neural tuning from either type of 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 derived from the right hemisphere.

Here, we first propose that cortical adjacency likely enables the integration of neural signals between functional regions located within functionally-distinct networks. We then discuss how the best combination of neural tuning that predicts behavioral categorization may depend on the stimulus and task, as well as elaborate on the limitations of our findings. Finally, we discuss how the model-based approach implemented here is generalizable to future studies examining the relationship between neural and behavioral categorization beyond visual categories.

Cortical adjacency likely enables the integration of neural signals between functional regions located within functionally distinct networks

Given that a large body of work in humans and non-human primates supports the causal role of domain-specific regions within high-level visual cortex in the perception of their preferred domain, the present findings showing that the combination of responses from right pFus-faces and OTS-bodies best predict the categorization of face-hand morphs may seem surprising. However, when considering the fact that pFus-faces and OTS-bodies are two regions that are located within a larger cortical expanse – lateral ventral temporal cortex – that is specialized for processing animate categories, the present findings are more intuitive. Indeed, our findings suggest that because face- and body-selective regions are cortically adjacent within a larger animate representation may be particularly beneficial for discriminating animate categories from one

another in ambiguous situations. For example, cortical adjacency likely enables the combination of signals between regions through short-range connections and the combination of these neural signals likely contributes to categorical judgments.

Consistent with our results and the interpretation of our findings, recent anatomical and functional connectivity studies in macaques also propose the combination of neural signals between category-specific networks selective for faces and bodies. Indeed, while recent research in macaques identifies distinct cortical networks that separately process face and body information[38,39], these studies also identify connections that are shared between these two networks[39,40]. These studies propose that the subset of these networks that overlap with one another may be integral for the combination of face and body information both anatomically[39] and functionally[40]. Thus, future studies examining the relationship between neural tuning and categorization while recording responses from single neurons within adjacent face- and body-selective regions can test if this combination occurs at the neuronal level.

The best combination of neural tuning that predicts behavioral categorization may depend on the stimulus and task

Our present findings revealed that neural tuning from certain combinations of face- and body-selective regions more accurately predicted behavior than other combinations of neural tuning of face- and body-selective regions. 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 the improved prediction of behavioral categorization from pFus-faces and OTS-bodies rather than mFus-faces and OTS-bodies may be related to the stimuli and the respective location of pFus-faces and mFus-faces in the visual processing hierarchy. As mFus-faces is situated at a higher position in the visual hierarchy compared to pFus-faces[14,41], we hypothesize that the present stimuli (simple, two-dimensional shapes) were more suitable for pFus-faces. We also propose that behavioral categorization of more complex stimuli may be attained from the linear combination of neural tuning from mFus-faces and OTS-bodies than the combination of pFus-faces and OTS-bodies, which can be tested in future research.

Future studies can also 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[42,43]. 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. Altogether, we propose that the cortical adjacency of functional regions in VTC likely enables flexible combinations of neural tuning that can accommodate different types of stimuli and task demands.

Limitations

Using simplified stimuli to investigate complicated experimental questions such as those related to the perception of visual ambiguity is beneficial because the images 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. Such a space enables a clear quantification of how neural and behavioral responses change when exposed to stimuli that cross category boundaries.

Despite the controlled feature space of these stimuli, a concern may be that the brain-behavior relationships modeled in the present study may not generalize to naturalistic stimuli[44–47]. We acknowledge this limitation and propose that our present results and model form a foundation from which to build more complex brain-behavior models[48] underlying the perception of visual ambiguity – especially for animate categories as silhouettes of animate objects are more accurately recognized than other types of silhouettes[49,50]. Nonetheless, we underscore that neurally, silhouettes are sufficient to selectively drive high-level category-selective regions within VTC, as face- and body-selective regions respond selectively to silhouettes of their preferred category, shown in the present study as well as in previous work[4,5,8,51–54]. Future studies can extend our work and test additional categories as well as formats in order to examine if a) the combination of neural responses between functional regions best predicts behavioral responses to visual ambiguous

stimuli and b) these brain-behavioral predictions are indeed more prominent for animate categories.

Beyond the representation of animate categories in VTC

The linear model-based approach we have implemented in the present study is applicable to additional factors beyond the representation of animate categories in VTC. For example, neurally, previous studies have examined visual categorization in brain areas outside of VTC[55–57], as well as have considered how other brain areas (for example, within prefrontal cortex[58]) may act as gatekeepers that guide the categorical decision process within cortical regions in VTC. Additionally, categorization is not limited to the visual domain. For instance, empirical studies examine the categorization of (a) movements in motor cortex[59], (b) odors in piriform cortex[60], and (c) sounds in the visual cortex of blind individuals[61]. The approach we have implemented here can be applied to these examples, as long as two requirements are fulfilled. First, there should be a stimulus continuum between (at least) two categories. Second, there must be a 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 (a) different viewpoints of faces[62], (b) distinct motion directions[63], or (c) specific parts of the body surface[64]. 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 alone.

Conclusions

In summary, we find that neural categorization tuning from multiple category-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 can accommodate efficient and flexible categorization.

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[65,66] 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[67].

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[68]. 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[29].

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[33,69].

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 one 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**). As in the 1-ROI model, this was done individually for each subject.

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: ND designed the stimuli; KSW, ND, and KGS designed and carried out the experiments; 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. 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).

22. Serre, T., Olivia, A. & Poggio, T. A feedforward architecture accounts for rapid categorization. *PNAS* **104**, 6424–6429 (2007).

23. 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).

24. 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).

25. 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).

26. Davidenko, N. Silhouetted face profiles: A new methodology for face perception research.

J. Vis. **7**, 1–17 (2007).

27. F, T., K, N., JT, V. & N, K. Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* **21**, 753–759 (1998).

28. Grill-Spector, K. & Malach, R. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol. (Amst).* **107**, 293–321 (2001).

29. 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).

30. 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).

31. Weiner, K. S. & Grill-Spector, K. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage* **52**, 1559–1573 (2010).

32. Peelen, M. V & Downing, P. E. Selectivity for the human body in the fusiform gyrus. *J. Neurophysiol.* **93**, 603–8 (2005).

33. Schwarzlose, R. F., Baker, C. I. & Kanwisher, N. Separate Face and Body Selectivity on the Fusiform Gyrus. *J. Neurosci.* **25**, 11055–11059 (2005).

34. Fan, L. *et al.* Connectivity-based parcellation of the human temporal pole using diffusion tensor imaging. *Cereb. Cortex* **24**, 3365–3378 (2014).

35. Aguirre, G. K., Zarahn, E. & D'Esposito, M. An Area within Human Ventral Cortex Sensitive to "Building" Stimuli. *Neuron* **21**, 373–383 (1998).

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

37. 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).

38. 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).

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

40. 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).

41. 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

42. Behrmann, M. & Plaut, D. C. A vision of graded hemispheric specialization. *Ann. N. Y. Acad. Sci.* **1359**, 30–46 (2015).

43. 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).

44. Naselaris, T., Stansbury, D. E. & Gallant, J. L. Cortical representation of animate and inanimate objects in complex natural scenes. *J. Physiol. Paris* **106**, 239–249 (2012).

45. Huth, A. G., Nishimoto, S., Vu, A. T. & Gallant, J. L. A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron* **76**, 1210–1224 (2012).

46. Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453–458 (2016).

47. Nishimoto, S. *et al.* Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* **21**, 1641–1646 (2011).

48. Rust, N. C. & Movshon, J. A. In praise of artifice. *Nat. Neurosci.* **8**, 1647–1650 (2005).

49. Lloyd-Jones, T. J. & Luckhurst, L. Outline shape is a mediator of object recognition that is particularly important for living things. *Mem. Cogn.* **30**, 489–498 (2002).

50. Tian, M., Yamins, D. & Grill-Spector, K. Learning the 3-D structure of objects from 2-D views depends on shape, not format. *J. Vis.* **16**, 7 (2016).

51. Bracci, S., Ietswaart, M., Peelen, M. V & Cavina-Pratesi, C. Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. *J. Neurophysiol.* **103**, 3389–3397 (2010).

52. Shai, A., Ungerleider, L. G., Martin, A., Schouten, J. L. & Haxby, J. V. Distributed representation of objects in the human. *Neurobiology* **96**, 9379–9384 (1999).

53. Tong, F. & Nakayama, K. Robust representations for faces: Evidence from visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **25**, 1016–1035 (1999).

54. Mendola, J. D., Dale, A. M., Fischl, B., Liu, A. K. & Tootell, R. B. H. The Representation of Illusory and Real Contours in Human Cortical Visual Areas Revealed by Functional Magnetic Resonance Imaging. *J. Neurosci.* **19**, 8560–8572 (2018).
55. 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).
56. Jiang, X. *et al.* Categorization Training Results in Shape- and Category-Selective Human Neural Plasticity. *Neuron* **53**, 891–903 (2007).
57. Ferrera, V. P., Yanike, M. & Cassanello, C. Frontal eye field neurons signal changes in decision criteria. *Nat. Neurosci.* **12**, 1458–1462 (2009).
58. Baldauf, D. & Desimone, R. Neural Mechanisms of Object-Based Attention. *Science (80-.).* **1268**, 424–428 (2014).
59. Seger, C. A. & Miller, E. K. Category Learning in the Brain. *Annu. Rev. Neurosci.* **33**, 203–219 (2010).
60. 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).
61. 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).
62. Tanaka, K. Inferotemporal Cortex and Object Vision. *Annu. Rev. Neurosci.* **19**, 109–139 (1996).
63. 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).
64. Mountcastle, V. The columnar organization of the neocortex. *Brain* (1997).
65. Phillips, P. J. J., Wechsler, H., Huang, J. & P.J, R. The FERET database and evaluation procedure for face-recognition algorithms. *Image Vis. Comput.* **16**, 295–306 (1998).
66. 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).
67. Wandell, B. A., Chial, S. & Backus, B. T. Visualization and measurement of the cortical

surface. *J. Cogn. Neurosci.* **12**, 739–752 (2000).

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

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

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