Ramp coding with population averaging predicts human cortical face-space representations and perception

Short title: face coding in human visual cortex

Johan D. Carlin* & Nikolaus Kriegeskorte

* to whom correspondence should be addressed:
Johan D. Carlin
MRC Cognition and Brain Sciences Unit
15 Chaucer Road
Cambridge CB2 7EF
UK
johan.carlin@mrc-cbu.cam.ac.uk
+44 (0) 1223 355294

Abstract

Face space provides a popular metaphor for the representation of individual faces. Although computer-graphics models of face space have a long history, their relationship to the cortical code for individual faces remains unclear. We used such a model to generate animations of faces with realistic 3D shape and texture and analyzed fMRI responses to each individual face. We developed and evaluated multiple neurobiologically plausible computational models of face-space coding, each of which predicts a representational geometry and a regional-mean activation profile. A population code of units with sigmoidal ramp tuning over the face-space dimensions explained both pattern and regional-mean fMRI effects better than alternative models, but only in conjunction with a readout-level population averaging mechanism. This model also accounted for perceptual similarity judgments. Our study demonstrates the importance of modeling readout-level population averaging and provides a computational account of the cortical representation underlying human face processing.

Introduction

Humans are expert at recognizing individual faces, but the neural mechanisms that support this ability are poorly understood. Multiple areas in human occipital and temporal cortex exhibit representations that discriminate between individual faces, as indicated by successful classification of brain responses in pattern information analyses of functional MRI (fMRI) data (1–10). However, the nature of this representation remains obscure because individual faces differ along multiple stimulus dimensions, which renders successful classification of any particular face pair inherently ambiguous. Here, we address this challenge by using representational similarity analysis (RSA) (11) to fit computational models of face coding to the multivariate discrimination performance of cortical regions in visual cortex as well as to perceptual similarity judgments. Our aim is to better understand the underlying computations for face coding by inspecting the properties of the model that best predicts cortical and perceptual face representations.

A key challenge for the application of RSA to face representations is that most cognitive and neuroscientific models of face processing do not make predictions at the level of distances between particular faces (12,13). However, such predictions can be obtained from models based on the notion that faces are encoded as vectors in a space (14). Most face-space implementations apply principal components analysis (PCA) to images or laser scans in order to obtain a space where each component is a dimension and the average face for the training sample is located at the origin (15,16). In such PCA face spaces, eccentricity is associated with judgments of distinctiveness while vector direction is associated with perceived identity (17–19). Initial evidence from macaque single unit recordings and human fMRI suggests that brain responses to faces are strongly modulated by face-space eccentricity, with most studies finding increasing responses with distinctiveness (20–23). However, there has been no attempt to develop a unified account for how a single underlying face-space representation can support both multivariate direction sensitivity and region-mean eccentricity sensitivity.

Here, we characterize how distances in a PCA face space relate to the psychophysical similarity judgments and to region-mean as well as multivariate fMRI effects. Our approach is inspired by circuits-level studies of early visual cortex, where insights are obtained from the manner in which visual responses deviate from the predictions of a simplified reference model (24–26). We use an analogous RSA approach where Euclidean distances in a PCA face space acts as a reference against which distances estimated from cortical representation and perception are compared. Anticipating our results, we demonstrate a substantial over-representation of eccentricity information in cortical face-space reconstructions, consistent with previous fMRI adaptation results (21). We then fit multiple computational models to the data in order to demonstrate that such warped representations are consistent with a sigmoidal ramp-based face-space coding scheme combined with a readout-level population averaging mechanism.

Results

Sampling face space with photorealistic but physically-controlled animations

We generated a set of realistic face animations by rendering horizontally rotating meshes obtained from a PCA model of 3D face shape and texture (27). Each frame of the animations was cropped with a feathered aperture and processed to equate low-level image properties across the stimulus set (Experimental Procedures). We generated 12 faces from a slice through the high-dimensional PCA model (Figure 1b). Euclidean distances between the Cartesian coordinates for each face were summarized in a distance matrix (Figure 1a), which served as the reference for comparisons against distances in the reconstructed perceptual and neural face spaces (Figure 1c-h). A physically distinct stimulus set with the same underlying similarity structure was constructed for each participant by randomizing the orientation of the slice through the PCA space (for examples, see Figure S1). This ensured that group-level effects of face-space direction and eccentricity were not strongly influenced by idiosyncrasies of particular face exemplars.

Figure 1. Face spaces obtained from a reference PCA model, perceptual judgments and fMRI response patterns in human visual cortex. (**a-b**) We generated 12 faces in a polar grid arrangement on a 2D slice through the reference PCA space (3 eccentricity levels and 4 directions). The grid was centered on the stimulus space norm (not shown). Euclidean distances between the 12 faces are illustrated in a distance matrix (a) and in a 2D reconstruction of these distances (b, multidimensional scaling, metric-stress criterion). (**c**) Group-average perceptual face space reconstruction (N=10) estimated from psychophysical similarity judgments. The distance matrix depicts the percentage of trials on which each pair of faces was rated as relatively more dissimilar. (**d-h**) Group-average cortical face-space reconstructions (N=10 except for occipital face area, N=9) estimated from fMRI response patterns in human visual cortex. The distance matrices depict a cross-validated estimate of multivariate discriminability where 0 corresponds to chance-level performance (Experimental Procedures). Each cortical region was defined in individual volunteers using independent data.

Reconstructed cortical and perceptual face spaces are warped with regard to the reference PCA space

Human volunteers participated in a perceptual judgment task followed by fMRI scans (Experimental Procedures). Perceptual dissimilarity judgments were correlated with Euclidean distances in the reference PCA space (Figure 1c, Figure 4a, gray bar). Face spaces reconstructed from fMRI responses also correlated with the reference PCA space (Figure 1d-h, Figure 4b-f, gray bars). However, although statistically significant (all p<=0.002), these correlations did not reach the estimated noise ceiling in the sample (shaded region in Figure 4b-f, Experimental Procedures), which indicates that a linear relationship with the reference PCA space could not capture all the explainable variance in cortical face spaces.

Visual inspection of the reconstructed face spaces suggests that relative to the reference PCA space, cortical face spaces systematically over-represent eccentricity relative to direction (Figure 1d-h). However, we also observed clear evidence for effects of direction independent of eccentricity. Discriminant distances for faces that differed in direction but not eccentricity exceeded chance-level performance (p<0.05) for typical and caricatured faces in all regions (Figure S2), while sub-caricatured faces were less consistently discriminable. Indeed, direction discrimination increased linearly with eccentricity in each region except the parahippocampal place area (sub>typical>caricature, all p<0.036, Figure S2), suggesting a dose-dependent relationship between face identity information and caricaturing. Thus cortical regions discriminate identity even in the absence of a difference in eccentricity, suggesting that cortical face representations cannot be reduced to a one-dimensional code based on distinctiveness.

Face-space warping reflects an over-representation of eccentricity information

We quantified the apparent warps in the cortical face space reconstructions by constructing a multiple regression RSA model where the effects of eccentricity and direction were estimated separately (Figure 2a). A group-level analysis of the resulting parameter estimates showed that both eccentricity and direction made reliable contributions to cortical and perceptual face spaces (Figure 2b-g, all p<0.047). However, eccentricity estimates were reliably greater than direction estimates (all p<0.001). This pattern could be observed in both cortical and perceptual face space reconstructions and in each individual volunteer (gray lines in Figure 2), although the eccentricity-direction difference was considerably larger in cortical face spaces. For instance, the fusiform face area (FFA) exhibited a 6.4 times greater effect of eccentricity than direction (Figure 2f), while this ratio was 1.5 for the perceptual face space (Figure 2b). In each case, inspection of the residual distances showed that little structure remained after modeling eccentricity and direction. Thus, face space warping with regard to the reference PCA space could be safely attributed to the estimated imbalance between eccentricity and direction coding, rather than being driven by other idiosyncratic components such as low-level features of the stimuli. Taken together, cortical face spaces exhibited a substantial over-representation of distinctiveness-related eccentricity information over identity-related direction information.

Figure 2. Face-space warping reflects an over-representation of eccentricity over direction information. (**a**) Squared distances in the reference PCA space were parameterized into a predictor coding variance associated with eccentricity (green) and a predictor coding variance associated with direction (blue). The resulting model was fitted each volunteer's face-space reconstructions (Experimental Procedures). Importantly, the scaling of the eccentricity and direction predictors ensures that equal parameter estimates corresponds to a preserved reference PCA space. (**b**) Multiple regression fit to the reconstructed perceptual face space with group-average parameter estimates (top), fitted distances (middle) and residuals (bottom). Gray lines reflect single volunteer parameter estimates and p values above each bar are obtained from T tests (N=10). Significant differences (paired T test, p<0.05) are indicated with horizontal connectors. (**c-g**) Multiple regression fit to the reconstructed cortical face spaces, plotted as in b (N=10 for all panels except d, N=9).

Cortical face-space warping could not be explained by region-mean activation preferences for caricatures. We performed a region-mean analysis of responses in each cortical area, which confirmed previous reports that fMRI responses increase with distinctiveness across much of visual cortex (Figure S3) (23). However, a control discriminant analysis where we corrected for region-mean biases showed similar results as the original analysis (Experimental, Procedures, Figure S4), suggesting that the eccentricity over-representation in the discriminant analysis could not be reduced to a single response pattern scaling or shifting with the conditions. Thus, although region-mean eccentricity effects were apparent in all visual areas, the warping of the cortical face spaces could not be attributed to region-mean effects alone. Furthermore, region-mean fMRI effects cannot explain the presence of smaller but reliable warps in the perceptual face space (Figure 2b).

A sigmoidal ramp-coding model with population averaging explains face-space warping

We developed a computational model to describe how eccentricity over-representation can be explained as an emergent property of partial averaging over a population of sigmoidal ramp-tuned model units. The model, which is inspired by known preferences for extreme feature values in single units recorded from area V4 and face-selective patches in the macaque visual cortex (28,29), proposes that the representational space is tiled with randomly oriented ramps, each of which exhibits a monotonically increasing response along its preferred direction (Figure 3a). We modeled the response along the preferred direction using a sigmoidal function with two free parameters (horizontal offset o and response saturation s, Experimental Procedures). The random orientation of the sigmoidal ramp summarizes our expectation that single neurons in the underlying face representation are tuned to linear combinations of the principal components in the reference PCA space. We approximated the effects of individual unit responses being partially averaged prior to readout with a parameter that controlled the extent to which each individual model unit's response vector was translated toward the population-mean response vector (population averaging p, Figure 3b-c). In practice, the consequence of such readout-level population averaging was that individual ramps exhibited a substantial U-shape in their response functions, with only a minor deflect in favor of their preferred face-space direction (Figure 3a, right panel). Such readout-level effects may arise from extrinsic measurement sources such as insufficient spatial precision in fMRI responses or from intrinsic neural circuits mechanisms such as pooling over inputs in a cortical hierarchy. We fitted the 3 global parameters to each volunteer's face-space reconstructions using a grid search procedure that identified the parameter values that maximized the median Pearson correlation between the reconstructed face space and Euclidean distances in the ramp model's response to the 12 faces over 100 initializations of the model (for example grid search outputs, see Figure S5).

Figure 3. A computational model for face-space coding based on sigmoidal ramp tuning and readout-level population averaging. The visualization uses the model parameters that were optimal for predicting the fusiform face area's face space (Figure 5). (**a**) Sigmoidal response function from the model's internal representation (left panel) and the readout (right panel) following translation toward the population-average response function (middle panel). (**b**) Two-dimensional generalization of the sigmoidal ramp tuning function to encode a direction in the face-space slice. An example unit is plotted in the left and right panels with the population-average response in the middle panel. (**c**) The model's representational dissimilarity structure was estimated as the Euclidean distance between the population response vectors elicited by two coordinates in the face-space slice. The two test face exemplars illustrate how dissimilarity increases more rapidly with radial (eccentricity) distance than with tangential (direction) distance in the model's representation.

The ramp model was a reliably better fit both to cortical and perceptual face spaces than the Euclidean distances in the reference PCA space (all p<0.006, Figure 4). Indeed, performance for the ramp model approached the noise-ceiling estimate for each reconstructed face space, suggesting that a partially averaged population of sigmoidal ramp-tuned model units provided a complete account of the reliable pattern-level fMRI and perceptual effects in the sample. By contrast, the fit to coding schemes based on low-level visual features did not reach the noise ceiling in the cortical face spaces (Figure 4b-f, purple bars), suggesting that the predictions of the ramp model were dissociable from accounts based on coding low-level physical similarity between the faces. In summary, both cortical and perceptual face spaces were consistent with an underlying representation based on sigmoidal ramp tuning and population averaging. The ability of this model to fit the reconstructed face spaces suggests that effects of distinctiveness can arise as an emergent property of an underlying representation where individual model units encode face space direction rather than eccentricity. Thus, over-representation of face-space eccentricity can be modeled as a readout-level phenomenon rather than as a feature of the underlying neuronal representation as such.

Figure 4. Distances in a neuronal model based on sigmoidal ramp tuning and population averaging fits the cortical and perceptual face spaces better than distances in the reference PCA space or other face coding schemes. Group-average distance matrices for the best-fitting models for the cortical and perceptual face spaces (top), and group-average prediction performance for the ramp model as well as competing face space coding schemes (bottom). Black lines indicate 1 standard error of the mean, and the shaded area indicates the noise-ceiling estimate, which is obtained from between-subject distance matrix correlation (Experimental Procedures). The free parameters of the ramp model were estimated in individual volunteers from a training split of the dataset and the resulting model based on the group-average parameters was correlated with the left-out volunteer's face-space reconstruction. This ensured that the ramp model's performance could be compared to the remaining face coding schemes, which are fixed models without free parameters. The ramp model is plotted together with the performance of the reference PCA space distances (gray) and two control models based on low-level features of the stimuli (purple bars, Experimental Procedures). Horizontal connectors and p values above bars were generated as in Figure 2.

Differences between cortical and perceptual face spaces are explained by different degrees of population averaging

The ramp model fit approached the noise ceiling both in cortical and perceptual face spaces even though these exhibited distinct levels of eccentricity over-representation. Inspection of the best-fitting ramp model parameters showed that this flexibility in the model's representation could be attributed to a higher level of population averaging in each of the cortical face spaces than in the perceptual space (all p<0.03, Figure 5). Overall, all cortical face spaces exhibited high proportions of population averaging with over 70% of the response being attributed to population average responses rather than unit-specific coding. This high level of averaging is consistent with the low spatial precision of the blood oxygen level-dependent fMRI response, which may introduce such readout averaging both by measuring neuronal activity indirectly through vascular responses and by sub-sampling these hemodynamic responses into a grid of voxels (30,31). By contrast, the parameters that controlled the shape of the sigmoidal ramp response function did not differ between any of the face space reconstructions (all p>0.1), suggesting that cortical and perceptual face spaces did not differ reliably in their ramp tuning functions. These results suggest that widely distributed cortical face space representations can support perceptual similarity judgments, with apparent differences between the cortical and perceptual representations accounted for by readout-level averaging in fMRI responses.

Figure 5. Cortical and perceptual face spaces were fitted using similar sigmoidal response functions but distinct levels of population averaging. The bars depict group-average ramp model parameters for the cortical and perceptual face spaces. Black lines depict 1 standard error, and significant differences (paired T test, p<0.05) are indicated with horizontal connectors. The perceptual face space is fitted with a lower proportion of population averaging relative to each of the cortical face spaces. No other pairwise comparisons were statistically significant.

Additionally, population-averaging effects may reflect mechanisms of the neural circuits that transform perceptual representations into behavioral outputs. Note that the perceptual face space reconstruction was fitted with a substantial amount of population averaging as well, suggesting that averaging occurs during perceptual processing and not merely as a consequence of fMRI response sampling. Thus, one interpretation of these population averaging effects is that the perceptual estimate reflects intrinsic averaging arising from pooling over inputs at multiple stages of the cortical hierarchy, while the difference between this averaging level and the higher level found in cortical reconstructions reflects the additional contribution of extrinsic averaging arising from spatially imprecise fMRI measurements.

Both sigmoidal ramp coding and population averaging are necessary for a complete account of cortical face-space representations

The two central features of the computational model we propose here are sigmoidal ramp-tuned model units with preferences for extreme feature values and readout-level population averaging. We tested whether these features are necessary to account for human face-space representations by developing an alternative model based on exemplar coding and by manipulating the availability of the population averaging computation to the ramp and exemplar models. The exemplar model proposes that the representational space is sampled by a Gaussian distribution of exemplars, where each unit prefers a Cartesian coordinate in the space rather than a direction and the response of each unit scales with an isotropic Gaussian rather than a sigmoidal ramp (Figure 6a, Experimental Procedures). We fitted this model similarly to the ramp model with two free parameters that controlled the width of the Gaussian distribution from which tuning centers were sampled and the width of a Gaussian controlling the fall-off of each unit's response with Euclidean distance from the tuning center.

Figure 6. Only a ramp model with population averaging provides a complete account of cortical face representations. (**a**) Response profile for an example unit from an exemplar model based on the group-average parameters that were optimal for predicting the fusiform face area. The filled circle indicates the coordinate of the example unit's peak response. Tuning centers from a subset of other model units are overlaid in black outlines for illustrative purposes along with the face exemplars (plotted as in Figure 1). (**b**) Prediction performance for the ramp and exemplar models fitted to the perceptual face-space reconstruction. Gray bars indicate performance with population averaging, white bars performance with no averaging. These cross-validated estimates were calculated as in Figure 4. (**c-g**) Left: prediction performance for the ramp and exemplar models fitted to the cortical face space reconstructions. Plotted as in panel b. It can be seen that the ramp model with population averaging has comparable prediction performance to the exemplar models, where population averaging has little effect. Middle and right: Comparison between observed region-mean fMRI response and population-mean model responses for the exemplar model (middle, no population averaging) and ramp model (right, with population averaging). Each point reflects a single-volunteer estimate for a given eccentricity level (arbitrary image intensity units on y axis). The slopes provide single-volunteer estimates of the linear effect of model population-mean response as a predictor for region-mean fMRI response. Region-mean fMRI responses generally increase with eccentricity, which is consistent with the ramp model but inconsistent with the exemplar model.

We found that population averaging was necessary for constructing a ramp model that could predict the explainable variance in cortical face representations. The ramp model without population averaging did not reach the noise-ceiling estimate and performed reliably worse than the full ramp model in each cortical region (all p<0.007, Figure 6c-g). This result is consistent with the Johnson-Lindenstrauss lemma, according to which coordinates in a space constructed from random projections approximately preserves distances in the original space. Thus, population averaging is necessary to achieve a ramp model that exaggerates eccentricity relative to direction in the manner exhibited by the cortical face-space reconstruction. By contrast, the ramp-model fit to perceptual judgments was similar with and without population averaging (p>0.19), consistent with the relatively smaller contribution of population averaging to the ramp-model fit. Population averaging had inconsistent effects on the exemplar model's prediction performance. Some reconstructions were better predicted with averaging (PPA, p=0.012), others without (perceptual judgments, p=0.043), while the remainder showed no reliable differences (all p>0.24). Thus, population averaging was a necessary component for successfully fitting the ramp model to cortical face-space reconstructions, but its advantages were specific to the ramp model and cortical face-space reconstructions.

The exemplar model's prediction performance was comparable to the ramp model with population averaging, but the exemplar model did not accurately predict region-mean fMRI responses. With sporadic exceptions (early visual cortex, p=0.043), we did not observe reliable differences in prediction performance between the ramp model with population averaging and the exemplar model with or without averaging (Figure 6b-g, left panels). Thus, pattern-level cortical and perceptual face-space reconstructions could not be used to reliably distinguish underlying representations based on exemplar coding from sigmoidal ramp coding with population averaging. However, an analysis of the population-mean predictions of the two models provided strong evidence in favor of the ramp model. Even though these models were fitted to distance matrices rather than region-mean fMRI responses, we found that they exhibited systematic population-mean response preferences as a function of face-space eccentricity. Consistent with the region-mean fMRI effects (Figure S3), the ramp model with population averaging exhibited increasing population-mean responses with eccentricity, while the exemplar model generally predicted decreasing responses with eccentricity (slopes in Figure 6c-g, middle and right panels). Thus, region-mean fMRI responses enabled us to adjudicate between the ramp and exemplar models, even though both models had comparable prediction performance in fits to pattern-level cortical and perceptual face space reconstructions. This highlights the power of fMRI in distinguishing between computational models that make equivalent predictions for behavior (see also (32)).

We explored whether the population-mean preference for sub-caricatures in the exemplar model could be altered, but found that such models do not predict cortical face-space reconstructions with the same performance as the other models we tested. The preference for sub-caricatures in the standard exemplar model arises as a necessary consequence of the Gaussian distribution of tuning centers, which over-samples the center relative to the periphery of the space. This tendency is further magnified by the free parameter controlling the width of this tuning-center distribution, where cortical fits consistently favored compressed distributions with tuning centers close to the center of the space (see FFA example in Figure 6a). We tested models with inverted Gaussian tuning-center distributions since such models might produce more appropriate population-mean response preferences for caricatures. However, the inverted-Gaussian exemplar model's prediction performance was reliably worse than the standard-Gaussian exemplar model for all cortical face-space reconstructions (Figure S6, all $p<0.027$). Thus, exemplar models provide accurate fits to pattern-level cortical face space reconstructions when the tuning-center distribution is Gaussian, but such distributions necessarily lead to inaccurate predictions for region-mean fMRI response preferences. In summary, the ramp model with population averaging was the only model we evaluated that provided a complete account of pattern-level face-space reconstructions, region-mean fMRI responses, and perceptual judgments.

Discussion

This study investigated the representation of individual faces in human visual cortex and in perceptual judgments by comparing estimated distances in reconstructed face spaces to distances in a reference PCA model of 3D shape and texture and to computational models of face-space coding. We found that cortical face spaces systematically over-represent face-space eccentricity relative to direction. Perceptual face spaces exhibited similar warping, but to a lesser degree. A simple neuronal model based on sigmoidal ramp tuning provided a good fit to both cortical and perceptual face spaces, provided that a parameter was included to account for readout-level population averaging of the individual model units.

Sigmoidal ramp tuning explains human face-space representation

Our findings suggest that sigmoidal ramp tuning supports human face-space coding. Although we showed that Euclidean distances in the reference PCA space correlates with cortical and perceptual face-space reconstructions, this relationship was imperfect relative to the ramp model's performance. Importantly, the advantage for the ramp model arises from its ability to capture readout-level population averaging effects, not because its internal representation in the absence of averaging differs substantially from coding Euclidean distances in the reference face space. Instead, the key contribution of our model is to provide a description of a face-space coding scheme that recapitulates these reference face-space distances while also exhibiting over-representation of eccentricity as a consequence of readout-level population averaging. Thus, the ramp model provides a computational account for how face-space coding is structured in cortical representation and for how readout-level effects might warp this representation.

Distinctiveness effects as an emergent property of a neuronal population with sigmoidal ramp tuning

Our results are consistent with a model where sensitivity to face-space eccentricity at the region-mean activation level is supported by a population of direction-tuned units rather than an explicit representation of face distinctiveness or associated psychological constructs. This view is also supported by evidence from single-unit recording studies, where cells generally are tuned to particular features with a preference for extreme values rather than responding to eccentricity regardless of direction (28,29). Our model exemplifies how neuronal representations may differ qualitatively from estimates of these representations at the level of fMRI voxels. Related effects have been reported in attention research, where response-gain and contrast-modulation effects at the single neuron level may sum to similar additive-offset effects at the fMRI-response level (33). In summary, direct interpretation of fMRI effects in terms of representations of distinctiveness may be misleading.

Cortical face space representations are widely distributed

We observed highly similar face space representations across visual cortex, including scene-selective, face-selective and early visual cortex. Such widely-distributed effects are consistent with previous face identity decoding studies (2,4). Unlike the clear anatomical clustering of region-mean selectivity for faces relative to other categories in regions such as FFA, multivariate within-category discrimination of face exemplars appears to involve a widespread network of visual areas.

It may appear surprising that a face-space coding model should explain representations in regions that are unlikely to encode faces specifically. However, the ramp-tuning model is agnostic with regard to the features encoded by the dimensions of the representational space, which need not be face-specific. For instance, local curvature increases with face space eccentricity and is encoded in a ramp-like manner at intermediate stages of visual processing in macaque V4 (28). Similarly, even a Gabor filter-based representation as envisioned by classic models of V1 simple cells (34,35) would likely possess some sensitivity to face-space direction because the contrast of local orientation content varies with major face features such as eyebrow or lip thickness. There are thus multiple mechanisms by which regions without specialized face processing may still exhibit the type of face-space coding we observed here. A corollary of this point is that the face-space effects we report are likely to reflect a general mechanism for object individuation in visual cortex rather than specialized processing for face recognition as such.

Population averaging has wide-spread applicability in computational modeling of perception and cortical representation

We found that manipulation of a global population averaging parameter was sufficient to reconcile disparate distance estimates from perceptual judgments and cortical representation with a single underlying representation based on sigmoidal ramp coding. By contrast, a mechanistic account of how population-averaging effects arise in perceptual judgments or fMRI responses is likely intractable at present since the circuit-level properties of these transformations are poorly understood, and the set of parameters involved would be large. For instance, a complete account of how fMRI responses sample neuronal activity remains a topic for on-going research (36–38), and a model that fits all likely parameters is unlikely to yield a unique solution. Instead of estimating such a complete model we approximate its effects at the population distance-matrix level by specifying a single parameter that controls the extent to which individual model unit responses are translated toward the population-average response. Despite its simplicity, the model captures nearly all the explainable variance in the current study, suggesting that population averaging is a reasonable approximation for how readout effects affect perception and cortical representations at the distance-matrix level.

Although the current study concerns face representations we expect similar modeling approaches to have broad applicability in related fields. The model we estimate is specified at the level of dimensional feature coding with unimodal tunings and does not explicitly encode any face-specific mechanisms. Thus, it could be applied to a wide range of perceptual representations including coding of color, shape and non-face objects. Consistent with this possibility, we found that the model could account for effects outside of face-specific cortex. We expect population averaging to prove a useful concept for modeling representations across a range of perceptual domains and recording modalities.

Materials and Methods

Sampling the reference PCA face space

We generated faces using a norm-based model of 3D face shape and texture, which has been described in detail previously (15,27). Briefly, the model comprises two principal components analyses (N=200 faces), one based on 3D shape estimated from laser scans and one based on texture estimated from digital photographs. The components of each PCA solution are considered dimensions in a space that describes natural variation in facial appearance. We yoked the shape and texture solutions in all subsequent analyses since we did not have distinct hypotheses for these.

We developed a method for sampling faces from the reference PCA space in a manner that would maximize dissimilarity variance. This is related to the concept of design efficiency in univariate general linear modeling (39), and involves maximizing the variance of hypothesized distances across the stimulus set. Because randomly sampled distances in high dimensional spaces tend to fall in a narrow range of distances relative to the norm (40), we reduced each participant's effective face space to 2D by specifying a plane which was centered on the norm of the space and extended at a random orientation. The face exemplars constituted a polar grid with 4 directions at 60 degrees separation and 3 eccentricity levels (scaled at 30%, 100% and 170% of the mean eccentricity in the training faces set). The resulting half-circle grid on a plane through the high-dimensional space is adequate for addressing our hypotheses concerning the relative role of direction and eccentricity coding under the assumption that the high-dimensional space is isotropic. In preliminary tests we observed that this method yielded substantially greater dissimilarity variance estimates than methods based on Gaussian or uniform sampling of the space.

Face animation preparation

We used Matlab software to generate a 3D face mesh for each exemplar. This mesh was rendered at each of the viewpoints of interest in the study in a manner that centered the axis of rotation on the bridge of the nose for each face. This procedure ensured that the eye region remained centered on the fixation point throughout each animation in order to discourage eye movements. Renders were performed at sufficient increments to enable 24 frames per second temporal resolution in the resulting animations. Frames were converted to gray-scale and cropped with a feathered oval aperture to standardize the outline of each face and to remove high-contrast mesh edges from the stimulus set. Finally, we performed a frame-by-frame histogram equalization procedure where the average histogram for each frame was imposed on each individual face. Thus, the histogram was allowed to vary across time but not across faces. Note that histogram matching implies that the animations also have identical mean gray-scale intensity and root-mean-square contrast.

A potential concern with these matching procedures is that they could affect the validity of the comparison to the reference PCA space. However, we found that the opposite appeared to be true: distances in the reference PCA space were more predictive of pixelwise correlation distances in the matched space than in the original frames. Thus, the matching procedure did not remove features that were encoded in the PCA space and may in fact have acted to emphasize such features.

Participants

10 healthy human volunteers participated in a similarity judgment task and fMRI scans. The psychophysical task comprised 3-4 separate days of data collection which were completed prior to 4 separate days of fMRI scans. All procedures were performed under a protocol approved by the Cambridge Psychology Research Ethics Committee (CPREC). Volunteers were recruited from the local area (Cambridge, UK) and were naïve

with regard to the purposes of the study. Five additional volunteers participated in initial data collection but were not invited to complete the study due to difficulties with vigilance, fixation stability, claustrophobia and/or head movements inside the scanner. The analyses reported here include all complete datasets that were collected for the study.

Perceptual similarity judgment experiment

We used a pair-of-pairs task to characterize perceptual similarity. Volunteers were presented with two vertically offset pairs of faces on a standard LCD monitor under free viewing conditions and judged which pair were relatively more dissimilar with a button press on a USB keyboard (two-alternative force choice). Each face rotated continuously between a leftward and a rightward viewpoint (45 degrees left to 45 degrees right over 3 seconds). Ratings across all possible pairings of face pairs (2145 trials: all pairings of the 66 possible pairs of the 12 faces) were combined into a distance matrix for each volunteer where each entry reflects the percentage of trials on which that face pair was rated as relatively more dissimilar.

Functional MRI experiment

We measured brain response patterns evoked by faces in a rapid event-related design. Volunteers fixated on a central point of the screen where a pseudo-random sequence of face animations appeared (7 degrees visual angle in height, 2s on, 1s fixation interval). We verified fixation accuracy online and offline using an infrared eye tracking system (Sensomotoric Instruments, 50Hz monocular acquisition). The faces rotated leftward and rightward on separate trials (18-45 degrees rotation), and volunteers responded with a button press to occasional face repetitions regardless of rotation (one-back task). This served to encourage attention to facial identity rather than to incidental low-level physical features. Consistent with a task strategy based on identity recognition rather than image matching, volunteers were sensitive to repetitions both within viewpoint (mean d'+-1 standard deviation 2.68+-0.62) and across viewpoint (2.39+-0.52).

The experiment was divided into 16 runs and each run comprised 156 trials bookended by 10s fixation intervals. The trial order in each run was first-order counterbalanced over the 12 faces using a De Bruijn sequence (41) with 1 additional repetition (diagonal entries in transfer matrix) added to each face in order to make the one-back repetition task more engaging and to increase design efficiency (39). The viewpoint in which each face appeared was randomized separately, since a full 24-stimulus De Bruijn sequence would have been over-long (576 trials), and the analyses we report concern identity rather than viewpoint effects. Although the resulting 24-stimulus sequences were not fully counter-balanced, we used an iterative procedure to minimize any resulting inhomogeneity by rejecting viewpoint randomizations that generated off-diagonal values other than 0 and 1 in the 24-condition transfer matrix (that is, each possible stimulus-to-stimulus transfer in the sequence could appear only once or not at all). These homogeneous trial sequences served to enhance cross-validation performance by minimizing over-fitting to idiosyncratic trial sequence biases in particular runs. We modeled the data in each run with one predictor per face exemplar.

Magnetic resonance imaging acquisition

Functional and structural images were collected at the MRC Cognition and Brain Sciences Unit (Cambridge, UK) using a 3T Siemens Tim Trio system and a 32-channel head coil. Functional runs used a 3D echoplanar imaging sequence (2mm isotropic voxels, 30 axial slices, 192 x 192mm field of view, 128 x 128 matrix, TR=53ms, TE=30ms, 15° flip angle, effective acquisition time 1.06s per volume) with GRAPPA acceleration (acceleration factor 2 x 2, 40 x 40 PE lines). Each volunteer's functional dataset (7376 volumes) was converted to NIFTI format and realigned to the mean of the first session's first experimental run using standard functionality in SPM8 (fil.ion.ucl.ac.uk/spm/software/spm8/). A structural T1-weighted volume was collected in the first session using a multi-echo MPRAGE sequence (1mm isotropic voxels)(42). The structural image was de-noised using previously described methods (43), and the realigned functional dataset's header was co-registered with the header of the structural volume using SPM8 functionality. The structural image was then skull-stripped using the FSL brain extraction tool (fmrib.ox.ac.uk/fsl), and a re-sliced version of the resulting brain mask was applied to the fMRI dataset to remove artifacts from non-brain tissue. We constructed design matrices for each run of the experiment by convolving the onsets of experimental events with the SPM8 canonical hemodynamic response function. Slow temporal drifts in MR signal were removed by projecting out the contribution of a set of nuisance trend regressors (polynomials of degrees 0-4) from the design matrix and the fMRI data in each run.

Cross-validated discriminant analysis

We estimated the neural discriminability of each face pair for each region of interest using a cross-validated version of the Mahalanobis distance (44). This analysis improves on the related Fisher's linear discriminant classifier by providing a continuous metric of discriminability without ceiling effects. Similarly to the linear discriminant, classifier weights were estimated as the contrast between each condition pair multiplied by the inverse of the covariance matrix of the residual time courses, which was estimated using a sparse prior (45). This discriminant was estimated separately for the concatenated design matrix and fMRI data in each possible leave-one-run-out training split, and the resulting weights were projected onto the contrast estimates from each training split's corresponding test run (16 estimates per contrast). The absolute value of each resulting distance estimate was square root transformed and then returned to original sign before being averaged across the splits to obtain the final neural discriminability estimate for that volunteer and region. When the same data is used to estimate the discriminant and evaluate its performance, this algorithm returns the Mahalanobis distance provided that a full rather than sparse covariance estimator is used (44). However, unlike a true distance measure, the cross-validated version that we use here is centered on 0 under the null hypothesis. This motivates group-level inference for above-chance performance using conventional T tests.

We developed a variant of this discriminant analysis where effects that might be broadly described as univariate-level are removed (Figure S4). This control analysis involved two modifications to how contrasts were calculated at the level of forming the discriminant and at the level of evaluating the discriminant on independent data. First, the mean pattern across voxels was subtracted from each parameter estimate in order to remove any region-mean offsets between the conditions. Second, the mean response pattern across pairs of conditions was subtracted from each member of the pair before calculating the contrast estimate, in order to correct for any scaling effects. The resulting control analysis is insensitive to patterns that differ in simple additive offsets as well as cases where a single pattern is multiplicatively scaled between the conditions.

Multiple regression RSA

We used a multiple regression model to estimate the relative contribution of eccentricity and direction to cortical and perceptual face-space representations. Multiple regression fits to distance estimates can be performed using squared values, since squared distances sum according to the Pythagorean theorem. We partitioned the squared distances in the reference PCA space into variance associated with eccentricity changes by creating a distance matrix where each entry reflected the minimum distance for its eccentricity group in the squared reference PCA matrix (that is, cases along the group's diagonal where there was no direction change). The direction matrix was then constructed as the difference between the squared reference PCA matrix and the eccentricity matrix (Figure 2a). These predictors were vectorized and entered into a multiple regression model together with a constant term. The absolute values of the cortical and perceptual distance matrices were squared and then transformed back to their original sign before being regressed on the predictor matrix using ordinary least squares. Finally, the absolute values of the resulting parameter estimates were square-root transformed and returned to their original signs. These sign transforms served to preserve symmetry about zero under the null hypothesis for the resulting parameter estimates.

Functional regions of interest

We used a conventional block-based functional localizer experiment to identify category-selective and visually-responsive regions of interest in human visual cortex. Volunteers fixated a central cross on the screen while blocks of full-color images were presented (36 images per block presented with 222ms on, 222ms off, 16 s fixation). Volunteers were instructed to respond to exact image repetitions within the block. Each run comprised 3 blocks each of faces, scenes, objects and phase-scrambled versions of the scene images. Each volunteer's data (8 runs of 380 volumes) was smoothed with a Gaussian kernel (6mm full width at half maximum) and responses to each condition were estimated using standard SPM8 first level modeling. Regions of interest were masked in individual volunteers using statistical thresholds that yielded separable activated regions. We defined the face-selective occipital and fusiform face areas with the contrast of faces over objects, the scene-selective parahippocampal place area and transverse occipital sulcus as the contrast of scenes over objects, and the early visual cortex as the contrast of scrambled stimuli over the fixation baseline. We also attempted to localize a face-selective region in the posterior superior temporal sulcus, a face-selective region in anterior inferotemporal cortex and a scene-selective region in retrosplenial cortex, but do not report results for these regions here since they could only be identified in a minority of the volunteers.

Ramp model

The ramp model comprises 1000 model units, each of which exhibits a monotonically increasing response in a random direction extending from the origin of the face space (Figure 3). The response *y* at position *x* along the preferred direction is described by the sigmoid

y[raw] = 1 / (1+exp((-x+o)/s));

where the free parameters are *o*, which specifies the horizontal offset of the response function (zero places the midpoint of the response function at the norm of the space, values greater than zero corresponds to responses shifted away from the norm), and *s*, which defines the amount of response function saturation (4 corresponds to a near-linear response in the domain of the face exemplars used here, while values near zero correspond to a step-like increase in response). The raw output of each model unit is then translated toward the population-mean response

y[final] = y[raw]-y[mean] * (1-p) + y[mean]

where *p* defines the strength of readout-level population averaging (0 corresponds to no averaging, 1 corresponds to each model unit returning the population-mean response).

Exemplar model

The exemplar model comprises 1000 model units, each of which prefers a Cartesian coordinate in the face space with response fall-off captured by an isotropic Gaussian. The free parameters are *w*, which controls the full width at half-maximum tuning width of the Gaussian response function, and *d*, which controls the width of the Gaussian distribution of tuning centers (0.1 places Z=2.32 at 10% of the eccentricity of the caricatures while 3 places this tail at 300% of the eccentricity of the caricatures).

We generated an inverted-Gaussian model where the distribution of distances was inverted at Z=2.32 and negative distances truncated to zero (1% of exemplars). This model was fitted with similar parameters as the original Gaussian exemplar model.

Estimating the noise ceiling

We estimated the noise ceiling for Z-transformed Pearson correlation coefficients based on methods described previously (44). This method estimates the explained variance that is expected for the true model given noise levels in the data. Although the true noise level of the data cannot be estimated, it is possible to approximate its upper and lower bounds in order to produce a range within which the true noise ceiling resides. The lower bound estimate is obtained by a leave-one-volunteer-out cross-validation procedure where the mean distance estimates of the training split are correlated against the left-out-volunteer's distances, while the upper bound is obtained by performing the same procedure without splitting the data. These estimates were visualized as a shaded region in figures after reversing the Z-transform (Figure 4).

Statistical inference

All statistical inference was performed using T-tests at the group-average level (N=10 in all cases except the occipital face area, N=9). Correlation coefficients were Z-transformed prior to statistical testing. Averages were reverse-transformed to original units before visualization for illustrative purposes.

Control models based on low-level physical feature coding

We used two control models to estimate whether coding based on pixelwise features or low-level visual properties would produce the same face space warping we observed in our data (Figure 4). The pixelwise distance model was generated by stacking all the pixels in each of the face animations into vectors and estimating the correlation distance between these intensity values. The Gabor wavelet pyramid is a neuroscientificallyinspired model that has previously been used to successfully predict responses in the early visual cortex (46). The model is composed of a bank of Gabor filters varying in spatial position and orientation, and predictions are generated by measuring the filter outputs to each face exemplar image and estimating the correlation distance between these model response patterns. We estimated these distances using the last frame for each animation.

Acknowledgments

We are grateful to Jenna Parker for assistance with data collection and to Marta Correia for assistance with 3D EPI protocols.

References

1. Anzellotti S, Caramazza A. From parts to identity: Invariance and sensitivity of face representations to different face halves. Cereb Cortex. :1–10.

2. Anzellotti S, Fairhall SL, Caramazza A. Decoding representations of face identity that are tolerant to rotation. Cereb Cortex [Internet]. 2014 Mar 5 [cited 2013 Mar 6];24:1988–95. Available from: http://www.cercor.oxfordjournals.org/cgi/doi/10.1093/cercor/bht046

3. Axelrod V, Yovel G. Successful decoding of famous faces in the fusiform face area. PLoS One [Internet]. 2015;10:e0117126. Available from: http://dx.plos.org/10.1371/journal.pone.0117126

4. Goesaert E, Op de Beeck HP. Representations of facial identity information in the ventral visual stream investigated with multivoxel pattern analyses. J Neurosci [Internet]. 2013 May 8 [cited 2013 May 8];33(19):8549–58. Available from: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1829-12.2013

5. Kriegeskorte N, Formisano E, Sorger B, Goebel R. Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc Natl Acad Sci. 2007;104:20600–5.

6. Natu VS, Jiang F, Narvekar A, Keshvari S, Blanz V, O'Toole AJ. Dissociable neural patterns of facial identity across changes in viewpoint. J Cogn Neurosci. 2010;22(7):1570–82.

7. Nestor A, Plaut DC, Behrmann M. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. Proc Natl Acad Sci [Internet]. 2011 May 31 [cited 2011 Jun 1];108:9998–10003. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1102433108

8. Nestor A, Behrmann M, Plaut DC. The neural basis of visual word form processing: A multivariate investigation. Cereb Cortex [Internet]. 2013 Jun 12 [cited 2013 Jun 5];23:1673–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22693338

9. Gao X, Wilson HR. The neural representation of face space dimensions. Neuropsychologia [Internet]. Elsevier; 2013 Jul 10 [cited 2013 Aug 11];51(10):1787–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23850598

10. Verosky SC, Todorov A, Turk-Browne NB. Representations of individuals in ventral temporal cortex defined by faces and biographies. Neuropsychologia [Internet]. Elsevier; 2013 Jul 16 [cited 2013 Jul 23];51:2100–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23871881

11. Kriegeskorte N, Mur M, Bandettini PA. Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci. 2008;2:1–28.

12. Haxby J V, Hoffman E, Gobbini M. The distributed human neural system for face perception. Trends Cogn Sci. 2000;4:223–33.

13. Bruce V, Young AW. Understanding face recognition. Br J Psychol [Internet]. 1986 Aug [cited 2010 Sep 23];77 (Pt 3):305–27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3756376

14. Valentine T. A unified account of the effects of distinctiveness, inversion, and race in face recognition. Q J Exp Psychol [Internet]. 1991 [cited 2011 Jun 14];43A(2):161–204. Available from: http://www.informaworld.com/index/776369317.pdf

15. Blanz V, Vetter T. A morphable model for the synthesis of 3D faces. Proc 26th Annu Conf Comput Graph Interact Tech - SIGGRAPH '99 [Internet]. New York, New York, USA: ACM Press; 1999;187–94. Available from: http://portal.acm.org/citation.cfm?doid=311535.311556

16. O'Toole AJ, Abdi H, Deffenbacher KA, Valentin D. Low-dimensional representation of faces in higher dimensions of the face space. J Opt Soc Am A [Internet]. 1993 [cited 2012 Jun 14];10(3):405–15. Available from: http://www.opticsinfobase.org/abstract.cfm?id=4552

17. Ross DA, Hancock PJB, Lewis MB. Changing faces: Direction is important. Vis cogn [Internet]. 2010 Jan [cited 2012 May 30];18(1):67–81. Available from: http://www.tandfonline.com/doi/abs/10.1080/13506280802536656

18. Schulz C, Kaufmann JM, Walther L, Schweinberger SR. Effects of anticaricaturing vs. caricaturing and their neural correlates elucidate a role of shape for face learning. Neuropsychologia [Internet]. 2012 Jun 27 [cited 2012 Aug 1];50:2426–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22750120

19. Wilson HR, Loffler G, Wilkinson F. Synthetic faces, face cubes, and the geometry of face space. Vision Res [Internet]. 2002 Dec;42(27):2909–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12450502

20. Leopold DA, Bondar I V, Giese MA. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature [Internet]. 2006 Aug 3;442(7102):572–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16862123

21. Loffler G, Yourganov G, Wilkinson F, Wilson HR. fMRI evidence for the neural representation of faces. Nat Neurosci. 2005;10:1386–90.

22. Davidenko N, Remus D a., Grill-Spector K. Face-likeness and image variability drive responses in human face-selective ventral regions. Hum Brain Mapp [Internet]. 2012 Aug 5 [cited 2011 Aug 7];33:2334–49. Available from: http://doi.wiley.com/10.1002/hbm.21367

23. Said CP, Dotsch R, Todorov A. The amygdala and FFA track both social and non-social face dimensions. Neuropsychologia [Internet]. Elsevier Ltd; 2010 Aug [cited 2010 Sep 7];48(12):3596–605. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20727365

24. Heeger D. Normalization of cell responses in cat striate cortex [Internet]. Visual Neuroscience. 1992. p. 181–97. Available from: http://journals.cambridge.org/abstract_S0952523800009640\nhttp://www.ncbi.nlm.nih.gov/pubmed/1504027

25. Carandini M, Heeger DJ, Movshon JA. Linearity and normalization in simple cells of the macaque primary visual cortex. J Neurosci. 1997;17(21):8621–44.

26. Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen B a, et al. Do we know what the early visual system does? J Neurosci [Internet]. 2005 Nov 16 [cited 2012 Oct 26];25(46):10577–97. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16291931

27. Paysan P, Knothe R, Amberg B, Romdhani S, Vetter T. A 3D Face Model for Pose and Illumination Invariant Face Recognition. 2009 Sixth IEEE Int Conf Adv Video Signal Based Surveill [Internet]. Ieee; 2009 Sep;296–301. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5279762

28. Pasupathy A, Connor CE. Shape Representation in Area V4 : Position-Specific Tuning for Boundary Conformation. J Neurophysiol. 2001;86:2505–19.

29. Freiwald WA, Tsao DY, Livingstone M. A face feature space in the macaque temporal lobe. Nat Neurosci. 2009;12:1187.

30. Kriegeskorte N, Bandettini PA, Cusack R. How does an fMRI voxel sample the neuronal activity pattern: Compact-kernel or complex-spatiotemporal filter? Neuroimage. 2009;49(3):1965–76.

31. Parkes LM, Schwarzbach J V., Bouts A a., Deckers RHR, Pullens P, Kerskens CM, et al. Quantifying the spatial resolution of the gradient echo and spin echo BOLD response at 3 Tesla. Magn Reson Med. 2005;54(6):1465–72.

32. Mack ML, Preston AR, Love BC. Decoding the brain's algorithm for categorization from its neural implementation. Curr Biol [Internet]. Elsevier Ltd; 2013;23(20):2023–7. Available from: http://dx.doi.org/10.1016/j.cub.2013.08.035

33. Hara Y, Pestilli F, Gardner JL. Differing effects of attention in single-units and populations are well predicted by heterogeneous tuning and the normalization model of attention. Front Comput … [Internet]. 2014 [cited 2014 Feb 28];8:1–13. Available from: http://www.frontiersin.org/Journal/10.3389/fncom.2014.00012/abstract

34. Daugman JG. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J Opt Soc Am A. 1985;2(7):1160–9.

35. Marcelja S. Mathematical description of the responses of simple cortical cells. J Opt Soc Am. 1980;70(11):1297–300.

36. Cardoso MMB, Sirotin YB, Lima B, Glushenkova E, Das A. The neuroimaging signal is a linear sum of neurally distinct stimulus- and task-related components. Nat

Neurosci [Internet]. Nature Publishing Group; 2012;15(9):1298–306. Available from: http://dx.doi.org/10.1038/nn.3170

37. Goense J, Logothetis N. Neurophysiology of the BOLD fMRI signal in awake monkeys. Curr Biol. 2008;18:631–40.

38. Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann a. Neurophysiological investigation of the basis of the fMRI signal. Nature [Internet]. 2001 Jul 12;412(6843):150–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11449264

39. Henson RNA. Analysis of fMRI timeseries: Linear time-invariant models, event-related fMRI and optimal experimental design. In: Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Price CJ, editors. Human Brain Function. New York: Academic Press; 2003. p. 793–822.

40. Burton AM, Vokey JR. The face-space typicality paradox: Understanding the face-space metaphor. Q J Exp Psychol [Internet]. 1998 [cited 2012 May 31];3:475–83. Available from: http://www.tandfonline.com/doi/abs/10.1080/713755768

41. Aguirre GK, Mattar MG, Magis-Weinberg L. de Bruijn cycles for neural decoding. Neuroimage. Elsevier Inc.; 2011 Mar 24;56(3):1293–300.

42. Van der Kouwe a. JW, Benner T, Salat DH, Fischl B. Brain morphometry with multiecho MPRAGE. Neuroimage. 2008;40(2):559–69.

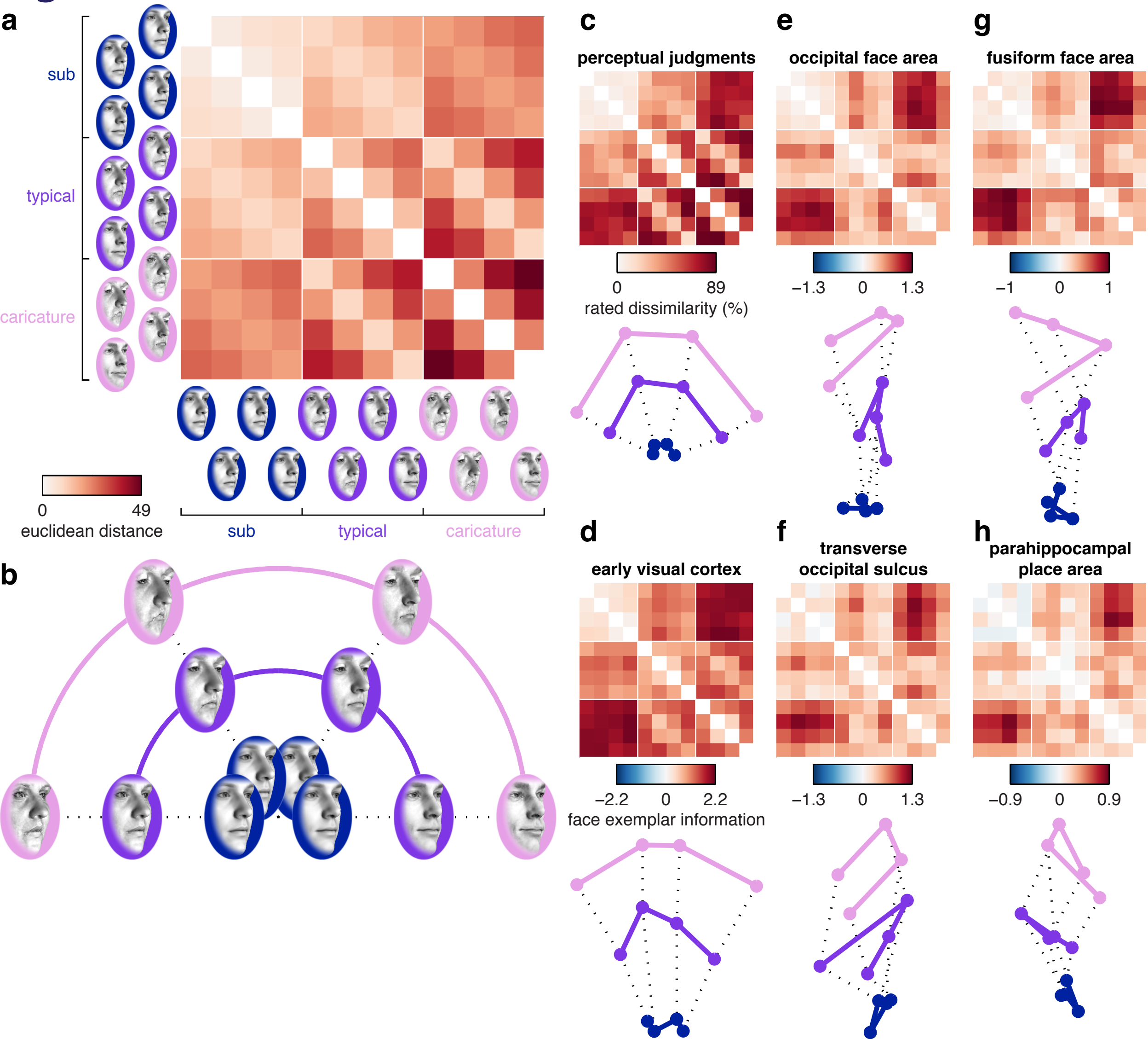
43. Manjón J V., Coupé P, Martí-Bonmatí L, Collins DL, Robles M. Adaptive non-local means denoising of MR images with spatially varying noise levels. J Magn Reson Imaging. 2010;31(1):192–203.

44. Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N. A toolbox for representational similarity analysis. PLoS Comput Biol. 2014;10:e1003553.

45. Misaki M, Kim Y, Bandettini PA, Kriegeskorte N. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. Neuroimage [Internet]. 2010 May [cited 2010 Aug 2];53:103–18. Available from: http://dx.doi.org/10.1016/j.neuroimage.2010.05.051

46. Kay KN, Naselaris T, Prenger RJ, Gallant JL. Identifying natural images from human brain activity. Nature. 2008;452:352–5.

Figure 1



a
sub
typical
caricature
0
49
euclidean distance
sub
typical
caricature
b
c
perceptual judgments
0
89
rated dissimilarity (%)
d
early visual cortex
−2.2
0
2.2
face exemplar information
e
occipital face area
−1.3
0
1.3
f
transverse
occipital sulcus
−1.3
0
1.3
g
fusiform face area
−1
0
1
h
parahippocampal
place area
−0.9
0
0.9

Figure 2

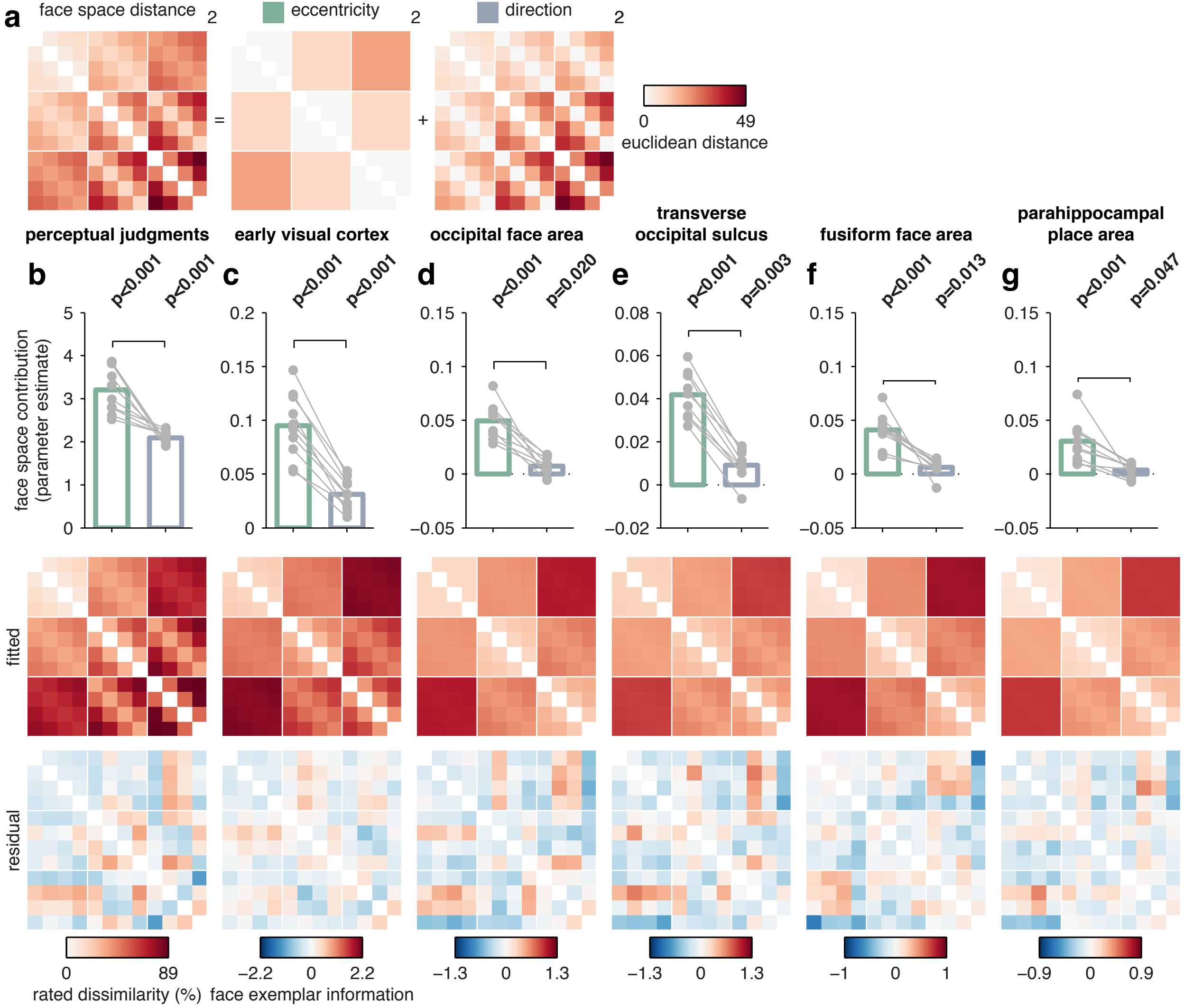

a
face space distance
2
=
eccentricity
2
+
direction
2
0
49
euclidean distance
perceptual judgments
early visual cortex
occipital face area
transverse occipital sulcus
fusiform face area
parahippocampal place area
b
p<0.001
p<0.001
c
p<0.001
p<0.001
d
p<0.001
p=0.020
e
p<0.001
p=0.003
f
p<0.001
p=0.013
g
p<0.001
p=0.047
face space contribution (parameter estimate)
5
4
3
2
1
0
0.2
0.15
0.1
0.05
0
0.15
0.1
0.05
0
−0.05
0.08
0.06
0.04
0.02
0
−0.02
0.15
0.1
0.05
0
−0.05
0.15
0.1
0.05
0
−0.05
fitted
residual
0
89
rated dissimilarity (%)
−2.2
0
2.2
face exemplar information
−1.3
0
1.3
−1.3
0
1.3
−1
0
1
−0.9
0
0.9

Figure 3

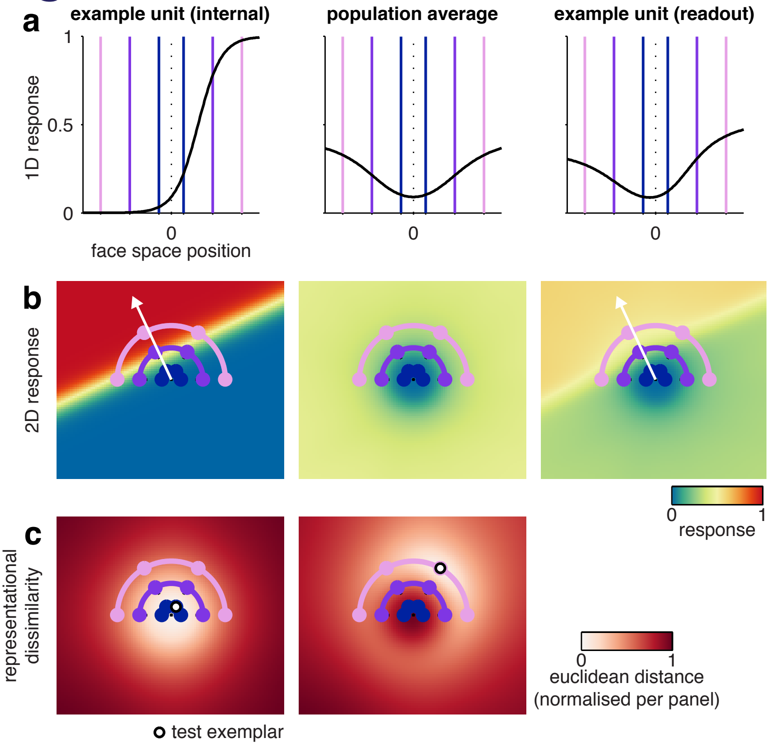

a
example unit (internal)
population average
example unit (readout)
1
0.5
0
1D response
0
face space position
0
0
b
2D response
0
1
response
c
representational
dissimilarity
0
1
euclidean distance
(normalised per panel)
test exemplar

Figure 4

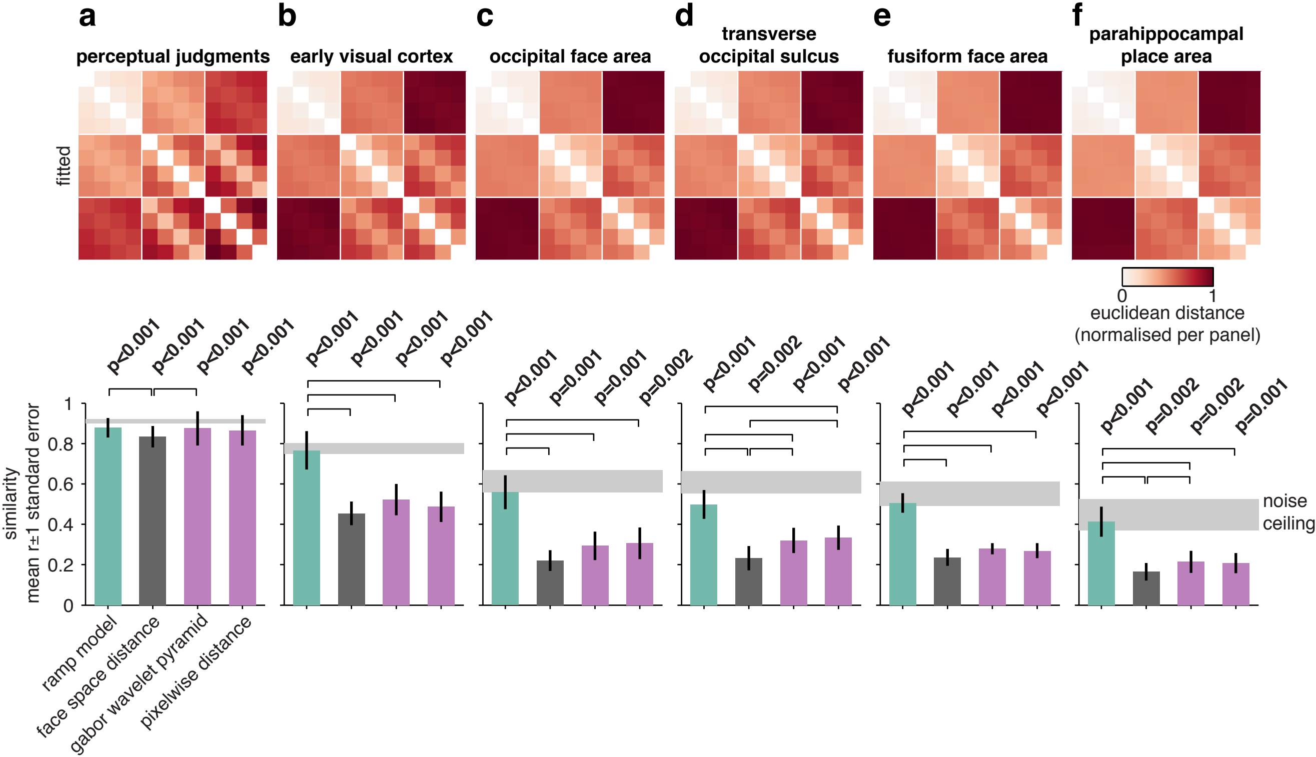
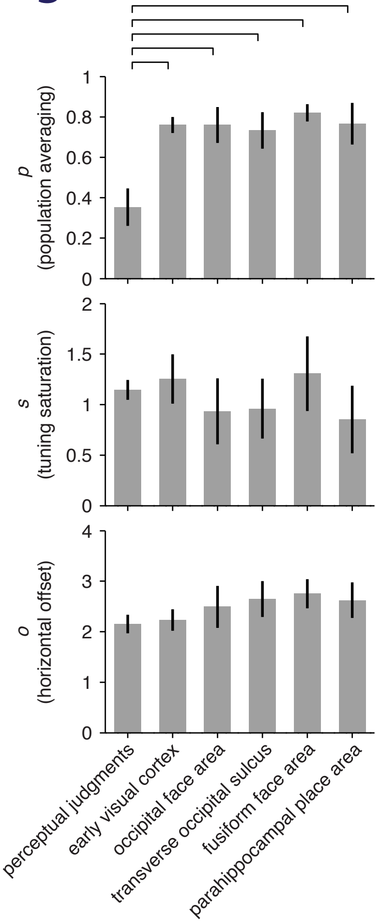
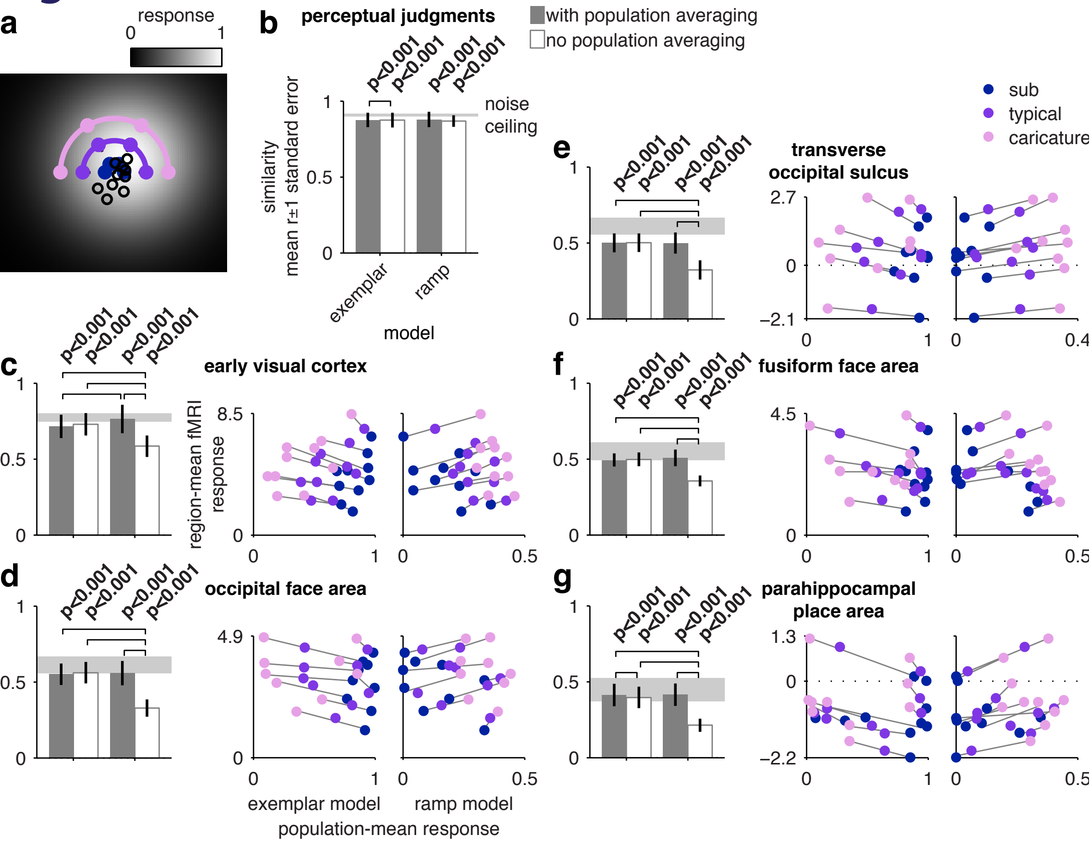

a
perceptual judgments
b
early visual cortex
c
occipital face area
d
transverse occipital sulcus
e
fusiform face area
f
parahippocampal place area
fitted
0
1
euclidean distance (normalised per panel)
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p=0.001
p=0.001
p=0.002
p<0.001
p=0.002
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p=0.002
p=0.002
p=0.001
similarity
mean r±1 standard error
0
0.2
0.4
0.6
0.8
1
noise ceiling
ramp model
face space distance
gabor wavelet pyramid
pixelwise distance

Figure 5



p
(population averaging)
0
0.2
0.4
0.6
0.8
1
s
(tuning saturation)
0
0.5
1
1.5
2
o
(horizontal offset)
0
1
2
3
4
perceptual judgments
early visual cortex
occipital face area
transverse occipital sulcus
fusiform face area
parahippocampal place area

Figure 6


a
response
0
1
b
perceptual judgments
with population averaging
no population averaging
p<0.001
p<0.001
p<0.001
p<0.001
1
0.5
0
noise
ceiling
similarity
mean r±1 standard error
exemplar
ramp
model
sub
typical
caricature
e
p<0.001
p<0.001
p<0.001
p<0.001
1
0.5
0
transverse
occipital sulcus
2.7
0
−2.1
0
1
0
0.4
c
p<0.001
p<0.001
p<0.001
p<0.001
1
0.5
0
early visual cortex
region-mean fMRI
response
8.5
0
0
1
0
0.5
f
p<0.001
p<0.001
p<0.001
p<0.001
1
0.5
0
fusiform face area
4.5
0
0
1
0
0.5
d
p<0.001
p<0.001
p<0.001
p<0.001
1
0.5
0
occipital face area
4.9
0
0
1
0
0.5
exemplar model
ramp model
population-mean response
g
p<0.001
p<0.001
p<0.001
p<0.001
1
0.5
0
parahippocampal
place area
1.3
0
−2.2
0
1
0
0.5

Supporting information captions

Figure S1. Example stimulus sets for 4 volunteers. Each stimulus set shares the same underlying distance matrix in the reference PCA space, while the randomization of the orientation of the plane on which the faces are sampled ensures that each set is visually distinct.

Figure S2. Direction discriminability as a function of eccentricity level. Each point reflects the mean performance for all directions at a given eccentricity level (4x4 block diagonals in Figure 1) for a single volunteer. Small random offsets have been added to each x coordinate for illustrative purposes. (**a**) Judged dissimilarity increases with eccentricity level. The black line indicates the least-squares fit. The p value for the slope estimate is obtained by group analysis of single volunteer slope estimates (n=10). (**b-f**) Cortical discrimination performance increases with eccentricity level in all regions except the parahippocampal place area (f). The p values above each x coordinate indicate significance at the group level, while markers denoting non-significant single volunteer discriminants are filled with gray (p>0.05, permutation test). The slope is plotted as in a.

Figure S3. Region-mean responses as a function of eccentricity level, estimated by a regression fit to the mean time course for each region. Each point reflects the mean parameter estimate for the 4 faces at a given eccentricity level for a single volunteer. Statistical inference is as Figure S2.

Figure S4. Effects of region-mean effect removal on multiple regression RSA fits. Effects are plotted as in Figure 2.

Figure S5. Example outputs of grid search procedure. Sections centered on the white crosshairs are plotted through the full 3-dimensional search space. The color map indicates the similarity between the ramp model distance matrix at that parameter value and the reconstructed face spaces for perceptual judgments (a) and the fusiform face area (b). It can be seen that both spaces are best fitted with similar parameters for the shape of the sigmoidal response function (tuning saturation, horizontal offset), but distinct levels of population averaging.

Figure S6. Fitting exemplar models with Gaussian and inverted-Gaussian distributions of tuning centers to cortical face space reconstructions. (**a**) Response profile for an example unit from an exemplar model with inverted-Gaussian unit-center distribution based on the group-average parameters that were optimal for predicting the FFA. The filled circle indicates the coordinate of the example unit's peak response. Tuning centers from a subset of other model units are overlaid in black outlines for illustrative purposes along with the face exemplars (plotted as in Figure 1). See Figure 6 for example unit from the standard Gaussian exemplar model. (**b-g**) Prediction performance for exemplar models fitted to the cortical and perceptual face space reconstructions. Gray bars indicate performance with population averaging, white bars performance with no averaging. These cross-validated estimates were calculated as in Figure 4. It can be seen that the inverted Gaussian is consistently outperformed by the standard Gaussian model.

Figure S1

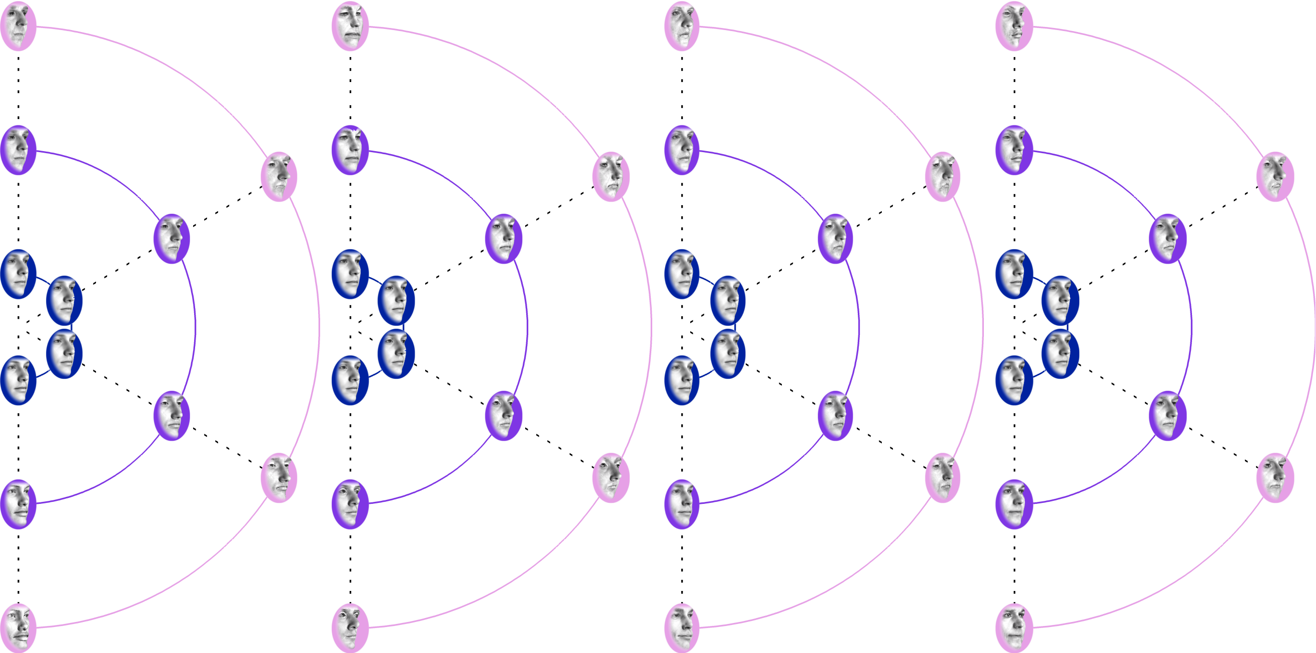


Figure S2

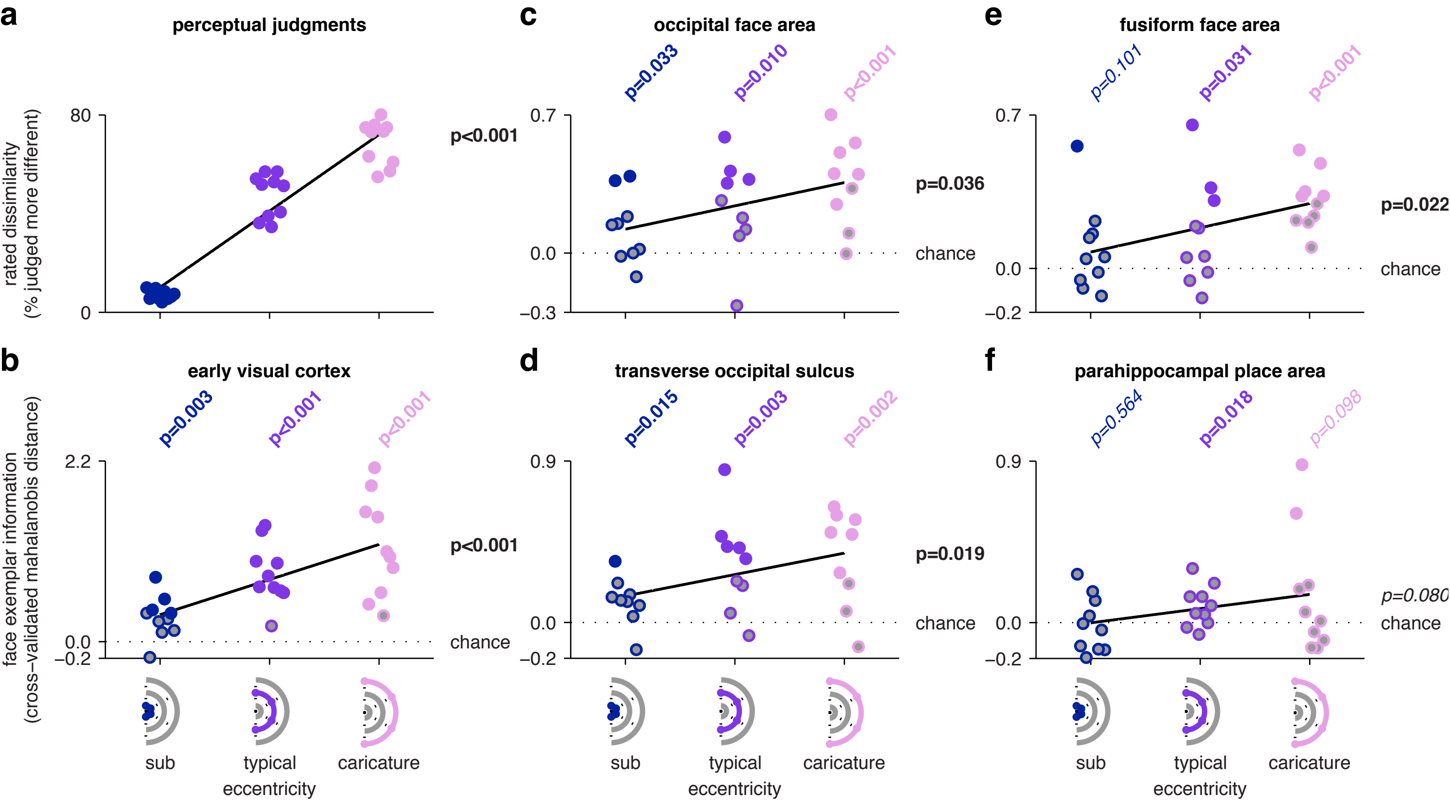

a
perceptual judgments
80
0
rated dissimilarity
(% judged more different)
p<0.001
c
occipital face area
p=0.033
p=0.010
p<0.001
0.7
0.0
−0.3
p=0.036
chance
e
fusiform face area
p=0.101
p=0.031
p<0.001
0.7
0.0
−0.2
p=0.022
chance
b
early visual cortex
p=0.003
p<0.001
p<0.001
2.2
0.0
−0.2
face exemplar information
(cross−validated mahalanobis distance)
p<0.001
chance
d
transverse occipital sulcus
p=0.015
p=0.003
p=0.002
0.9
0.0
−0.2
p=0.019
chance
f
parahippocampal place area
p=0.564
p=0.018
p=0.098
0.9
0.0
−0.2
p=0.080
chance
sub
typical
eccentricity
caricature
sub
typical
eccentricity
caricature
sub
typical
eccentricity
caricature

Figure S3

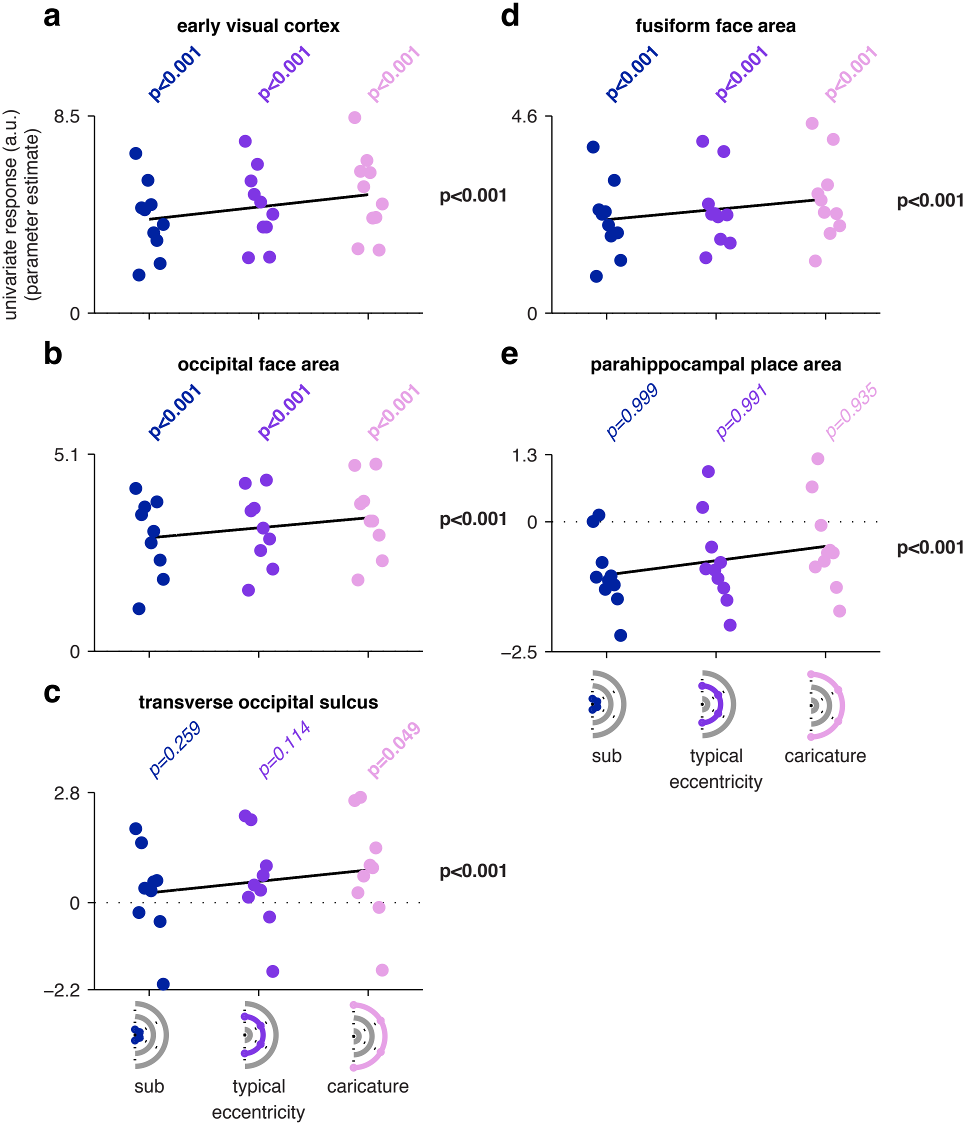

a
early visual cortex
p<0.001
p<0.001
p<0.001
8.5
0
univariate response (a.u.)
(parameter estimate)
p<0.001
d
fusiform face area
p<0.001
p<0.001
p<0.001
4.6
0
p<0.001
b
occipital face area
p<0.001
p<0.001
p<0.001
5.1
0
p<0.001
e
parahippocampal place area
p=0.999
p=0.991
p=0.935
1.3
0
–2.5
p<0.001
c
transverse occipital sulcus
p=0.259
p=0.114
p=0.049
2.8
0
–2.2
p<0.001
sub
typical
eccentricity
caricature
sub
typical
eccentricity
caricature

Figure S4

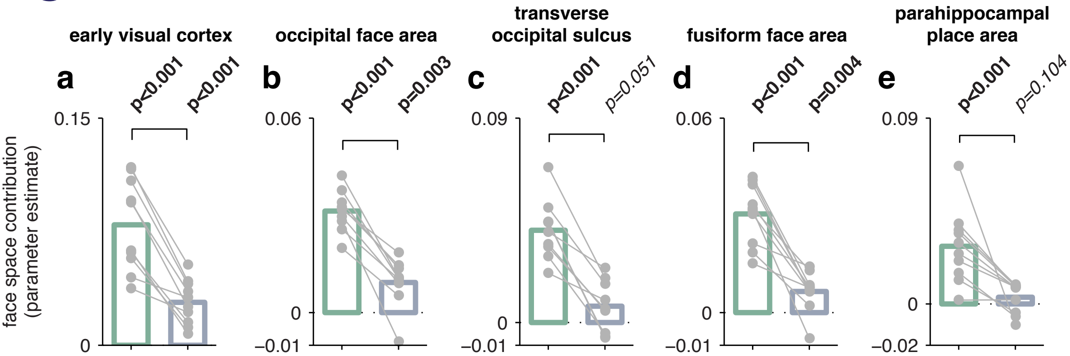

early visual cortex
occipital face area
transverse occipital sulcus
fusiform face area
parahippocampal place area
a
b
c
d
e
p<0.001
p<0.001
p<0.001
p=0.003
p<0.001
p=0.051
p<0.001
p=0.004
p<0.001
p=0.104
face space contribution (parameter estimate)
0.15
0
0.06
0
−0.01
0.09
0
−0.01
0.06
0
−0.01
0.09
0
−0.02

Figure S5

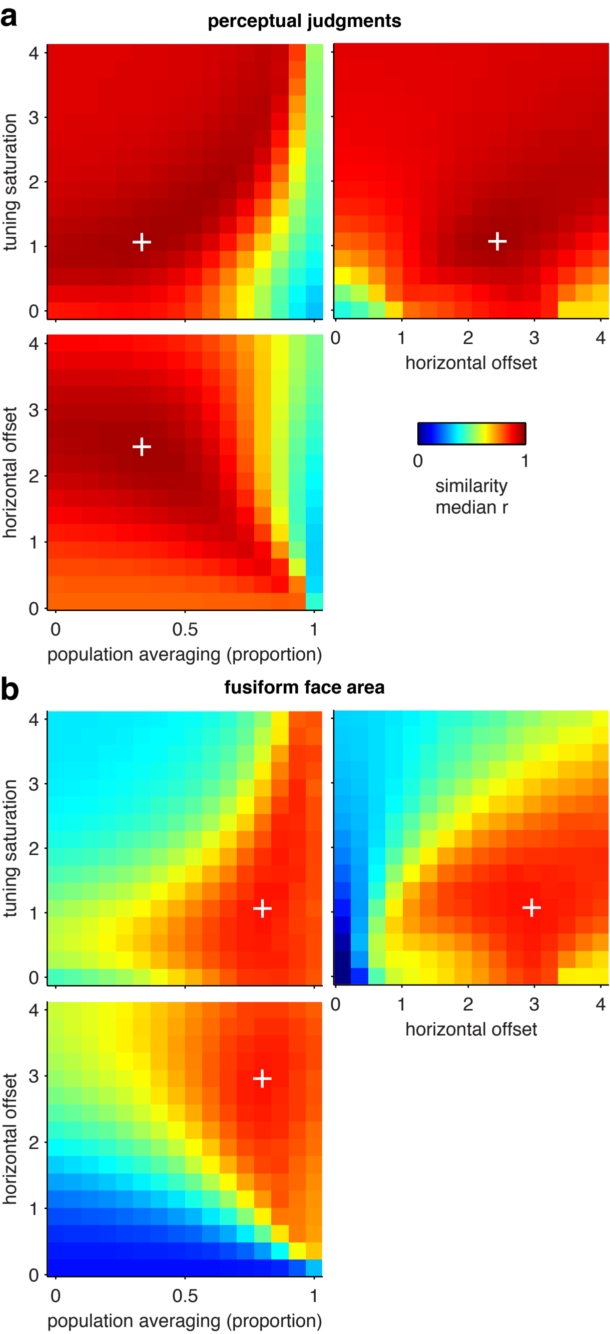
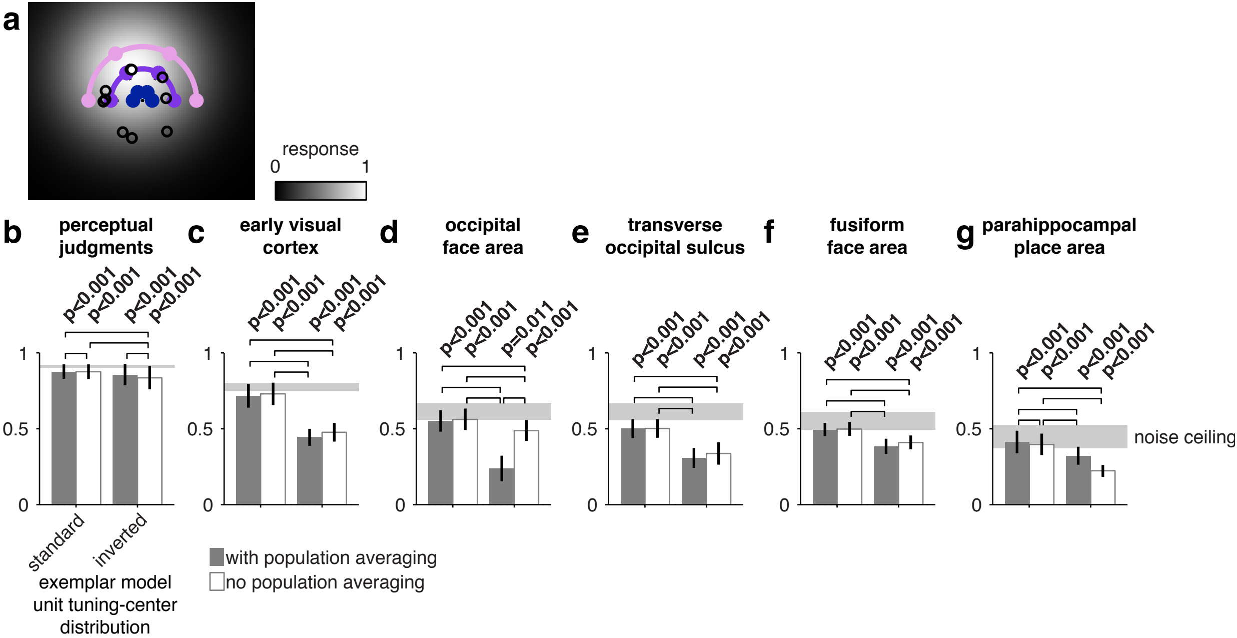

a
perceptual judgments
tuning saturation
4
3
2
1
0
horizontal offset
4
3
2
1
0
0
0.5
1
population averaging (proportion)
0
1
2
3
4
horizontal offset
0
1
similarity
median r
b
fusiform face area
tuning saturation
4
3
2
1
0
horizontal offset
4
3
2
1
0
0
0.5
1
population averaging (proportion)
0
1
2
3
4
horizontal offset

Figure S6


a
response
0
1
b
perceptual judgments
c
early visual cortex
d
occipital face area
e
transverse occipital sulcus
f
fusiform face area
g
parahippocampal place area
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p=0.011
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
p<0.001
similarity
mean r±1 standard error
1
0.5
0
standard
inverted
exemplar model unit tuning-center distribution
■ with population averaging
□ no population averaging
noise ceiling