

# A Unifying Model for Discordant and Concordant Results in Human Neuroimaging Studies of Facial Viewpoint Selectivity

 Cambria Revsine,<sup>1,2</sup>  Javier Gonzalez-Castillo,<sup>3</sup>  Elisha P. Merriam,<sup>1</sup>  Peter A. Bandettini,<sup>3,4</sup> and  Fernando M. Ramírez<sup>1,3</sup>

<sup>1</sup>Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, <sup>2</sup>Department of Psychology, University of Chicago, Chicago, Illinois 60637, <sup>3</sup>Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, and <sup>4</sup>Functional MRI Core Facility, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892

Recognizing faces regardless of their viewpoint is critical for social interactions. Traditional theories hold that view-selective early visual representations gradually become tolerant to viewpoint changes along the ventral visual hierarchy. Newer theories, based on single-neuron monkey electrophysiological recordings, suggest a three-stage architecture including an intermediate face-selective patch abruptly achieving invariance to mirror-symmetric face views. Human studies combining neuroimaging and multivariate pattern analysis (MVPA) have provided convergent evidence of view selectivity in early visual areas. However, contradictory conclusions have been reached concerning the existence in humans of a mirror-symmetric representation like that observed in macaques. We believe these contradictions arise from low-level stimulus confounds and data analysis choices. To probe for low-level confounds, we analyzed images from two face databases. Analyses of image luminance and contrast revealed biases across face views described by even polynomials—i.e., mirror-symmetric. To explain major trends across neuroimaging studies, we constructed a network model incorporating three constraints: cortical magnification, convergent feedforward projections, and interhemispheric connections. Given the identified low-level biases, we show that a gradual increase of interhemispheric connections across network-layers is sufficient to replicate view-tuning in early processing stages and mirror-symmetry in later stages. Data analysis decisions—pattern dissimilarity measure and data recentering—accounted for the inconsistent observation of mirror-symmetry across prior studies. Pattern analyses of human fMRI data (of either sex) revealed biases compatible with our model. The model provides a unifying explanation of MVPA studies of viewpoint selectivity and suggests observations of mirror-symmetry originate from ineffectively normalized signal imbalances across different face views.

**Key words:** face recognition; fMRI; MVPA; RSA; symmetry; viewpoint

## Significance Statement

The recognition of identity regardless of viewpoint is critical for social interactions. In primates, the representation of mirror-symmetric face views is thought to be a key intermediate processing step leading from strictly view-tuned to viewpoint-invariant representations. Human neuroimaging studies, however, have reached contradictory conclusions regarding the representation of viewpoint information in face-selective areas, despite being concordant in early visual areas. We show that stimulus confounds and data analysis choices explain these contradictory observations. We propose a network model that replicates observations of view-tuning in early processing stages regardless of analysis choices. The variable observation of mirror-symmetry in later stages is explained by choice of pattern dissimilarity measure and data recentering. Analyses of fMRI data confirmed biases compatible with our model.

Received Feb. 14, 2023; revised Feb. 6, 2024; accepted Feb. 15, 2024.

Author contributions: F.M.R. designed research; C.R. and F.M.R. performed research; C.R. and F.M.R. contributed unpublished reagents/analytic tools; C.R. and F.M.R. analyzed data; F.M.R., C.R., J.G.-C., E.P.M., and P.A.B. wrote the paper.

We thank Charles Zheng, Gang Chen, and Daniel Handwerker for helpful feedback and suggestions. This work utilized the computational resources of the National Institutes of Health HPC Biowulf cluster (<http://hpc.nih.gov>) and was supported by the Intramural Research Program at National Institute of Mental Health (ZIAMH002783 and ZIAMH002966).

The authors declare no competing financial interests.

Correspondence should be addressed to Fernando M. Ramírez at [fernando.ramirez@nih.gov](mailto:fernando.ramirez@nih.gov), [toporam@gmail.com](mailto:toporam@gmail.com).

<https://doi.org/10.1523/JNEUROSCI.0296-23.2024>

Copyright © 2024 the authors

## Introduction

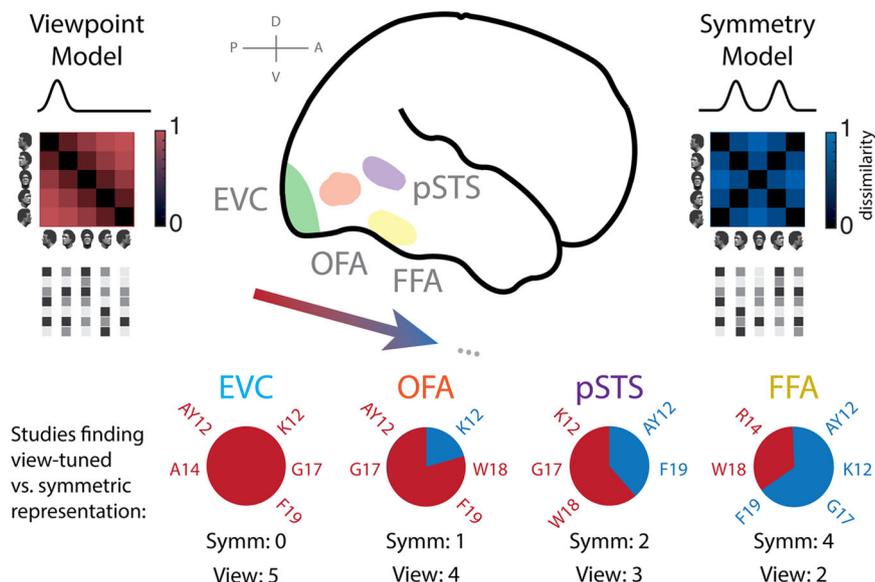
Our ability to recognize faces regardless of viewpoint remains a topic of intensive research. From a computational perspective (Marr, 1982), viewpoint-invariant recognition is challenging due to vast variation among images corresponding to the same identity. Traditional hierarchical theories hold that object recognition is achieved in a gradual stagewise process along the ventral visual stream (Grill-Spector and Malach, 2004). Theory also suggests that exploiting object symmetries, such as bilateral symmetry of the head, reduces the complexity of this task (Poggio and Anselmi, 2016). Current research aims to understand how object symmetries are represented in the brain to aid recognition.

In macaques, face recognition is thought to be supported by a hierarchically organized system of face-selective patches along the ventral visual pathway (Moeller et al., 2008; Tsao et al., 2008). Freiwald and Tsao (2010) characterized the tuning properties of neurons recorded from three stages of this hierarchy. Neurons in the middle lateral/fundus face patch (ML/MF) maximally responded to faces in one preferred view. In contrast, anterior lateral (AL) neurons exhibited bimodal tuning functions maximally responsive to mirror-symmetric face views. Finally, the anteromedial face patch (AM) exhibited virtual viewpoint invariance. It was proposed that AL is an intermediate mirror-symmetrically tuned processing stage along the face-processing hierarchy.

Several studies attempted to characterize the form of view-tuning of neural populations in human face-selective areas with functional magnetic resonance imaging (fMRI) and MVPA. These studies measured blood oxygen level-dependent (BOLD) responses in the occipital face area (OFA), fusiform face area (FFA), posterior superior temporal sulcus (pSTS), and early visual cortex (EVC) and characterized them as predominantly view-tuned or mirror-symmetric (Fig. 1). While all studies

in Table 1 observed view-tuning in EVC, conclusions in increasingly anterior face-selective areas proved increasingly harder to reconcile. For example, while some studies concluded that neurons in human FFA are tuned to a single preferred view, like the macaque ML/MF face patches, others concluded that FFA is mirror-symmetrically tuned, like the AL face patch.

How might such inconsistencies emerge between studies using similar stimuli and experimental designs? One possibility lies in different analysis methods, including the choice to demean the data and what dissimilarity measure to use when conducting representational similarity analysis (RSA; Kriegeskorte et al., 2008). We previously observed that studies reporting mirror-symmetry either demeaned the data prior to RSA using the correlation distance to measure pattern dissimilarities or relied on measures related to the Euclidean distance. We also noted that stronger responses for frontal than profile face views were observed in multiple studies (reviewed in Ramírez, 2018). Critically, regardless of the form of view-tuning of the underlying neurons, stronger responses for frontal face views will lead to “mirror-symmetry” with analysis pipelines that demean the data prior to RSA or use the Euclidean distance to measure dissimilarities (Ramírez, 2017). We recently also showed that stimuli from previous studies exhibit low-level confounds consistent with the reported overrepresentation of the frontal face views (Ramírez et al., 2020). Building on these observations, geometrical reasoning (Wickens, 2014) led us to hypothesize that cortical interhemispheric crossings might explain common and inconsistent trends observed across neuroimaging studies of viewpoint selectivity (Fig. 2). We reasoned that the influence of low-level stimulus features (e.g., luminance and contrast) in visual areas must depend on the strength of contra- and ipsilateral hemifield responses, and qualitatively different biases in signal strength across conditions along the visual hierarchy would differently interact with data analysis choices.

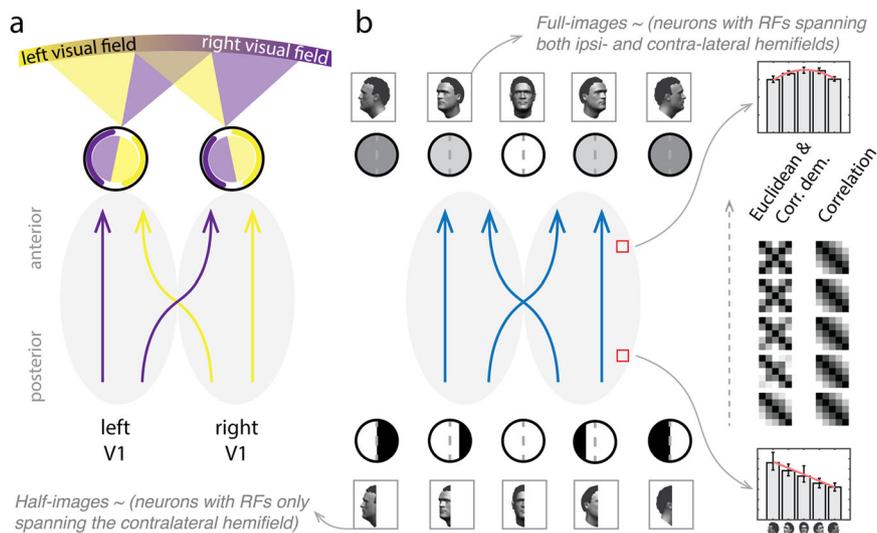


**Figure 1.** Commonalities and inconsistencies across fMRI-MVPA studies investigating viewpoint representations in humans. Four regions of interest are depicted on a sagittal view of the brain. In EVC (shown in green), 5/5 studies reported marked view-tuning, as depicted by the dissimilarity matrix shown in red (viewpoint model) and the unimodal neural tuning function shown immediately above. In OFA (in orange), 5/5 studies reported evidence of view-tuning. One study, however, reported additional evidence of some degree of mirror-symmetry, represented by the blue dissimilarity matrix (symmetry model) and bimodal tuning function immediately above. In pSTS (in purple), 5/5 studies reported evidence of view-tuning, with 2/5 observing some degree of mirror-symmetry. Finally, in the FFA (in yellow), 6/6 studies reported evidence of view-tuning, while 4/6 of these studies also reported evidence of mirror-symmetry, ranging from weak to strong. In sum, while marked view-tuning was consistently observed in posterior brain regions, mirror-symmetry was inconsistently observed, albeit with increasing frequency, in increasingly anterior areas along the ventral stream. OFA, occipital face area; pSTS, posterior superior temporal sulcus; FFA, fusiform face area.

**Table 1. Human fMRI-MVPA studies of viewpoint selectivity**

Study name	Abbreviation	Task design	Pattern estimation method	MVPA type	RSA distance measure	Data re-centering	ROI def. (bilateral/unilateral)	Areas exhibiting mirror-symmetric effects
Axelrod and Yovel, 2012 Exp. 1	AY12a	No fixation One-back, identity	None Concatenation of raw* intensity values for each condition within runs	Multiclass linear SVC, spatiotemporal patterns	Error rates	Yes a. Time series z-scored b. Patterns demeaned across voxels	Bilateral EVC, FFA; right OFA, pSTS	FFA, pSTS (not OFA)
Axelrod and Yovel, 2012 Exp. 2	AY12b	Yes fixation Color change detection of fixation spot	None Concatenation of raw* intensity values for each condition within runs	Multiclass linear SVC, spatiotemporal patterns	Error rates	Yes a. Time series z-scored b. Patterns demeaned across voxels	Bilateral EVC; right OFA, FFA	FFA (not OFA)
Kietzmann et al., 2012	K12	Yes fixation One-back, identity	Averaging for each block of percent signal changes of TRs 5–12 (10–24 s.)	RSA	Correlation	Yes a. Time series z-scored b. Regressed out mean spatial pattern	Bilateral EVC, OFA, FFA	OFA, FFA (not pSTS)
Anzellotti et al., 2014	A14	No fixation Identity detection	GLM with canonical HRF	1. RSA 2. Linear SVC	Correlation	No	Right EVC, OFA, Not tested FFA	
Ramírez et al., 2014	R14	Yes fixation Luminance change detection	GLM with canonical HRF	1. RSA 2. Linear SVC	a. Correlation b. Euclidean	No	Bilateral EVC; right OFA, pSTS, FFA	None
Guntupalli et al., 2017	G17	No fixation One-back, identity	GLM with canonical HRF	1. RSA 2. Linear SVC, spatiotemporal patterns	a. Correlation b. Classification accuracy	No	Bilateral EVC, OFA, FFA; right pSTS	SVC: none RSA: FFA (neither OFA nor pSTS)

SVC, support vector classification; GLM, general linear model; HRF, hemodynamic response function; EVC, early visual cortex; FFA, fusiform face area; OFA, occipital face area; pSTS, posterior superior temporal sulcus. Note that while time series z-scoring implies mean-centering across conditions, pattern demeaning across voxels does not. Regression of the mean spatial pattern, like mean centering, also changes the covariance structure of the data. \*fMRI data were motion corrected, spatially normalized, smoothed, and detrended.



**Figure 2.** Proposed account of commonalities and inconsistencies across fMRI-MVPA studies. **a**, Schematic of visual hemifields and their mapping onto cerebral hemispheres, axial view. Locations in left and right visual fields map onto primary visual cortex of the contralateral cerebral hemisphere. A known property of visual cortex is increasingly bilateral hemifield representations, due to interhemispheric connections, as one proceeds along the visual hierarchy. This property is central to the model proposed here. **b**, (Low-level image properties) + (interhemispheric crossings)  $\approx$  RSA results. Faces in different views exhibit different distributions of luminance and contrast. These properties have been found to exhibit symmetric distributions about the frontal view. Full circles (top row) indicate the mean luminance of the image of the face view shown immediately above. Anterior brain areas, which integrate input from both hemifields, are expected to exhibit quadratic (i.e., symmetric) biases of the form illustrated in the bar plot shown at the top right. In contrast, responses across views for half-images (bottom row) are expected to exhibit antisymmetric biases. Black and white circles indicate the rough luminance distribution of each half-image for each face view—note the dark hair and bright skin. For right V1 (this would imply a roughly linear (i.e., antisymmetric) trend if responses were proportional to the luminance of the left side of the stimulus (see bar plot at bottom right). If RSA outcomes reflect such trends for luminance and contrast, as we propose here, then dissimilarity matrices at earlier processing stages would exhibit marked view-tuning regardless of pattern dissimilarity measure. In turn, representations in later processing stages would exhibit mirror-symmetry either with the Euclidean distance or angular distances (e.g., correlation distance) only if the data are mean-centered across conditions prior to RSA. Instead, for RSA with the correlation distance, a viewpoint-specific representation would be expected throughout cortex, as shown by the dissimilarity matrices in the rightmost column.

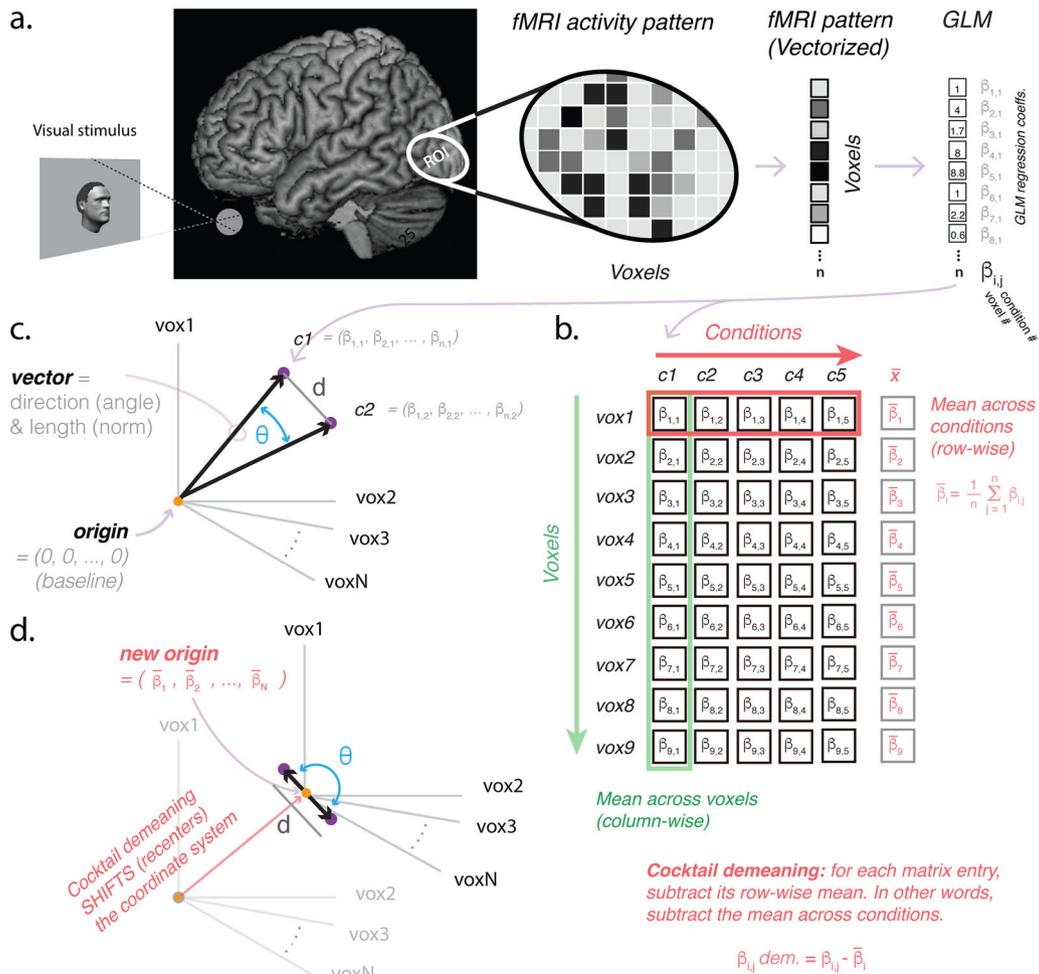
To test this hypothesis, we constructed a network model incorporating three biologically motivated constraints: convergent feed-forward projections, interhemispheric crossings increasing in frequency along the hierarchy, and foveal cortical magnification (CM). Connections between network units were otherwise randomly generated and hence not tuned to viewpoint information. We evaluated this minimally structured model in its capacity to reproduce observations across fMRI-MVPA studies when given as input images from two popular face databases. The model explained why conclusions in EVC are consistent across studies regardless of data recentering and RSA distance measure, while results in higher-tier areas are sensitive to these analysis choices.

## Materials and Methods

All simulations and statistical analyses were implemented using MATLAB (R2018b) and executed in Biowulf, the high-performance computing facility at NIH, as well as tested on a MacBook Pro (2021, 16 GB RAM) running MacOS (13.6.3). The implemented network architecture consisted of eight layers and considered exclusively feedforward

connections between units in adjacent layers. Two further constraints on the architecture were (1) CM of central image locations in the input layer and (2) a gradual increase in the number of interhemispheric projections in subsequent levels of the hierarchy. A preliminary version of this study was published in abstract form (Revsine et al., 2020).

In the simulations described here, when a specific instantiation of the model receives an image as input, it outputs a set of distributed response patterns—one per network layer. We obtained patterns associated with faces belonging to various identities, each presented in five viewpoints (see below, Input images for details). Response patterns in each network layer were then subjected to three variants of RSA. A first factor distinguishing these variants was the distance measure used to describe the dissimilarity between activity patterns. We computed dissimilarity matrices according to the Euclidean ( $RSA_{Euc}$ ) and the correlation distance ( $RSA_{corr}$ ). A second distinguishing factor was the choice whether to demean the simulated response patterns prior to computing pattern dissimilarities. By “demeaning” throughout this paper, we refer to the practice of subtracting in each measurement channel—for example, fMRI voxel—the mean activity level observed across conditions from that associated with each condition. This data transformation has been previously referred to as “cocktail demeaning,” or “cocktail blank removal” (Fig. 3).



**Figure 3.** Cocktail-mean subtraction changes correlations among brain patterns. **a**, fMRI spatial activity pattern for a visual stimulus in an ROI. The spatial pattern is also represented as a vector by concatenating the response magnitude in each voxel. Regression coefficients obtained from a GLM (see Materials and Methods) for one example experimental condition are shown to the right. **b**, Regression coefficients for nine voxels and five experimental conditions are shown arranged as a matrix. To the right, the row-wise mean (i.e., cocktail mean) is shown in red for each voxel. Demeaning the data across voxels (column-wise), shown in green, is not the problematic form of demeaning relevant here. Demeaning the data across conditions (i.e., cocktail demeaning), shown in red, is the problematic form of data recentering relevant here. **c**, Representation in N-dimensional space of fMRI pattern-vectors for two conditions (c1 and c2). The coordinates of the origin in this space are specified by the zero vector. The Euclidean distance,  $d$ , between the endpoint of these vectors is shown with a gray line, and the angle  $\theta$  subtended between c1 and c2 is shown in blue. **d**, Representation of the same experimental conditions after cocktail demeaning, which shifts the origin of the coordinate system. Critically, the angle between c1 and c2 has markedly changed after cocktail demeaning (compare  $\theta$  in **c**, **d**) and hence their correlation, since the latter is the cosine of the angle between c1 and c2 after zero-centering each vector (e.g., by row-wise demeaning, as shown in panel **b**). In contrast, note that the Euclidean distance between c1 and c2 remained unchanged after cocktail demeaning.

Throughout this paper, unless explicitly stated, by “demeaning” we refer to “cocktail demeaning.” Only RSA using the correlation distance was conducted both on native and demeaned activity patterns (see below, Pattern similarity analyses for details). We chose these three analysis variants because they reflect key differences in the analysis pipelines used by previous studies. To explore the impact of network connectivity density on pattern analyses, our model included a parameter that controlled the number of connections received by each network unit.

Below, we describe in further detail the (1) images provided as input to the model, (2) network architecture, (3) two model variants explored, and (4) three pattern analysis variants conducted on the distributed response patterns provided as output by the model.

*Input images*

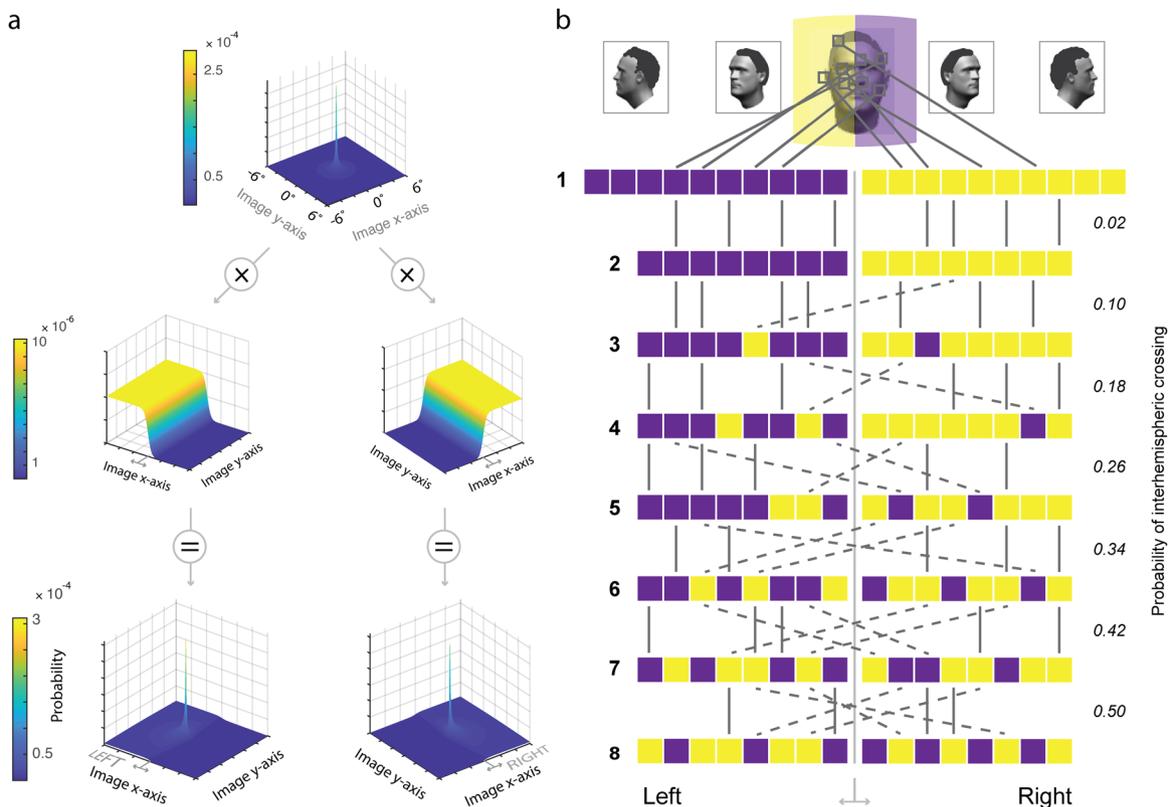
Images provided as input to the model were obtained from two popular databases: Karolinska Directed Emotional Faces (KDEF; Lundqvist et al., 1998) and Radboud Faces Database (RaFD; Langner et al., 2010). These databases have been used by previous studies investigating viewpoint-invariant face recognition and RaFD specifically used in studies combining fMRI and MVPA (Weibert et al., 2018; Flack et al., 2019). Both databases include multiple individuals photographed in the same five viewpoints:  $-90^\circ$  (right-profile view),  $-45^\circ$ ,  $0^\circ$ ,  $45^\circ$ , and  $90^\circ$  (left-profile view). The RaFD database contains images of 57, and the KDEF database of 70, adult individuals. Photographs of two individuals from the KDEF database were excluded because they exhibited lighting conditions inconsistent with the remaining identities (mean luminance  $>4$  SD away from the mean). Both databases include images of each identity displaying various

emotions and eye-gaze directions. Only faces with neutral expressions and consistent eye-gaze and head direction were provided as input to the model.

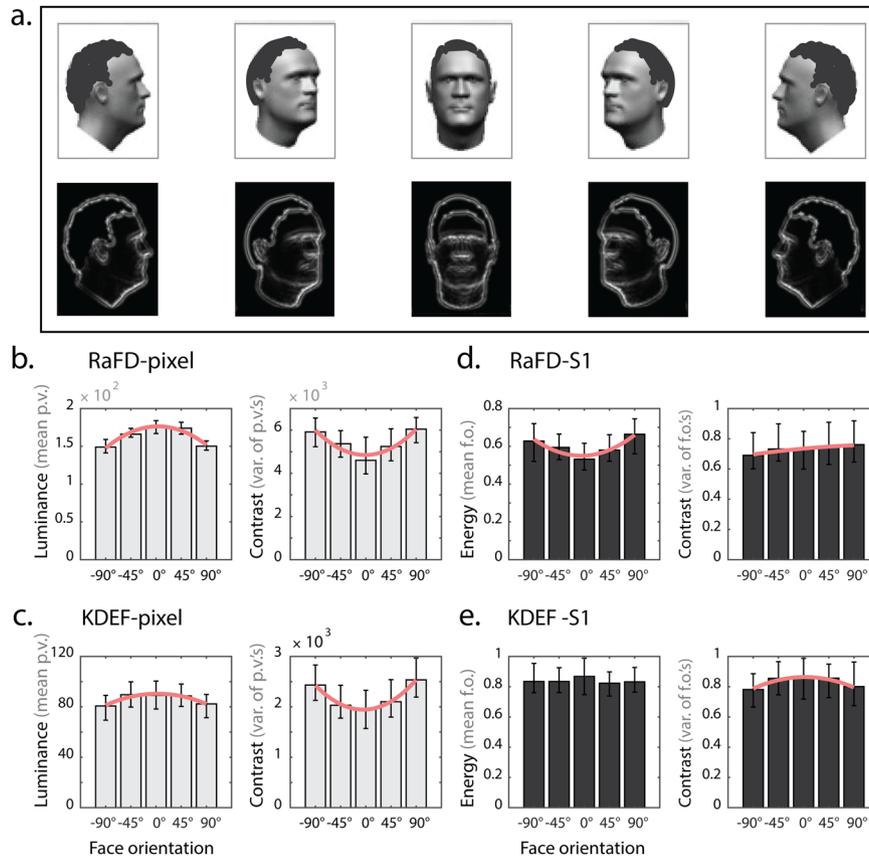
Images were converted from RGB to grayscale values (range, 0–255) and resized to ensure the height of the head measured on average  $8.7^\circ$  of visual angle, well within the range used in previous studies [range,  $4.2^\circ$ – $12.5^\circ$  (vertical dimension)]. Image size was specified with respect to the input array of the network model (441 pixels  $\times$  441 pixels), which spanned  $12.1^\circ \times 12.1^\circ$  of visual angle. The center of the fovea was located at the central pixel of the array. Images were centered and overlaid on a uniform gray background. An example face image is shown in Figure 5. Note that face images throughout this manuscript are used for illustration only, computer generated, and not photographs from the analyzed databases.

*Network architecture*

The network consists of eight hierarchically organized layers and is composed of two hemispheres. Each network layer consists of multiple units; 4,096 units in layer 1 (2,048 per hemisphere) and 1,024 units in each of the remaining layers (2–8; 512 per hemisphere). Units project exclusively onto units in the subsequent layer. Each network layer is intended to correspond to a different processing stage along the ventral visual stream. While early visual areas can be roughly mapped onto the first two or three layers of the network, high-level face-selective areas such as OFA and FFA are assumed to correspond to later layers, with OFA and pSTS presumably in a hierarchical level preceding FFA (but see Rossion et al., 2011). However, we do not suggest an explicit mapping between brain areas and layers of our network model.



**Figure 4.** Feedforward, randomly-connected, two-hemisphere network architecture. **a**, Probability distributions used to specify image locations sampled by layer 1 units. Top row, Distribution used to model CM of central image locations in V1. Middle row, Distributions used to model left-hemifield (LH) and right-hemifield (RH) representations. Bottom row, Product of CM with LH and RH distributions. These distributions served to specify, by random sampling, image locations providing input to units in each hemisphere of layer 1. **b**, Full network, consisting of eight layers: layer 1 (4,096 units) and layers 2–8 (each 1,024 units). Feedforward connections between units in consecutive layers define this architecture. Input to the left hemisphere of layer 1 (shown in purple) originates from RFs located almost exclusively on the right side of each image. The opposite is observed for the right hemisphere of layer 1 (in yellow). Ipsilateral projections within each network hemisphere are indicated by solid lines. Contralateral projections are indicated by dashed lines. The probability of a contralateral projection increases in steps of 0.08, beginning at 0.02 between layers 1 and 2, and reaching 0.5 between layers 7 and 8.



**Figure 5.** Distribution of mean luminance and contrast as a function of viewpoint for two face databases. **a**, Top row, Example face identity shown in five orientations ( $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ). Second row, Images above shown after pooling local orientation and frequency filters from the S1 layer of HMAX model. **b**, **c**, Median and interquartile range are shown for mean luminance (mean pixel value) or contrast (pixel variance) of face identities, always as a function of viewpoint. Bar plots in light gray correspond to pixel-level representation. **d**, **e**, Plots in black correspond to S1 level representation. Best fitting second-order polynomial is shown in red only if either the linear or quadratic regression coefficients are significantly different from zero at the population level and their associated absolute values also significantly different. p.v., pixel value; f.o., filter output. Please note that all face images shown throughout this paper were computer generated and used for illustration purposes only, they are not instances of photographs from the analyzed databases.

One prominent aspect of our network architecture is that it considers distinct left and right network hemispheres, emulating the structure of the brain (Figs. 2, 4). Connections between units in consecutive layers originate from either the ipsilateral or contralateral hemisphere of the previous layer. In successive layers of the network, connections are increasingly likely to cross hemispheres, reflecting the increasing probability of neural projections crossing through the corpus callosum and anterior commissure at successive stages of the visual hierarchy (Berlucchi, 2014). This aspect of the model is supported by a well-documented increase of receptive field (RF) sizes along the ventral visual hierarchy, such that RFs in higher-tier areas tend to be large, usually include the fovea, and cross the vertical meridian into the ipsilateral hemifield (Gross et al., 1972; Desimone et al., 1984; Rolls, 2012). Human fMRI measurements are consistent with this concept (Tootell et al., 1998; Hemond et al., 2007; Dumoulin and Wandell, 2008; Henriksson et al., 2008; Amano et al., 2009). The RF of each unit in layer 1 is defined as the collection of image locations that provide input to that unit. Units in the left hemisphere of layer 1 receive input almost entirely from pixel locations in the right hemifield of the image and vice versa for units in the right hemisphere. A second biological property incorporated by the model is CM of the foveal representation. Image locations toward the center of each image are sampled with markedly increased probabilities, following the CM function described in Duncan and Boynton (2003).

**Probability distributions: layer 1.** To specify the input received by each layer 1 network unit, we randomly sampled pixel locations from the  $441 \times 441$  array assumed in our model to represent input to the retinae. The exact number of pixel locations providing input to a particular

unit in layer 1 was specified by a parameter controlling the connectivity density of the network (see below, Density for details). For clarity, we will first explain the random process by which we specify the pixel locations providing input to each layer 1 unit, which comes almost exclusively from the contralateral hemifield. To this aim, we relied on two probability distributions—termed cortical magnification left hemifield (CM-LH) and cortical magnification right hemifield (CM-RH). Each of these distributions was tailored to reflect two properties of human primary visual cortex (V1), namely, (1) CM and (2) separate hemifield representations. CM-LH and CM-RH were obtained by combining, in each case, a distribution used to specify the preferred hemifield of a unit, with a second distribution used to model the desired overrepresentation of the fovea (Fig. 4a). CM-LH and CM-RH are mirror-symmetric versions of each other. Importantly, both distributions exhibit markedly increased probabilities at locations of the input array corresponding to the fovea. These probability distributions were used to specify the multinomial probability assignments according to which pixel locations were then randomly drawn to define the RF of each unit.

In more detail, as a first step, a two-dimensional distribution was specified using the following formula for the linear CM factor of V1 described in Duncan and Boynton (2003):

$$f(x, y) = 9.81 \times (\sqrt{x^2 + y^2})^{-0.83}. \quad (1)$$

In this formula,  $\sqrt{x^2 + y^2}$  denotes eccentricity from the origin in degrees of visual angle, which is in our model at the center of the fovea. Because the value of this function at the origin tends to infinity, we arbitrarily set this value to 100. The distribution was then normalized to ensure the sum

of discrete probabilities was equal to one. This distribution models heightened spatial resolution at the foveal confluence. Probabilities monotonically decrease as a function of distance from the fovea. We will refer to this distribution as the CM distribution (Fig. 4a, top row).

To model the existence of a left and right visual hemifield representation in V1, we then defined two 2D distributions, one corresponding to the left hemifield and the other to the right hemifield (Fig. 4a, middle row). To define these distributions, we relied, respectively, on the following logistic functions:

$$f(x, y) = L \times \frac{1}{1 + e^{-k(x-x_0)}}, \quad (2)$$

and

$$f(x, y) = L \left( 1 - \frac{1}{1 + e^{-k(x-x_0)}} \right). \quad (3)$$

In these functions,  $L$  specifies the maximum value of the curve,  $x_0$  the value of the midpoint along the  $x$ -axis, and  $k$  the logistic growth rate. Here,  $L$  was set to 10,  $x_0$  to 0, and  $k$  to 4. These values achieve a steep transition across the vertical meridian from low to high probabilities. Probability densities were maximally concentrated on the left hemifield for the left-hemifield distribution and on the right hemifield for the right-hemifield distribution. The output values of functions (2) and (3), when evaluated on any arbitrary pair of  $x$ - and  $y$ -values, are independent of the value of  $y$ , and hence the resulting probability distributions can be understood as a concatenation of 1D logistic distributions along the  $y$ -dimension (Fig. 4a, middle row).

Finally, the CM distribution and the left-hemifield logistic distributions were pointwise multiplied and normalized to ensure discrete probabilities summed to one. We term the resulting distribution CM-LH (Fig. 4a, bottom row, left). This distribution is characterized by probabilities maximally concentrated at the center of the input array, a property inherited from the CM distribution, and heightened probabilities in the left hemifield, inherited from the left-hemifield logistic distribution. In a separate but otherwise identical procedure, CM was pointwise multiplied by the right-hemifield logistic distribution and normalized to produce the CM-RH distribution (Fig. 4a, bottom row, right).

To explore the relevance of incorporating CM into the model, we tested an additional variant excluding this specific component. In this reduced model, only the left-hemisphere and right-hemisphere logistic distributions were used to specify input projections to units in layer 1.

**Network architecture: layers 2–8.** Units in layers 2–8 of the network, like layer 1 units, receive feedforward input exclusively from the immediately preceding processing stage. Unlike layer 1 units, however, which receive projections almost exclusively from the contralateral hemifield of the input image, inputs to units in the remaining layers are specified according to a binomial distribution. The latter distribution is used to determine whether each projection received by units beyond layer 1 originates from the ipsilateral or contralateral hemisphere in the preceding network layer. As was the case in layer 1, the number of input connections received by each unit is again controlled by the density parameter  $d$  (see below, Density). By parametrically changing the binomial probability  $p$  over successive layers, we enforce on the model a gradual increase in the frequency of interhemispheric connections along the network hierarchy.

When defining connections between layers 1 and 2, binomial probabilities were set to a value of 0.02. This implies that there is a 0.02 probability that the feedforward projection to that layer 2 unit corresponds to a contralateral projection and a 0.98 probability that it corresponds to an ipsilateral connection. Binomial probabilities of a contralateral connection in subsequent layer transitions, as specified by parameter  $p$ , were sequentially updated in steps of 0.08 until reaching a probability of 0.5 at the transition between layers 7 and 8 (Fig. 4b). The exact unit of the pertinent source hemisphere that projected to the target unit, as defined by the outcome of that binomial trial, was determined in a subsequent step. This was done by random sampling from a uniform distribution where all units in that network hemisphere have equal probability of being drawn.

**Density.** The number of input connections each network unit receives from units in the preceding processing stage is controlled by a parameter,  $d$ , which we term density. This value is defined once for each instantiation of the model. This implies that all network units, by definition, receive the same number of input projections. For a density value of one, each unit in the target layer receives input from one source unit in the preceding layer. This unique source of input will determine its activation level to any given image. In contrast, for higher values of  $d$ , each unit in the target layer receives  $d$  projections from—most probably—multiple units in the previous layer. For density values larger than one, each unit integrates its multiple inputs by averaging the activation values of its source units. Network density values explored in the model were defined as  $2^q$ , with  $q = 0, 1, 2, 3, 4, 5$ .

#### Model variants

Two model variants were explored. In the first variant, termed pixel-level model variant ( $MV_{\text{pixel}}$ ), grayscale face images served as input to the network model (for example images, see Fig. 5). RFs of units in layers 1–8 were specified as explained in the “Network architecture” section. Local luminance was the primary source of information propagated through this first variant of our model. In the second model variant, termed  $MV_{S1}$ , instead of grayscale images, we relied on the S1 representation of the HMAX model as implemented by Serre et al. (2005). An advantage of this implementation over the original formulation by Riesenhuber and Poggio (1999) is that filter parameters (orientation, effective width, wavelength) were adjusted to match the tuning profiles of S1 units with those of V1 parafoveal simple cells. As in earlier versions of HMAX, the implementation by Serre et al. first applies a filter bank to each input image. Gabor filter banks of four orientations ( $0, \pi/4, \pi/2$ , and  $3\pi/4$  radians) and 16 RF scales (0.19–1.07° of visual angle;  $7 \times 7$  pixels to  $39 \times 39$  pixels in steps of 2 pixels) thus lead to 64 filters, each associated with a unique pair of RF-size and orientation. All 64 filters are applied to each image location and a measure hereby obtained of the energy matching the spatial frequency and orientation of each filter. In each image location, we summed filter output across all orientations and frequencies. We thus obtained a single intensity value per image location summarizing the total energy in the image matching the basis filters (for an example S1 filtered image, see Fig. 5). This form of representation served as input to  $MV_{S1}$ .

The pixel-level representation captures biases in mean luminance across stimuli neglected by the S1-level representation. The S1-level representation was informed by electrophysiological measurements from macaque V1 and in this sense is more biologically plausible. Recent evidence revealing interactions between luminance and contrast responses in EVC (Vinke and Ling, 2020) suggest that these two representations provide complementary information and motivate their inclusion here. Our model does not aim to demonstrate that visual cortex is sensitive to luminance and contrast; it assumes it, based on prior empirical evidence from well-controlled studies.

**Event-related design.** Images provided as input to the network were presented emulating an fMRI event-related design. Each stimulation event was defined by a unique combination of face identity and orientation. In the case of the RaFD database, this led to  $57$  (identities)  $\times$   $5$  (orientations) = 285 event types. In the case of the KDEF database, this led to  $68$  (identities)  $\times$   $5$  (orientations) = 340 event types.

#### Pattern generation and analyses

**Network activation patterns and their simulated measurement.** In our network model, each unit models a single fMRI voxel. In turn, the collection conformed by all units in each hemisphere of a network layer define a region of interest (ROI)—for example, layer 1, left hemisphere. Then, for each image provided as input to a specific instantiation of the network, patterns of activation are defined by concatenating the activation levels observed in each unit of the ROI under consideration. For example, response patterns to input images associated with a right hemisphere ROI in layer 1 would consist of 2,048 entries, one entry per unit of the right layer 1 network hemisphere. In a similar fashion, the patterns associated with right hemisphere ROIs of layers 2–8 would each consist of 512 entries.

fMRI time series, one per voxel, naturally exhibit different levels of gain. Some fMRI voxels exhibit higher BOLD signal levels than others, due to, among many reasons, partial voluming (González Ballester et al., 2002), cortical folding (Polimeni et al., 2010; Gagnon et al., 2015), and specifics of the underlying vasculature (Bandettini and Wong, 1997; Schmid et al., 2019). A static measurement gain field (mGF) can be defined that summarizes the level of signal gain in each voxel (Ramírez et al., 2014, 2020). To model the influence of such gain field on our simulated response patterns, we randomly generated an mGF. For each network unit, we randomly sampled a scalar in the interval  $[0, 1]$  from a uniform distribution. Responses in each unit associated with each experimental condition were pointwise multiplied by its associated entry in the generated mGF, leading to simulated measured activation patterns in response to all images. We next used RSA to analyze patterns of activity across voxels within each ROI.

**Pattern similarity analyses.** RSA (Kriegeskorte et al., 2008) was used to analyze the simulated activation patterns. As in previous fMRI studies summarized in Table 1, our goal was to characterize the form of viewpoint tuning of responses to face stimuli presented in various viewpoints. We conducted three RSA variants to investigate if, and if so, how, the form of the inferred viewpoint representations differed. These RSA variants differed with regards to two factors: (1) distance measure used to represent pattern dissimilarities (Euclidean, or correlation) and (2) choice whether to demean the data prior to computing pattern dissimilarities.

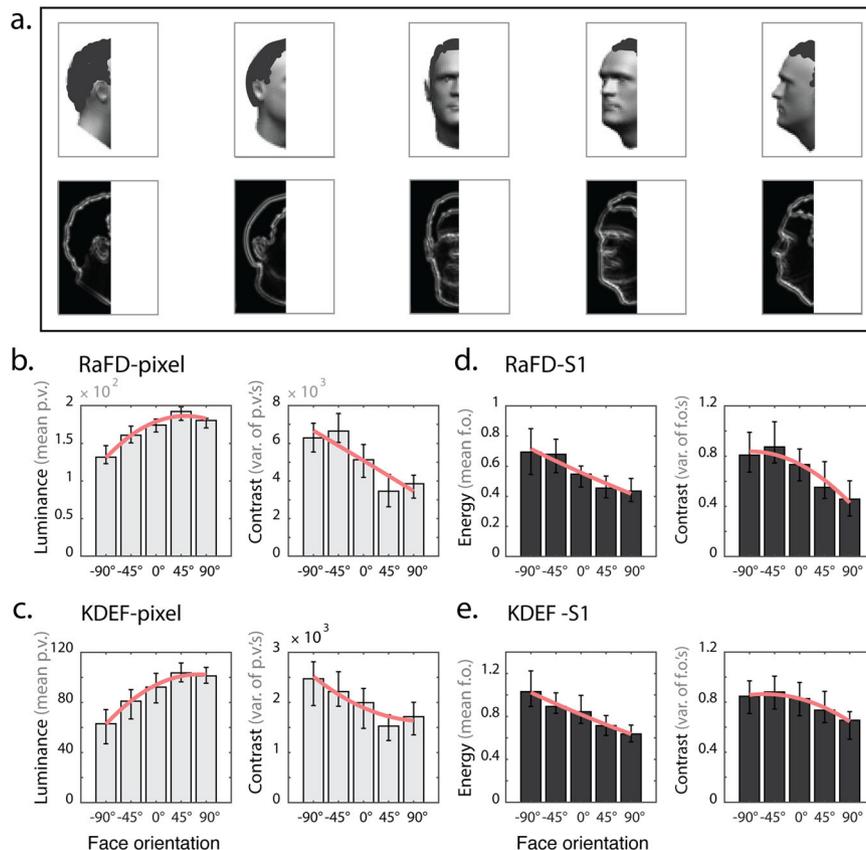
The RSA procedure consisted of two main steps. In the first step, we computed empirical dissimilarity matrices (eDSMs) from the simulated activity patterns elicited by each face identity in each region of interest. Please note that Kriegeskorte et al. (2008) refer to such matrices as representational dissimilarity matrices (RDMs). eDSMs were indexed by facial viewpoint ( $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ ), leading to 5 by 5 dissimilarity matrices. Two distance functions were used to define eDSMs: the correlation distance (eDSM<sub>corr</sub>) and the Euclidean distance (eDSM<sub>Eucl</sub>). Thus, two

of our RSA variants were uniquely specified by the choice of pattern dissimilarity measure (i.e., RSA<sub>corr</sub> and RSA<sub>Eucl</sub>). The third RSA variant considered the same simulated activity patterns as before, but computed pattern dissimilarities after first demeaning the data. In the latter procedure, also previously referred to as “cocktail demeaning” (Haxby et al., 2001; Garrido et al., 2013), the mean response across conditions is subtracted from that observed for each condition in each measurement channel (e.g., each fMRI voxel). See Figure 3 for further details. This operation has been implicitly and explicitly implemented in previous neuroimaging studies (Table 1 and Ramírez, 2017). Since angular distances, but not Euclidean distances, are affected by data recentering, only RSA<sub>corr</sub> was conducted on demeaned data. Thus, the option whether to demean the data defined the third RSA variant considered here—RSA<sub>corrDem</sub>.

In the second step, the computed eDSMs were compared with two representational models, each expressed as a model dissimilarity matrix (mDSM): (1) viewpoint model, which assumes neuronal response patterns in some brain area reflect head angular disparity, and (2) mirror-symmetry model, which assumes neuronal populations respond similarly to mirror-symmetric face orientations, for example,  $-90^\circ$  and  $90^\circ$  views (Fig. 1). These model templates capture two forms of view-tuning observed in macaques (see Introduction). As measure of representational similarity (or, in other words, agreement between the rank-order of the entries in the model and empirical DSMs), the Spearman rank-order correlation was computed between eDSMs and mDSMs. These correlations only considered upper triangular matrix entries and excluded the main diagonal (Ritchie et al., 2017).

#### Image-level analyses

For whole- (Fig. 5) and half-image analyses (Fig. 6), we computed the mean and variance of the pixel- and S1-level representation of each face identity presented in the five orientations considered in this study. In the case of the pixel-level representation, the mean is proportional to mean image luminance, and the variance to image contrast. Although strictly speaking



**Figure 6.** Distribution of mean luminance and contrast as a function of viewpoint for half-images for two face databases. Layout as in Figure 5. Here, however, only the left half of each image was analyzed. In contrast to Figure 5, where stronger quadratic than linear trends of mean luminance and contrast were usually observed across face views, linear trends proved dominant here regardless of database and representational format. Best fitting second-order polynomials are shown in red following criteria in Figure 5.

luminance is a photometric quantity, and here we are instead dealing with pixel values, we assume that a well-defined mapping exists between pixel values and luminance levels and use the term luminance accordingly throughout the manuscript. In the case of the S1-level representation, the mean corresponds to the mean energy matching the relevant Gabor filters (see above, Model variants for details), and the variance corresponds to the variability of the pooled filter responses for the relevant input image.

For each combination of face identity and orientation, we calculated the corresponding mean luminance and contrast and formed luminance and contrast profiles for each identity by concatenating the values associated with the five face orientations (Fig. 5). A general linear model (GLM) considering three regressors—namely, constant, linear, and quadratic—was fitted to each profile using ordinary least squares (OLS) error minimization. This model is a special case of the generalized linear model where the identity matrix is used as link function. Please note that the axis of symmetry of the quadratic regressor is by definition aligned to the front-on view ( $x = 0^\circ$ ) and the linear and quadratic regressors are orthogonal, each profile is hence projected onto mutually orthogonal subspaces. The squared regression coefficients divided by the total variance of each profile leads to partial coefficients of determination ( $R^2$ ) reflecting the proportion of variance explained by the linear and quadratic components of each profile.

Statistical tests were conducted either on the regression coefficients or the difference in partial  $R^2$  observed between the linear and quadratic regressors. The latter difference was used to characterize the dominant trend observed on these profiles—that is, linear (antisymmetric) or quadratic (symmetric). We also explored an additional decomposition further considering cubic and quartic regressors and thus exhaustively partitioning the variance of each profile into two mutually orthogonal symmetric and antisymmetric components. Dominant trends were tested as above by examining the variance jointly explained by the even (quadratic and quartic) and odd (linear and cubic) regressors, in other words, by comparing the variance explained by projection of the profiles onto these two mutually orthogonal subspaces. Here, we loosely refer to trends described by even polynomials as symmetric and those described by odd polynomials as antisymmetric (but see Petitjean, 2021).

### Statistical analyses

For the sake of consistency, statistical tests reported throughout this paper were initially implemented by means of bootstrap tests on medians making minimal assumptions about the nature of the underlying distribution of the tested statistics. All statistical analyses considered 10,000 bootstrap resamples. When testing in a population (i.e., face database) for systematic effects on GLM regression coefficients (for details, see Image-level analyses, immediately above), we tested the null hypothesis of zero median of the pertinent average regression coefficient by bootstrapping each database; this is, by resampling with replacement instances from that database. All statistical tests were two-sided, unless a directional hypothesis is explicitly stated. To evaluate the dominance of linear over quadratic trends (and vice versa) irrespective of the sign of a polynomial trend across different face identities in the population, we tested the null hypothesis of zero median across bootstrap resamples of the average difference of the partial  $R^2$  associated with the linear and quadratic regressors obtained for each face identity. Similarly, to more generally test the dominance of symmetric over antisymmetric components, we tested the null hypothesis of zero median across bootstrap resamples of the average difference of the partial  $R^2$  associated with the even (square and quartic) versus odd polynomial (linear and cubic) regressors.

Statistical tests for RSA were also initially implemented by bootstrap tests of medians. To evaluate a relative increment (or decrement) of viewpoint versus mirror-symmetry when comparing early network layers (defined here as layers 1 and 2) and late layers (defined as layers 7 and 8), we compared in each combination of population (RaFD, KDEP) and model variant ( $MV_{S1}$ ,  $MV_{\text{pixel}}$ ) the difference in correlation between the simulated empirical DSMs and our two model DSMs (symmetry and viewpoint). A relative increment (or decrement) when comparing early and late layers would indicate an interaction between hierarchical level of the network and the inferred prevalent

form of view-tuning—for example, a shift from a view-tuned representation in early layers to a predominantly mirror-symmetric representation in late layers. To test if the viewpoint and symmetry models each significantly correlate with the simulated data in early and late network layers, for each combination of population and model variant, we tested the null hypothesis of zero median of the Spearman correlation between the simulated data and the tested model. Finally, to compare the level of the network hierarchy at which the observed representational structure shifted from predominantly view-tuned to mirror-symmetric, we ran bootstrap tests on medians of the average zero-crossing of the difference between the viewpoint and mirror-symmetric models.

Our initial bootstrap statistical significance tests make minimal distributional assumptions. However, because this analysis approach can be suboptimal, we confirmed all our findings relying on bootstrap tests on medians with one-sample  $t$  tests on means (when testing the statistical significance of regression coefficients against zero), paired  $t$  tests when comparing the difference in magnitude of the linear and quadratic regression weights, and sign-permutation tests (when testing the statistical significance of the difference in partial  $R^2$  associated with the linear and quadratic terms of our GLM). Importantly, the observations reported in this paper proved robust to choices concerning the tested parameters (mean, or median) as well as the preferred statistical procedures (bootstrap tests on means or medians, one-sample and paired  $t$  tests, sign tests, and/or sign-permutation tests).

### fMRI data and analyses

Our model implies that RSA analyses of previous studies should exhibit a yet untested, but very specific, pattern of biases as a function of the location of an ROI along the ventral stream when subjected to  $RSA_{\text{corr}}$ ,  $RSA_{\text{corrDem}}$ , and  $RSA_{\text{Euc}}$ . To test these predictions, we analyzed a relevant human fMRI dataset ( $n = 8$ ; 5 males, 3 females). For a detailed methods description, see Ramírez et al. (2014). A preliminary report of this study was published in abstract form (Ramírez et al., 2010). MRI data were acquired on a 3 T Siemens Trio scanner with a 12-channel head coil. Functional images were acquired with a gradient-echo EPI sequence [TR, 2,500 ms; TE, 30 ms; flip angle,  $70^\circ$ ; matrix,  $128 \times 96$ ; FOV,  $256 \times 192$  mm; 30 slices (2 mm thick, no gap, interleaved acquisition)], resulting in a 2 mm isotropic voxel resolution. Slices were positioned along the slope of the temporal lobe and covered ventral visual cortex. The functional localizer comprised 260 volumes and each of the runs of the main experiment 298 volumes.

EPI data were preprocessed and subsequently analyzed in single-subject space. That is, we did not perform spatial normalization of participants to a common reference space. RSA analyses were performed on brain activation patterns associated with faces in the same five facial viewpoints explored with our model ( $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ). All MVPA analyses were conducted on unsmoothed data. In each of five experimental runs, stimuli were presented in randomized order following a mini-block design. Experimental conditions included faces and vehicles in the five viewpoints listed above. Activations associated with each experimental condition were obtained after convolving the stimulus presentation times in each run with a canonical HRF. For each subject, we modeled cortical responses for each of the five experimental runs separately. Brain activations for each face view were obtained by generalized linear modeling with SPM2 (<http://www.fil.ion.ucl.ac.uk/spm>). Here, we only analyzed activation patterns associated with the face stimuli. In each region of interest, voxels were selected following standard methods (for details, see Ramírez et al., 2014), and brain patterns formed by concatenating parameter estimates for each condition across voxels.

RSA analyses in each ROI were conducted following the procedures described in the “Pattern similarity analyses” section for the analysis of simulated brain patterns. The input patterns for RSA analyses were obtained by averaging across the five experimental runs the regression coefficients for each of the five facial viewpoints. In each ROI, empirical DSMs were computed according to the three distance measures explored in this paper: namely,  $RSA_{\text{corr}}$ ,  $RSA_{\text{corrDem}}$ , and  $RSA_{\text{Euc}}$ . The similarity observed in each ROI between empirical and model DSMs was measured with the Spearman rank-order correlation.

Average correlation coefficients across subjects for the RSA analyses reported in Table 4 were obtained after Fisher  $z$ -transforming

correlations for each subject and average correlations then back-transformed. Statistical significance testing was performed with one-sample  $t$  tests and paired  $t$  tests against zero, always on Fisher  $z$ -transformed correlations. Because prior studies have revealed view-tuned and mirror-symmetric effects in human visual cortex, tests against zero were one-tailed. In contrast, because the evidence is mixed regarding which representation is prevalent in each brain area, paired  $t$  tests comparing the viewpoint and mirror-symmetric models were two-tailed.

#### Code accessibility

Code to reproduce all simulations reported in this paper is available in the following repository: <https://github.com/toporam/model-crossings>.

## Results

Of seven studies investigating viewpoint representations in humans using fMRI-MVPA, five focused on brain activation patterns found to be reliable within individual subjects (Table 1). Our model primarily addresses these five studies. The remaining two studies relied on across-subject RSA approaches unsuited to draw inferences about spatially structured brain patterns reliable at the single-subject level (Sabuncu et al., 2010; Haxby et al., 2011; Yamada et al., 2015; Feilong et al., 2018). We have previously addressed the interpretation of these types of studies (Ramírez et al., 2020). Of the five within-subject fMRI-MVPA studies in Table 1, three considered a task that required subjects to fixate on a known image location (Axelrod and Yovel, 2012, their Experiment 2; Kietzmann et al., 2012; Ramírez et al., 2014) and for this reason the prime focus of our model. Axelrod and Yovel (2012, their Experiment 1) relied instead on a one-back identity task that allowed subjects to freely move their eyes. If the antisymmetric pattern of activations across face views reported in EVC, attributed by the authors to systematic eye movements toward the facial features (e.g., eyes and mouth), is assumed to generalize to the study by Guntupalli et al. (2017), which relied on a similar task, our model provides a straightforward account of these two additional experiments. Two further contributions included images of either left- or right-ward oriented faces and hence unsuited to investigate mirror-symmetric viewpoint representations (Natu et al., 2010; Foster et al., 2022). The central intuition behind the account proposed here is that, given previously observed low-level image biases across face views, considering a system exhibiting CM and a gradually increasing degree of interhemispheric connectivity is sufficient to account for key common and inconsistent trends observed among within-subject studies.

We present our results in four steps. First, we report analyses demonstrating the prevalence in two popular face databases of biases in the distribution of low-level features (i.e., mean luminance and contrast) across face views qualitatively like those previously

noted in a smaller set of face stimuli (Ramírez et al., 2020). Second, we present simulation results probing the impact of CM on the manifestation of these low-level biases in the input layer of our network model. Third, we present RSA results of response patterns to face images obtained from our network model. We focus on the impact of two analysis choices—that is, pattern dissimilarity measure and data demeaning—on the inferred form of the underlying viewpoint representations: view-tuned or mirror-symmetric (Fig. 1). Throughout this paper, when using the term “demeaning,” we refer to “cocktail demeaning,” unless specifically stated otherwise (see above, Pattern similarity analyses and Fig. 3 for details). Finally, we report RSA analyses identical to those conducted on the activity patterns provided as output by our model, but on empirical data from an fMRI study that measured brain responses to face stimuli presented in the same five facial viewpoints probed here (Ramírez et al., 2014). We target multiple regions of interest along the ventral visual processing stream, including EVC, the lateral occipital (LO) portion of the lateral occipital complex (LOC; Malach et al., 1995), as well as the right occipital face area (rOFA) and the right fusiform face area (rFFA; Sergent et al., 1992; Puce et al., 1996; Kanwisher et al., 1997; McCarthy, 1997).

### Image-level analyses: whole and half-images

To test whether low-level feature imbalances of the type assumed by our model are prevalent among face stimuli, we investigated the distribution of mean luminance and contrast across facial viewpoints ( $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ) of two popular face databases—the KDEF and RaFD databases. Figure 5 presents results for analyses of full images. The goal of these analyses was to uncover biases across conditions expected to manifest in brain areas populated by neurons with RFs that cross the vertical meridian into the ipsilateral hemifield. Two representational formats of images from these two databases were investigated. The first format was the pixel-level representation. The second format focused on the pooled output of frequency- and orientation-tuned filters of the S1 layer of HMAX in the biologically constrained implementation used by Serre et al. (2005) to model RFs of V1 simple cells (see Materials and Methods). We term this second format S1-level representation.

Pixel-level analyses revealed systematic quadratic biases about the frontal face-view in both databases (Fig. 5). Such biases were evident for both mean luminance and contrast. Statistical tests confirmed that when each database was treated as a random sample of face identities from some population, quadratic trends on regression coefficients were significantly different from zero (one-sample  $t$ -tests; all  $p < 0.001$ ). See Table 2 for detailed statistics. For both analysis variants and databases, quadratic trends were also found to be on average significantly stronger than their linear

**Table 2. Statistics of low-level feature analyses for whole images**

Database	Image and LLF	$M$ linear	$t$	$p$	$M$ quad.	$t$	$p$	$\Delta  Abs $	$t$	$p$	$\Delta R^2$	$p$
RaFD: df = 56	Pixel, mean	2.12	3.23	0.002	-23.45	-22.52	<0.001	-19.09	-17.99	<0.001	-0.81	<0.001
	Pixel, var	38.76	1.00	n.s.	845.98	7.23	<0.001	-774.66	-8.13	<0.001	-0.54	<0.001
	S1, mean	0.02	3.90	<0.001	0.07	7.16	<0.001	-0.05	-6.11	<0.001	-0.38	<0.001
	S1, var	0.05	5.98	<0.001	-0.02	-1.24	n.s.	-0.06	-4.99	<0.001	-0.34	<0.001
KDEF: df = 67	Pixel, mean	0.61	1.06	n.s.	-8.80	-11.26	<0.001	-6.97	-9.05	<0.001	-0.59	<0.001
	Pixel, var	43.38	1.68	n.s.	461.59	9.99	<0.001	-385.09	-12.81	<0.001	-0.70	<0.001
	S1, mean	-0.01	-2.14	n.s.	-0.001	-0.13	n.s.	-0.03	-4.62	<0.001	-0.25	<0.001
	S1, var	0.01	1.17	n.s.	-0.06	-4.08	<0.001	-0.06	-5.74	<0.001	-0.30	<0.001

First two columns indicate database and low-level-feature tested. Next six columns report the mean,  $t$ , and  $p$ -value of  $t$ -tests against zero. Next three columns ( $\Delta |Abs|$ ) report the mean,  $t$ , and  $p$ -value of paired  $t$ -tests of the magnitude of the linear and quadratic regression coefficients. The last two columns report for each low level feature the mean difference in partial  $R^2$  of the linear and quadratic terms, as well as its statistical significance according to sign-permutation tests. df, degrees of freedom of  $t$ -tests. Tests for the linear and quadratic regression coefficients were Bonferroni corrected. Significance level for all tests is  $\alpha = 0.05$ .

**Table 3. Statistics of low-level feature analyses for half-images**

Database	Image and LLF	$M$ linear	$t$	$p$	$M$ quad.	$t$	$p$	$\Delta$  Abs	$t$	$p$	$\Delta R^2$	$p$
RaFD: $df = 56$	Pixel, mean	36.68	18.22	<0.001	-19.46	-17.58	<0.001	17.22	11.95	<0.001	0.47	<0.001
	Pixel, var	-2,446.20	-23.95	<0.001	-151.41	-1.38	n.s.	1,799.70	13.49	<0.001	0.59	<0.001
	S1, mean	-0.21	-9.22	<0.001	0.02	1.87	n.s.	0.15	7.99	<0.001	0.53	<0.001
	S1, var	-0.31	-10.30	<0.001	-0.10	-6.66	<0.001	0.19	5.45	<0.001	0.36	<0.001
KDEF: $df = 67$	Pixel, mean	31.74	19.40	<0.001	-10.40	-13.15	<0.001	21.18	12.52	<0.001	0.65	<0.001
	Pixel, var	-671.35	-11.26	<0.001	142.28	2.23	n.s.	259.70	5.20	<0.001	0.26	<0.001
	S1, mean	-0.33	-16.78	<0.001	0.01	0.93	n.s.	0.25	11.93	<0.001	0.73	<0.001
	S1, var	-0.16	-7.56	<0.001	-0.12	-6.72	<0.001	0.05	2.30	0.024	0.15	0.039

Table format is exactly as Table 2.

**Table 4. Statistics for RSA analyses of empirical fMRI data**

RSA type	ROI	Mean corr. viewpoint	$t_{(7)}$	$p$	Mean corr. mirror	$t_{(7)}$	$p$	$\Delta$ (View, mirror)	$t_{(7)}$	$p$
$RSA_{corr}$	EVC	0.54	5.19	<0.001	-0.30	-4.40	n.s.	0.72	5.60	<0.001
	LO	0.42	3.41	0.006	-0.14	-1.04	n.s.	0.52	3.28	0.014
	OFA	0.21	2.38	0.024	-0.21	-1.76	n.s.	0.40	2.44	0.045
	FFA	0.26	3.72	0.004	0.00	0.05	n.s.	0.26	3.30	0.013
$RSA_{corrDem}$	EVC	0.60	6.18	<0.001	-0.21	-4.75	n.s.	0.72	6.23	<0.001
	LO	0.45	3.63	0.004	0.03	0.38	n.s.	0.42	2.85	0.025
	OFA	0.20	1.81	n.s.	0.09	0.89	n.s.	0.11	0.59	n.s.
	FFA	0.11	1.39	n.s.	0.22	2.10	0.037	-0.11	-0.81	n.s.
$RSA_{Euc}$	EVC	0.47	6.22	<0.001	-0.27	-3.87	n.s.	0.66	5.79	<0.001
	LO	0.33	3.61	0.004	-0.07	-0.64	n.s.	0.39	2.73	0.029
	OFA	0.15	2.45	0.022	0.01	0.09	n.s.	0.14	0.94	n.s.
	FFA	0.02	0.43	n.s.	0.19	2.56	0.019	-0.17	-3.07	0.018

ROI naming conventions as in Figure 10.

counterparts both in terms of their magnitude (paired  $t$ -tests; all  $p < 0.001$ ) and their associated partial  $R^2$  (sign-permutation tests; all  $p < 0.001$ ). As can be observed in Figure 5, the sign of these quadratic biases was congruent across databases.

However, for S1-level analyses of the distribution of mean filter energy across facial viewpoints, only the RaFD database revealed a significant quadratic trend (one-sample  $t$ -test;  $t_{(56)} = 7.2$ ;  $p < 0.001$ ; Fig. 5). Note that this analysis is based on averages of S1 filter responses instead of pixel values, and the observed trends are therefore not necessarily proportional to mean luminance. These two measures provide distinct information about the images. When analyzing the variance of the S1 filter energy observed across image locations (an analysis similar to image contrast across views but, again, now based on S1 filter outputs instead of pixel values), only the KDEF database revealed a significant quadratic trend (one-sample  $t$ -test;  $t_{(67)} = -4.1$ ;  $p < 0.001$ ). In contrast, only the linear regression coefficient proved significantly different from zero for the RaFD database (one-sample  $t$ -test;  $t_{(56)} = 6.0$ ;  $p < 0.001$ ). Importantly, for both databases and analysis variants, the quadratic trends were found to be significantly stronger than their linear counterparts in terms of their magnitude (paired  $t$ -tests; all  $p < 0.001$ ) as well as their associated partial  $R^2$  (sign-permutation tests; all  $p < 0.001$ ). In sum, analyses on whole images revealed significant mirror-symmetric biases across face views. Critically, quadratic components always significantly outweighed their linear counterparts in terms of partial  $R^2$ . On average, the linear term accounted for only 13% of the variance, while the quadratic term accounted for 62% of the variance. These results show that mirror-symmetric low-level biases about the frontal view are common among face stimuli, in line with previous observations (Ramírez et al., 2020). We next asked if the opposite pattern of results is observed for half-images, as assumed by our model (Fig. 2). We analyzed half-images to see if the linear trend in this case outweighed the quadratic trend.

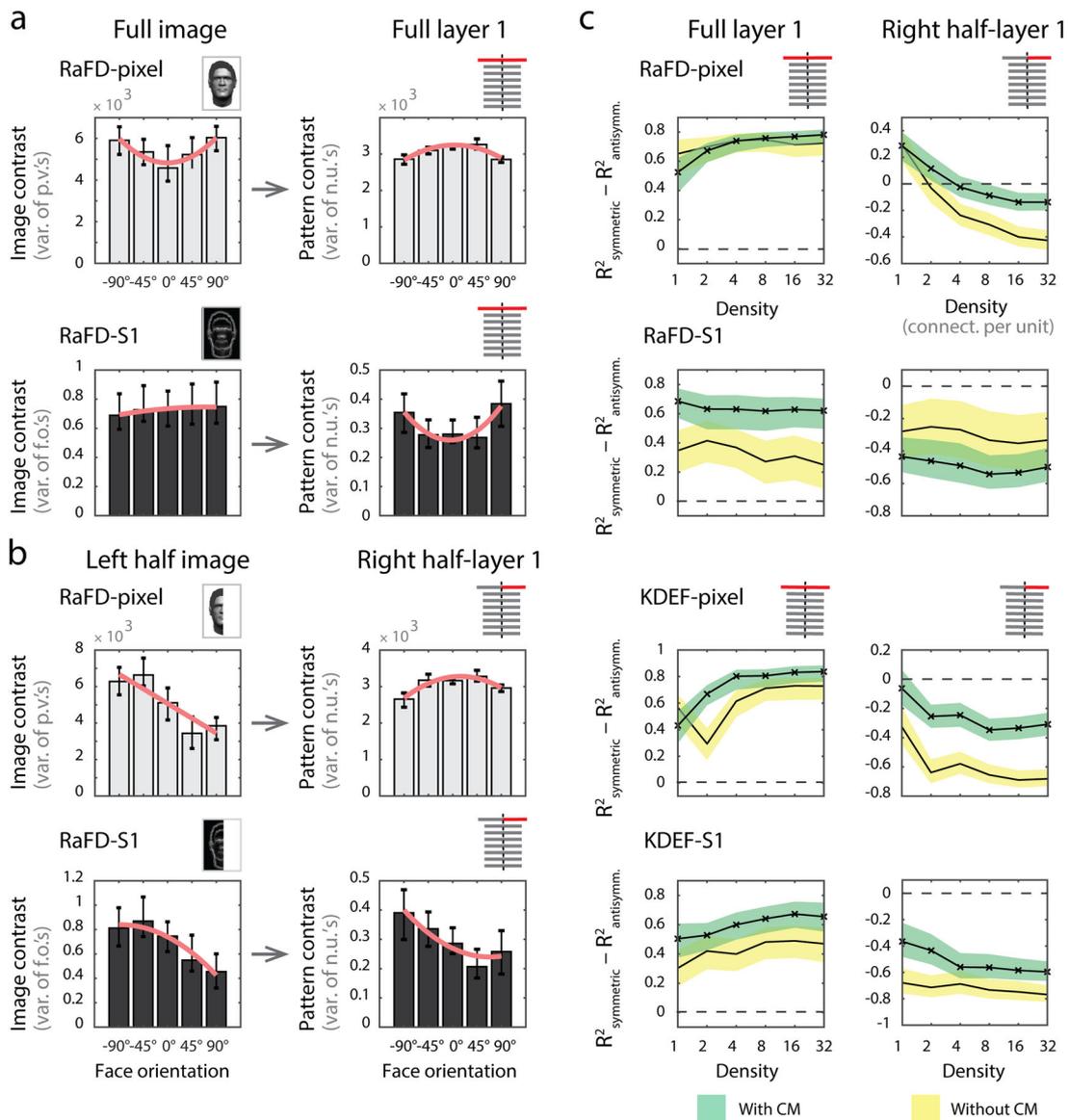
Half-image analyses otherwise identical to those reported for full images in Figure 5 are summarized in Figure 6 and Table 3. These analyses aim to characterize low-level biases across face views due to mean luminance and contrast. Such biases are likely to be observed in areas with RFs circumscribed to one visual hemifield, such as V1. As hypothesized, a clear dominance of the linear over the quadratic trend was observed for the pixel model variant for both low-level features, regardless of database, when analyzing mean regression coefficients (paired  $t$ -tests; all  $p < 0.001$ ). A similar pattern of results was observed for the mean and variance of the S1 filter outputs (paired  $t$ -tests; all  $p < 0.025$ ). See Table 3 for detailed statistics. For completeness, we decomposed the profiles for each face identity into linear and quadratic components, as done for analyses of the full images. We found that the linear component was significantly larger than its quadratic counterpart in terms of partial  $R^2$  regardless of low-level feature, database, and representational format (sign-permutation tests; all  $p < 0.05$ ). On average, the linear term accounted for 66% of the variance, while the quadratic term accounted for only 19% of the variance—the opposite to what was observed when analyzing full images. This implies a double dissociation for full- and half-images with regard to their dominant polynomial trends—namely, quadratic over linear for full images and linear over quadratic for half-images. In sum, we found evidence of low-level biases in half-images consistent with the antisymmetric biases assumed by our model, as well as evidence of the dominant symmetric biases about the frontal face view anticipated for full images.

#### Impact of CM and network density on layer 1 activation profiles

If biases across face views like those described in Figures 5 and 6 were present, respectively, in FFA and V1, this would suggest a simple account of key similarities and inconsistencies observed

across the fMRI-MVPA studies summarized in Table 1 (see also Fig. 1). However, it may be argued that image-level analyses inadequately represent biases that exist at the level of images when it comes to consider their influence on V1 activation profiles. Crucially, both image-level representations considered here lack a fundamental property of primate V1, namely, CM. To assess whether biases qualitatively like those reported in Figures 5 and 6 are observed in a representation considering CM, we conducted analyses homologous to those reported in Figures 5 and 6, but now on activation patterns associated with the same images in layer 1 of the feedforward randomly connected two-hemisphere network proposed here (Fig. 4).

First, we assessed whether the mean activation observed across all layer 1 units (roughly analogous to analyzing whole images) or, alternatively, across only one hemisphere of layer 1 (roughly analogous to analyzing half-images) exhibit trends qualitatively like those found at the pixel and S1 level. When jointly considering units from both hemispheres of layer 1, as in full-image results, we again found marked mirror-symmetric biases as a function of viewpoint for both mean and variance of layer 1 activation patterns. However, we also noted that the positive quadratic trend across views previously observed for contrast (pixel-level, whole image) now exhibited a significant negative quadratic trend (Fig. 7a, top). In a similar vein, we noted



**Figure 7.** Impact of CM and network density on layer 1 activation profiles. **a, b,** Results are organized according to portion of the image (whole, or left-half) or network hemispheres analyzed (both, only right). Image- and network-level analyses are shown, respectively, on the left and right column. Analyses computed on pixel- and S1-level representations are shown, respectively, on the top and bottom row. **a,** Top, left; Median image contrast across face identities for each viewpoint. Error bars indicate interquartile ranges. Best fitting second-order polynomials are overlaid on each plot. Top, right; Median variance of activation patterns associated with each face identity in layer 1. Unlike image-level analyses shown to the left, network analyses shown to the right consider CM of central image locations. **b,** Median contrast across face identities as a function of viewpoint for half-images. Panel is organized as panel **a.** Note differences in form and direction of trends when contrasting image- and network-level analyses. **c,** Difference in partial  $R^2$  of symmetric (quadratic and quartic) and antisymmetric (linear and cubic) trends for layer 1 activation patterns as a function of network density (x-axis) and CM (green, CM; yellow, no CM). As in panel **a,** activation profiles for each face identity were formed by concatenating the variance of activation patterns for each face view. Shaded areas indicate 95% confidence intervals. Positive values indicate stronger symmetric than antisymmetric trends. Note consistency in the direction of dominant trends regardless of database, CM, and number of hemispheres. p.v., pixel value; f.o., filter output; n.u., network unit.

that the dominant linear trend found for contrast for RaFD-S1 changed to positive quadratic (Fig. 7*a*, bottom). These observations demonstrate that although CM does influence the precise form and strength of the mirror-symmetric biases noted when analyzing image-level representations, marked mirror-symmetric biases of the form assumed by our model are again observed when probing layer 1 activation profiles. In turn, analyses of patterns from only one hemisphere of layer 1 revealed dominant antisymmetric biases consistent with those shown in Figure 6 for half-images (Fig. 7*b*).

Next, we evaluated if the pattern of results in layer 1 of our network model is specific to the lowest density level, as reported in Figure 7*a,b*, or prevalent over a wider range of densities. High sensitivity to network density would in our view challenge our model as a plausible explanation of empirical data. In contrast, robustness over a wider range of densities would lend plausibility to the model. To address this issue, we quantified the relative dominance of symmetric and antisymmetric trends across face views in each network layer as a function of network density. Profiles describing the signal strength across face views for each face identity were specified by concatenating the variance of the activation pattern associated with each face view for that identity. The relative strength of symmetric (even) and antisymmetric (odd) polynomial trends on these activation profiles was evaluated by comparing the distribution across face profiles of the partial  $R^2$  associated with the symmetric and antisymmetric components. Please note that this variance decomposition projects face profiles onto mutually orthogonal subspaces. We did this for both databases (RaFD and KDEF) and representational formats (pixel and S1). To assess the impact of CM on network behavior, we also ran a set of simulations ignoring this property. As observed in Figure 7*c*, while CM and network density were found to modulate the relative influence of the analyzed symmetric and antisymmetric trends, critically, the direction of these biases was consistent across network density levels. The only exception was the observation for the lowest density level ( $d=1$ ) of a dominant symmetric trend for half-layer network analyses for  $MV_{\text{pixel}}$  for the RaFD database (Fig. 7*c*, top row, right column; see Discussion). Except for this observation, as expected, a predominance of the antisymmetric component was evident for densities larger than two regardless of image database and model variant. Symmetric and antisymmetric biases of the form assumed by our model were found with and without CM.

### RSA of simulated activity patterns

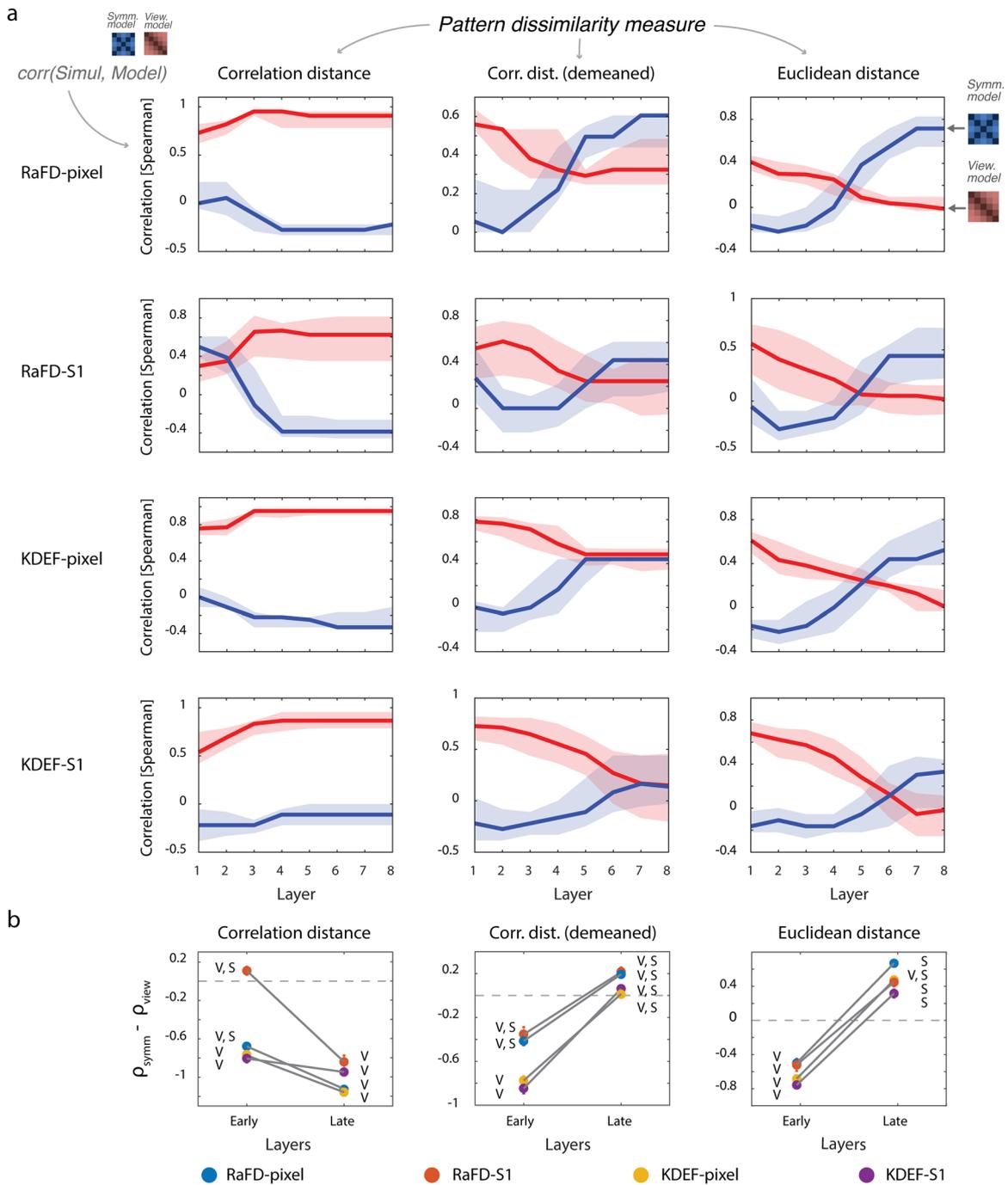
The biases described thus far are agnostic regarding the spatial structure of the underlying distributed response patterns. The reported analyses are only informative regarding the mean and variance of the simulated brain patterns as a function of viewpoint. According to our account, such biases may be sufficient to explain the dominant trends observed across the fMRI-MVPA studies considered here. Naturally, we next asked if analyses of the patterns associated with face images of the two datasets reveal biases sufficient to account for the commonalities and inconsistencies summarized in Figure 1.

To address this question, we subjected activation patterns in each network layer to RSA (see Materials and Methods for details). We computed (simulated) empirical dissimilarity matrices (eDSMs) describing the dissimilarity relationships of the five facial viewpoints for each face identity. eDSMs were computed according to two distance measures—correlation distance and Euclidean distance. Because angular relationships across

pattern-vectors are expected to change after data demeaning only with angular, but not Euclidean distances (Fig. 3), we also report RSA results on cocktail-demeaned data when using the correlation distance to measure pattern dissimilarities. For analyses on thus demeaned data, the mean response across experimental conditions was computed for each network unit and subtracted from the value associated with each condition, as previously done for fMRI voxels (Table 1). We argue that these three analysis strategies (RSA<sub>corr</sub>, RSA<sub>Euc</sub>, and RSA<sub>corrDem</sub>) capture key analysis choices across studies that may largely account for the observed inconsistencies. Pattern analysis results are summarized in Figures 8 and 9. We first report simulations considering a network density of 16 connections per unit. As expected, analyses relying on the correlation distance led to the observation of markedly view-tuned representations throughout the eight network layers (Fig. 8*a*, left). For both face databases and model variants, the monotonically view-tuned model exhibited higher correlations with the simulated eDSMs than the mirror-symmetric model. The only exception was for the S1 model variant for the RaFD database, which exhibited in layer 1 higher correlations with the mirror-symmetric than the view-tuned model. This bias, observed with RSA<sub>corr</sub>, reflects a genuine mirror-symmetric bias in the spatial structure of images from this database.

Analyses relying on the correlation distance but—crucially—now conducted on demeaned data, as well as analyses relying on the Euclidean distance, led to marked increments in later network layers in the association between the simulated eDSMs and the mirror-symmetric model (Fig. 8*a*). As further shown in Figure 8*b*, correlations between the mirror-symmetric mDSM and the simulated data were found to be significantly larger than zero (sign-permutation tests; all  $p < 0.001$ ), as well as to increase significantly in later layers (defined as 7 and 8) with respect to earlier layers (defined as 1 and 2; sign-permutation tests; all  $p < 0.001$ ). Moreover, in 2/4 combinations of database and model variant for RSA<sub>corrDem</sub>, and in 4/4 combinations for RSA<sub>Euc</sub>, the mirror-symmetric model significantly outperformed the view-tuned model in later layers (sign-permutation tests; all  $p < 0.001$ ). For RSA<sub>Euc</sub> and RSA<sub>corrDem</sub>, no instance was found where the view-tuned model significantly outperformed the mirror-symmetric model. In sum, these observations show that a parsimonious model incorporating CM and gradually increasing interhemispheric crossings, given the observed low-level image biases, leads to geometric configurations in multivariate pattern space that account for inconsistencies observed in higher-level processing stages such as FFA. Specifically, these inconsistencies arise due to the influence on RSA outcomes of data analysis choices—namely, pattern dissimilarity measure and whether to demean the data prior to RSA. Our observations also explain the view-tuned pattern of response consistently reported in EVC (Table 1).

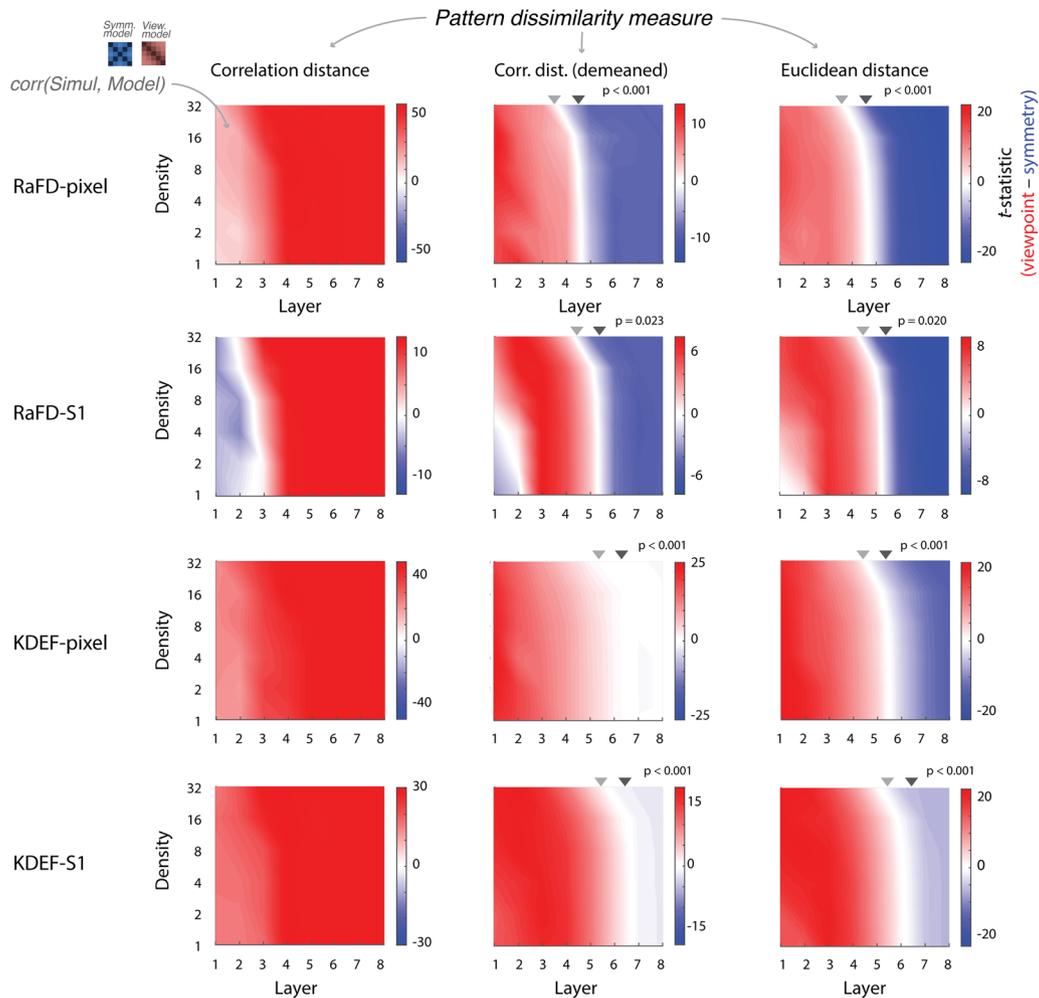
Next, we assessed whether the pattern of results observed for density 16 is consistent across a broader range of network densities. Identical analyses as those reported in Figure 8, but for a wider range of densities, are summarized in Figure 9. These analyses show a broadly consistent pattern of results within the range of network densities explored. However, we observed two unexpected interactions. First, RSA<sub>corrDem</sub> for the S1 model variant for the RaFD database revealed an interaction between network layer, density, and the observed representational structure. In more detail, only for early network layers (1 and 2) and low densities, the mirror-symmetric model outperformed the viewpoint model (Fig. 9, second row, middle column). In contrast, when



**Figure 8.** RSA results across network layers for density level = 16. **a**, Each subplot summarizes RSA outcomes for network layers 1–8 for a single combination of image database, model variant, and RSA type. Rows indicate image database (KDEF, RaFD) and model variant (pixel, S1) combinations. Columns indicate RSA type: correlation distance ( $RSA_{\text{corr}}$ ), correlation distance on demeaned data ( $RSA_{\text{corrDem}}$ ), and Euclidean distance ( $RSA_{\text{Euc}}$ ). For each network layer ( $x$ -axis), median correlations are shown between simulated empirical dissimilarity matrices and the view-tuned (in red) and the mirror-symmetric models (in blue). Shaded areas indicate interquartile ranges. Note consistently view-tuned representations across layers for  $RSA_{\text{corr}}$  (except for RaFD-S1 at layer 1). In contrast, however, for  $RSA_{\text{corrDem}}$  and  $RSA_{\text{Euc}}$  a relative decrease of view-tuning and increase of mirror-symmetry is observed along the hierarchy. **b**, Average correlation difference between the symmetry and viewpoint models ( $y$ -axis), in early (1 and 2) and late (7 and 8) network layers. “V” and “S” indicate that correlations with the viewpoint and symmetry models, respectively, are significantly above zero. Note for  $RSA_{\text{corrDem}}$  and  $RSA_{\text{Euc}}$  a clear shift toward mirror-symmetry in later network layers. Gray lines indicate statistically significant mean increments or decrements when comparing early and late network layers (sign-permutation tests; all  $p < 0.001$ ).

the data were not demeaned, the mirror-symmetric model outperformed the viewpoint model in these same layers regardless of network density (Fig. 9, second row, left column). This observations with  $RSA_{\text{corr}}$  and  $RSA_{\text{corrDem}}$  reflect a genuine mirror-symmetric bias in the spatial structure of images from this database. The interaction between inferred form of view-tuning, network density, and RSA distance measure is explained in the

Discussion (see below, Simulation results). Second, for 4/4  $RSA_{\text{Euc}}$  and 4/4  $RSA_{\text{corrDem}}$  analyses, the network layer at which the observed representations changed from predominantly view-tuned to mirror-symmetric occurred slightly earlier for the highest density ( $d = 32$ ) than that for the remaining densities. Statistical analyses showed that these differences are significant in the two populations (sign-permutation tests; all  $p < 0.025$ ).



**Figure 9.** RSA results for multiple network densities. Layout as in Figure 8. Here, y-axis denotes density level of the analyzed patterns. Each plot color codes the paired *t*-statistic comparing correlation coefficients of the simulated empirical DSMs with the viewpoint and symmetry models. Areas in red indicate consistently higher correlations with the viewpoint model. Areas in blue indicate consistently higher correlations with the symmetry model. Results are broadly concordant with those shown in Figure 8 for a network density of 16 (see text for one exception). Dark gray and light gray triangles, respectively, indicate for the lowest and highest network densities ( $d = 1$  and  $32$ ) the point at which a shift occurs from a predominantly view-tuned to a predominantly mirror-symmetric representation. Triangles are shown when the mean difference in the point of zero-crossing for the two network densities was significantly larger than zero (sign-permutation tests).

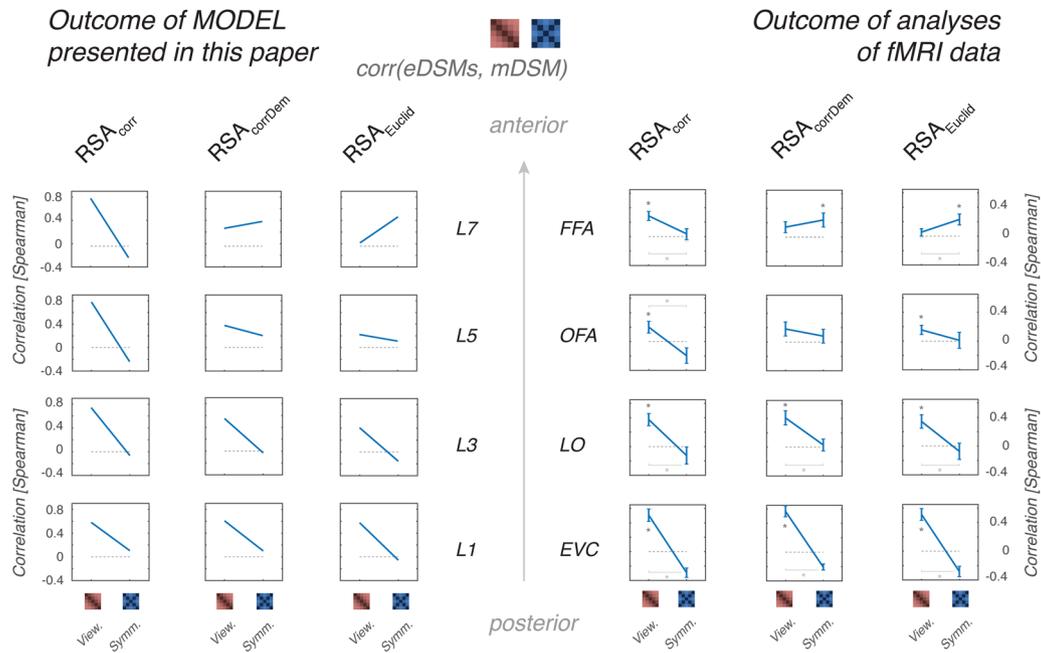
These observations indicate that network density can significantly influence the outcome of RSA results. Overall, and as expected, analyses reported in Figure 9 revealed view-tuning with  $RSA_{corr}$  regardless of network layer, and  $RSA_{corrDem}$  and  $RSA_{Euc}$  led to the observation of view-tuning in early layers and mirror-symmetry in later layers. Crucially, data demeaning induced mirror-symmetry in the simulated eDSMs only in later network layers (Figs. 8b, 9). On the contrary, data demeaning induced view-tuning in early network layers. This can be observed in Figure 9 when comparing the color of higher network densities in layer 1 for  $RSA_{corr}$  and  $RSA_{corrDem}$  for the RaFD-S1 model; while  $RSA_{corr}$  looks blue (i.e., mirror-symmetric),  $RSA_{corrDem}$  looks red (i.e., view-tuned).

**Empirical evaluation of key predictions derived from the network model**

Thus far, we found that overall trends in RSA analyses probing the activity patterns provided as output by our model consistently confirmed biases of the form we initially hypothesized (Fig. 2). Our model thus provides a plausible explanation for the key commonalities and inconsistencies across the fMRI studies described in Figure 1. Our model, however, makes many

simplifying assumptions and implies that RSA outcomes of previous studies should exhibit a yet untested, but very specific, pattern of biases as a function of the location of an ROI along the ventral stream when subjected to  $RSA_{corr}$ ,  $RSA_{corrDem}$ , and  $RSA_{Euc}$ . Critically, our model predicts that RSA outcomes in FFA should reveal a significant interaction of inferred representation (view-tuned or mirror-symmetric) as a function of the pattern dissimilarity measure used for RSA.

To lend further support to our model and test these predictions, we conducted analyses of empirical fMRI data associated with faces shown in the same five facial viewpoints probed with our network model (i.e.,  $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ). We subjected fMRI activity patterns from EVC, the LO portion of the LOC, the rOFA, and the rFFA to RSA analyses identical to those implemented in Figures 8 and 9 on simulated patterns. These results are summarized in Figure 10 and Table 4. As predicted, we observed evidence of view-tuning regardless of RSA approach in EVC (one-sample *t*-tests;  $t_{(7)}$ , all  $t < 5.2$ ; all  $p < 0.001$ ). In stark contrast, in FFA with  $RSA_{corr}$ , we found only evidence of view-tuning (one-sample *t*-test;  $t_{(7)} = 3.72$ ;  $p = 0.004$ ), while  $RSA_{Euc}$  (one-sample *t*-test;  $t_{(7)} = 2.56$ ;  $p = 0.019$ ) and  $RSA_{corrDem}$  (one-sample *t*-test;  $t_{(7)} = 2.10$ ,  $p = 0.037$ ) revealed



**Figure 10.** Empirical evaluation of model predictions for face stimuli in multiple processing stages along the posterior-anterior axis of the visual hierarchy. The outcome of RSA analyses based on the response patterns derived from our model for face images in different viewpoints ( $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ) is shown on the left side of this figure (see text for details). Each row shows the mean correlation with the mirror-symmetric and viewpoint model templates in layers 1, 3, 5, and 7 according to the three approaches to RSA probed in this paper, namely, using the correlation distance as measure of pattern dissimilarity ( $RSA_{corr}$ ), the correlation distance on demeaned data ( $RSA_{corrDem}$ ), or the Euclidean distance ( $RSA_{Euclid}$ ). On the right side of the figure, the outcome of identical RSA analyses of empirical fMRI data for face stimuli in the same five rotational angles are shown for EVC, the LO portion of the LOC, the OFA, as well as the FFA. Error bars correspond to standard errors of means. Stars indicate values significantly different from zero. In turn, bars with stars indicate that the difference between the two models is statistically significant. Significance level for all tests is  $\alpha = 0.05$ . See Materials and Methods and Table 4 for details.

significant associations only with the mirror-symmetric model. Importantly, while correlations with the mirror-symmetric model in FFA with  $RSA_{Euclid}$  were significantly higher than the viewpoint model (paired  $t$ -test;  $t_{(7)} = -3.07$ ;  $p = 0.018$ ), the opposite was observed with  $RSA_{corr}$ ; correlations with the viewpoint model were significantly higher than the mirror-symmetric model (paired  $t$ -test;  $t_{(7)} = 3.30$ ;  $p = 0.013$ ). No significant difference was observed with  $RSA_{corrDem}$  between the viewpoint and mirror-symmetric models.

To summarize, we found evidence in rFFA of the specific interaction of inferred representation as a function of analysis choices predicted by our model. Similarly, and also as predicted, we observed in EVC only evidence of a view-tuned representation regardless of analysis choices. The observed pattern of results in areas located between EVC and FFA was broadly consistent with the gradual change in RSA outcomes as a function of the location of an ROI along the ventral visual processing stream predicted by our model (compare Figs. 2*b*, 10).

## Discussion

This study evaluated the ability of a parsimonious model to explain concordant and discordant results across fMRI-MVPA studies investigating facial viewpoint representations in human visual cortex. Our simulations and empirical observations explain why conclusions in EVC are consistent across studies regardless of data demeaning and pattern dissimilarity measure, while conclusions in higher-tier areas depend on these choices. Our results highlight that prior fMRI-MVPA studies may have misinterpreted low-level feature imbalances across conditions as evidence of mirror-symmetrically tuned neuronal populations. To evaluate if that is indeed the case, additional fMRI

experiments should be performed to test if mirror-symmetric representations disappear when these imbalances are removed. Below we discuss our findings, their broader significance, and strengths and limitations of our model.

### Low-level biases as a function of viewpoint

Image-level analyses of two popular databases replicated previous observations of mirror-symmetric biases about the frontal face view (Ramírez et al., 2020). Analyses of half-images instead revealed antisymmetry. Our simulations revealed that although biases persisted with and without CM, their exact nature depended on this parameter. CM should be therefore considered in models aiming to distinguish the impact of low-level features from other variables of interest, like category structure and form of view-tuning.

### Simulation results

Our simulations replicated key empirical observations. As observed in FFA, we found mirror-symmetry in later network layers with  $RSA_{Euclid}$ , but view-tuning with  $RSA_{corr}$ . This is consistent with observations using classification rates from multiclass linear support-vector machines as proxy for pattern dissimilarity (Axelrod and Yovel, 2012). This metric is sensitive to the Euclidean distance between pattern and vectors. In contrast, another study relying on  $RSA_{corr}$  found in FFA only evidence of view-tuning and no mirror-symmetry (Ramírez et al., 2014). Noteworthy, both studies observed stronger responses for frontal than lateral face views in FFA. Such trends are expected to lead to the observation of mirror-symmetry with  $RSA_{Euclid}$ .

A geometric principle behind our model is that cocktail demeaning does not change Euclidean distances among fMRI pattern-vectors but does change their angular distances.

Moreover, cocktail demeaning can induce mirror-symmetry on eDSMs if mirror-symmetric biases in signal strength exist in the data (Ramírez, 2017). Our results support these concepts and help explain a reported gradual increase of mirror-symmetry, first noticed in OFA, but found to be prevalent in the ventral and dorsal streams with  $RSA_{\text{corrDem}}$  (Kietzmann et al., 2012). This finding is reminiscent of a reported gradual increase in BOLD sensitivity to ipsilateral stimuli along the ventral stream (Hemond et al., 2007).

All studies in Table 1 observed view-tuning in EVC (regardless of analysis choices). Our model reproduced these observations. View-tuning is observed in early network layers because the length of pattern-vectors associated with different face views, like their angular distances, increase monotonically with head angular disparity. Hence, both  $RSA_{\text{corr}}$  and  $RSA_{\text{Euc}}$  lead to view-tuned eDSMs since the former is sensitive to pattern-vector angles and the latter to their lengths. Cocktail demeaning in the presence of antisymmetric biases on vector lengths induces antisymmetry on RSA analyses with angular distances, and eDSMs hence remain antisymmetric (“view-tuned”) after demeaning the data.

In contrast, RSA outcomes in later network layers depend on distance measure and cocktail demeaning because of the inconsistent trends observed on vector lengths and angles—mirror-symmetry on lengths and antisymmetry on angles. Cocktail demeaning across conditions in the presence of a quadratic trend on vector lengths induces mirror-symmetry on ensuing RSA analyses with angular distances. Thus, the low-level confounds described in Figure 6 imply a bias to observe view-tuning in EVC even if the spatial structure of some set of face images exhibited substantial mirror-symmetry. In turn, confounds described in Figure 5 imply a bias to observe mirror-symmetry in higher-tier visual areas with  $RSA_{\text{Euc}}$  and  $RSA_{\text{corrDem}}$ , ironically, even if the spatial structure of the images happened to exhibit antisymmetry. See Ramírez (2017) for details.

### Alternative views on findings of “mirror-symmetry”

It has been argued that studies revealing mirror-symmetry, especially in FFA, suggest a prevalence of mirror-symmetrically tuned neuronal populations like macaque area AL (Axelrod and Yovel, 2012; Kietzmann et al., 2012). It is worth noting, however, that experiments finding mirror-symmetry were prone to reveal low-level biases across face views, and/or mid- and high-level representations of face orientation, but unsuited to study mirror-symmetric face identity representations—since they were by design unable to detect face-identity information. Critically, no evidence has been to our knowledge produced in humans of face-identity information that generalizes across the axis of symmetry. We hence argue that inferences of mirror-symmetric tuning in humans are inconclusive.

Our model demonstrates that, given the low-level biases observed in popular face databases, a network considering uncontroversial aspects of brain architecture and common analysis choices can explain key observations in Table 1. Evidence exists of the influence of low-level features in early-visual areas and FFA (Yue et al., 2011), as well as mid- and high-level features in FFA (reviewed in Ramírez, 2018). It remains unclear if this list includes reflection-invariant neuronal populations.

One study probing FFA response patterns using a model-selection framework found no evidence of mirror-symmetric tuning and favored the interpretation that human FFA is view-tuned like macaque area ML/MF (Ramírez et al., 2014). This study concluded that FFA may implement a code reflecting

head angular disparity. These conclusions depend on modeling assumptions inspired by monkey research (Wang et al., 1996, 1998) that lack confirmation in humans, such as spatial clustering of similarly view-tuned neurons in FFA.

### Model strengths and limitations

Strengths of our model include its simplicity, ability to explain a range of observations, biologically motivated constraints, and generalization across a range of network densities and databases. However, the primate brain exhibits feedback connections (Salin and Bullier, 1995; Briggs, 2020), category structure (Op de Beeck et al., 2019), lateralization (Rossion, 2014), and neuronal tuning properties that support behavior (Hubel and Wiesel, 1962; Gross et al., 1972; DiCarlo et al., 2012; Afraz et al., 2015). Our model does not incorporate these properties. Ours is not a realistic model of the visual system but a biologically informed device to comprehend the behavior of pattern analyses under minimal assumptions. The model is not directly informative regarding the form-of-tuning of neurons in any visual area. The model does reveal, however, sources of variance that must be considered to draw valid inferences regarding neural population codes with fMRI-MVPA.

Our results do not imply that mirror-symmetry is—or is not—an important neural property supporting viewpoint-invariant face recognition. Future work considering direct electrophysiological recordings is needed to clarify if genuine mirror-symmetry exists in human face-selective areas, and, if so, in which areas. Although low-level biases like those described here extend to the study by Freiwald and Tsao (2010), and evidence suggests that luminance and contrast influence responses in ML/MF (Ohayon et al., 2012; Dubois et al., 2015), it seems unlikely these biases explain the single-cell mirror-symmetric identity effects observed in AL or the viewpoint-invariant face identity representation observed in the AM face patch.

### Overrepresentation of frontal face views

The influence of low-level features in visual cortex is well documented (Andrews et al., 2015). Yue et al. (2011) found that, on average, fMRI responses in EVC and FFA were stronger for frontal than those for lateral face views. From the strong correlation observed between EVC and FFA responses, they argued the latter may be caused by the former. Several studies have observed an overrepresentation of the frontal face views in ventral visual cortex (reviewed in Ramírez, 2018). The question remains whether this overrepresentation is due to low-level features, mid-level features, or endogenous processes (Kastner et al., 1998; Birn et al., 2006; Silver et al., 2007). Our predictions in higher-tier areas remain essentially unchanged if the source of symmetric biases on signal-strength were endogenous.

### Comparing brain measurements and computational models

RSA has been used to compare brain representations and network models and taken to support convolutional neural networks (CNNs) as models of human vision (Khaligh-Razavi and Kriegeskorte, 2014; Cichy et al., 2016). Recent work, however, highlights fundamental differences between how CNNs and humans represent visual information (Serre, 2019; Xu and Vaziri-Pashkam, 2021; Wichmann and Geirhos, 2023). We showed that a model that incorporates biological constraints, but does not compute anything, accounts for inconsistencies across studies due to different image statistics and RSA choices. Incorporating similar constraints in CNNs, and considering the impact of the measurement process on RSA (Ramírez and

Merriam, 2020), may help clarify these discrepancies. Deep networks may also help formulate hypotheses and provide insights about how view-tuning and mirror-symmetry emerge in macaques (Osadchy et al., 2006; Leibo et al., 2017; Farzmaḥdi et al., 2023). The low-level biases and constraints explored here importantly bear on the interpretation of such models.

### MVPA: interpretational challenges

MVPA encompasses numerous methods that consider the multivariate structure of neuroimaging signals when studying brain function (Haynes and Rees, 2006; Norman et al., 2006; Tong and Pratte, 2012). Key challenges include how to estimate and normalize brain patterns, choose dissimilarity measure, and compare empirical and model DSMs (Misaki et al., 2010; Diedrichsen et al., 2011; Smith et al., 2011; Garrido et al., 2013; Davis et al., 2014; Ramírez et al., 2014; Diedrichsen and Kriegeskorte, 2017; Ramírez, 2017; Cai et al., 2019; Ritchie et al., 2021; Kaniuth and Hebart, 2022). Here, we focused on the impact of cocktail demeaning and pattern dissimilarity measure. While this study is not prescriptive, we have argued that angular and Euclidean distances provide fundamentally different descriptions of the data, and angular distances uniquely suited to reveal information regarding the form-of-tuning of indirectly measured neural populations (Ramírez, 2018). Evidence also indicates that, when representing neural population responses as vectors, vector angles—and not their lengths—carry virtually all available face identity and orientation information in macaque face-patches (Meyers et al., 2015). Since data demeaning across conditions changes angular relationships among pattern-vectors and conflates information about signal strength and spatial structure, we recommend avoiding this transformation prior to RSA, especially if signal imbalances are present in the data.

### Conclusion

We proposed a parsimonious model considering feedforward projections, cortical magnification, interhemispheric crossings, and a gain field—an often-neglected aspect of the fMRI measurement process. This model accounted for a range of unexplained similarities and inconsistencies across fMRI-MVPA studies investigating facial viewpoint representations in humans. This account was broadly confirmed when analyzing relevant fMRI data. Our results suggest an important source of current disagreement is due to analysis choices. We do not claim that luminance and contrast are necessarily the sole explanation for all published fMRI results. We claim that low-level imbalances across conditions are sufficient to account for findings of mirror-symmetry in humans. We demonstrate how choice of distance measure and data demeaning can lead to the impression that results across studies are inconsistent, when instead they offer different descriptions of the same process. Our results also suggest low-level feature imbalances across conditions may have been mistaken as evidence of mirror-symmetrically tuned neuronal populations.

### References

- Afraz A, Boyden ES, DiCarlo JJ (2015) Optogenetic and pharmacological suppression of spatial clusters of face neurons reveal their causal role in face gender discrimination. *Proc Natl Acad Sci U S A* 112:6730–6735.
- Amano K, Wandell BA, Dumoulin SO (2009) Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. *J Neurophysiol* 102:2704–2718.
- Andrews TJ, Watson DM, Rice GE, Hartley T (2015) Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *J Vis* 15:3.
- Anzellotti S, Fairhall SL, Caramazza A (2014) Decoding representations of face identity that are tolerant to rotation. *Cereb Cortex* 24:1988–1995.
- Axelrod V, Yovel G (2012) Hierarchical processing of face viewpoint in human visual cortex. *J Neurosci* 32:2442–2452.
- Bandettini PA, Wong EC (1997) A hypercapnia-based normalization method for improved spatial localization of human brain activation with fMRI. *NMR Biomed* 10:197–203.
- Berlucchi G (2014) Visual interhemispheric communication and callosal connections of the occipital lobes. *Cortex* 56:1–13.
- Birn RM, Diamond JB, Smith MA, Bandettini PA (2006) Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *NeuroImage* 31:1536–1548.
- Briggs F (2020) Role of feedback connections in central visual processing. *Annu Rev Vis Sci* 6:313–334.
- Cai MB, Schuck NW, Pillow JW, Niv Y (2019) Representational structure or task structure? Bias in neural representational similarity analysis and a Bayesian method for reducing bias. *PLoS Comput Biol* 15:e1006299.
- Cichy RM, Khosla A, Pantazis D, Torralba A, Oliva A (2016) Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci Rep* 6:27755.
- Davis T, LaRocque KF, Mumford JA, Norman KA, Wagner AD, Poldrack RA (2014) What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *NeuroImage* 97:271–283.
- Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4:2051–2062.
- DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? *Neuron* 73:415–434.
- Diedrichsen J, Kriegeskorte N (2017) Representational models: a common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS Comput Biol* 13:e1005508.
- Diedrichsen J, Ridgway GR, Friston KJ, Wiestler T (2011) Comparing the similarity and spatial structure of neural representations: a pattern-component model. *Neuroimage* 55:1665–1678.
- Dubois J, de Berker AO, Tsao DY (2015) Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J Neurosci* 35:2791–2802.
- Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. *NeuroImage* 39:647–660.
- Duncan RO, Boynton GM (2003) Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron* 38:659–671.
- Farzmaḥdi A, Zarco W, Freiwald W, Kriegeskorte N, Golan T (2023) Emergence of brain-like mirror-symmetric viewpoint tuning in convolutional neural networks. 2023.01.05.522909. Available at: <https://www.biorxiv.org/content/10.1101/2023.01.05.522909v1>
- Feilong M, Nastase SA, Guntupalli JS, Haxby JV (2018) Reliable individual differences in fine-grained cortical functional architecture. *NeuroImage* 183:375–386.
- Flack TR, Harris RJ, Young AW, Andrews TJ (2019) Symmetrical viewpoint representations in face-selective regions convey an advantage in the perception and recognition of faces. *J Neurosci* 39:3741–3751.
- Foster C, Zhao M, Bolkart T, Black MJ, Bartels A, Bühlhoff I (2022) The neural coding of face and body orientation in occipitotemporal cortex. *NeuroImage* 246:118783.
- Freiwald WA, Tsao DY (2010) Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330:845–851.
- Gagnon L, et al. (2015) Quantifying the microvascular origin of BOLD-fMRI from first principles with two-photon microscopy and an oxygen-sensitive nanoprobe. *J Neurosci* 35:3663–3675.
- Garrido L, Vaziri-Pashkam M, Nakayama K, Wilmer J (2013) The consequences of subtracting the mean pattern in fMRI multivariate correlation analyses. *Front Neurosci* 7:174.
- González Ballester MÁ, Zisserman AP, Brady M (2002) Estimation of the partial volume effect in MRI. *Med Image Anal* 6:389–405.
- Grill-Spector K, Malach R (2004) The human visual cortex. *Annu Rev Neurosci* 27:649–677.
- Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the macaque. *J Neurophysiol* 35:96–111.

- Guntupalli JS, Wheeler KG, Gobbini MI (2017) Disentangling the representation of identity from head view along the human face processing pathway. *Cereb Cortex* 27:46–53.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430.
- Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, Gobbini MI, Hanke M, Ramadge PJ (2011) A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72:404–416.
- Haynes J-D, Rees G (2006) Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 7:523–534.
- Hemond CC, Kanwisher NG, de Beeck HPO (2007) A preference for contralateral stimuli in human object- and face-selective cortex. *PLoS One* 2:e574.
- Henriksson L, Nurminen L, Hyvärinen A, Vanni S (2008) Spatial frequency tuning in human retinotopic visual areas. *J Vis* 8:5.
- Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160:106–154.
- Kaniuth P, Hebart MN (2022) Feature-reweighted representational similarity analysis: a method for improving the fit between computational models, brains, and behavior. *NeuroImage* 257:119294.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
- Kastner S, De Weerd P, Desimone R, Ungerleider LG (1998) Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science* 282:108–111.
- Khaligh-Razavi S-M, Kriegeskorte N (2014) Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput Biol* 10:e1003915.
- Kietzmann TC, Swisher JD, König P, Tong F (2012) Prevalence of selectivity for mirror-symmetric views of faces in the ventral and dorsal visual pathways. *J Neurosci* 32:11763–11772.
- Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:1–28.
- Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A (2010) Presentation and validation of the Radboud Faces Database. *Cogn Emot* 24:1377–1388.
- Leibo JZ, Liao Q, Anselmi F, Freiwald WA, Poggio T (2017) View-tolerant face recognition and Hebbian learning imply mirror-symmetric neural tuning to head orientation. *Curr Biol* 27:62–67.
- Lundqvist D, Flykt A, Öhman A (1998) The Karolinska directed emotional faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet, ISBN 91-630-7164-9. Available at: <https://doi.apa.org/doiLanding?doi=10.1037%2F10732-000>
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* 92:8135–8139.
- Marr D (1982) *Vision*, 3rd ed. San Francisco: W.H. Freeman.
- McCarthy G (1997) Face-specific processing in the human fusiform gyrus. *J Cogn Neurosci* 9:605–610.
- Meyers EM, Borzello M, Freiwald WA, Tsao D (2015) Intelligent information loss: the coding of facial identity, head pose, and non-face information in the macaque face patch system. *J Neurosci* 35:7069–7081.
- Misaki M, Kim Y, Bandettini PA, Kriegeskorte N (2010) Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *NeuroImage* 53:103–118.
- Moeller S, Freiwald WA, Tsao DY (2008) Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science* 320:1355–1359.
- Natu VS, Jiang F, Narvekar A, Keshvari S, Blanz V, O'Toole AJ (2010) Dissociable neural patterns of facial identity across changes in viewpoint. *J Cogn Neurosci* 22:1570–1582.
- Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10:424–430.
- Ohayon S, Freiwald WA, Tsao DY (2012) What makes a cell face selective? The importance of contrast. *Neuron* 74:567–581.
- Op de Beeck HP, Pillot I, Ritchie JB (2019) Factors determining where category-selective areas emerge in visual cortex. *Trends Cogn Sci* 23:784–797.
- Osadchy M, Le Cun Y, Miller ML (2006) Synergistic face detection and pose estimation with energy-based models. In: *Toward category-level object recognition* (Ponce J, Hebert M, Schmid C, Zisserman A, eds), pp 196–206. Berlin, Heidelberg: Springer; Lecture Notes in Computer Science.
- Petitjean M (2021) Symmetry, antisymmetry, and chirality: use and misuse of terminology. *Symmetry* 13:603.
- Poggio TA, Anselmi F (2016) *Visual cortex and deep networks: learning invariant representations, illustrated edition*. Cambridge, Massachusetts: The MIT Press.
- Polimeni JR, Fischl B, Greve DN, Wald LL (2010) Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. *NeuroImage* 52:1334–1346.
- Puce A, Allison T, Asgari M, Gore JC, McCarthy G (1996) Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J Neurosci* 16:5205–5215.
- Ramírez FM (2017) Representational confusion: the plausible consequence of demeaning your data. Available at: <https://www.biorxiv.org/content/10.1101/195271v1>
- Ramírez FM (2018) Orientation encoding and viewpoint invariance in face recognition: inferring neural properties from large-scale signals. *Neuroscientist* 24:582–608.
- Ramírez FM, Cichy RM, Allefeld C, Haynes J-D (2014) The neural code for face orientation in the human fusiform face area. *J Neurosci* 34:12155–12167.
- Ramírez F, Cichy RM, Haynes J-D (2010) Orientation-encoding in the FFA is selective to faces: evidence from multivoxel pattern analysis. [Abstract] *J Vis* 10:669.
- Ramírez FM, Merriam EP (2020) Forward models of repetition suppression depend critically on assumptions of noise and granularity. *Nat Commun* 11:4732.
- Ramírez FM, Revsine C, Merriam EP (2020) What do across-subject analyses really tell us about neural coding? *Neuropsychologia* 143:107489.
- Revsine C, Merriam EP, Ramírez FM (2020) Low-level features, view tuning, and mirror symmetry: a parsimonious model accounts for commonalities and inconsistencies across neuroimaging studies. [Abstract] *J Vis* 20:1387–1387.
- Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2:1019–1025.
- Ritchie JB, Bracci S, Op de Beeck H (2017) Avoiding illusory effects in representational similarity analysis: what (not) to do with the diagonal. *NeuroImage* 148:197–200.
- Ritchie JB, Lee Masson H, Bracci S, Op de Beeck HP (2021) The unreliable influence of multivariate noise normalization on the reliability of neural dissimilarity. *NeuroImage* 245:118686.
- Rolls ET (2012) Invariant visual object and face recognition: neural and computational bases, and a model, VisNet. *Front Comput Neurosci* 6:35.
- Rossion B (2014) Understanding face perception by means of human electrophysiology. *Trends Cogn Sci* 18:310–318.
- Rossion B, Dricot L, Goebel R, Busigny T (2011) Holistic face categorization in higher order visual areas of the normal and prosopagnosic brain: toward a non-hierarchical view of face perception. *Front Hum Neurosci* 4:225.
- Sabuncu MR, Singer BD, Conroy B, Bryan RE, Ramadge PJ, Haxby JV (2010) Function-based intersubject alignment of human cortical anatomy. *Cereb Cortex* 20:130–140.
- Salin PA, Bullier J (1995) Corticocortical connections in the visual system: structure and function. *Physiol Rev* 75:107–154.
- Schmid F, Barrett MJP, Jenny P, Weber B (2019) Vascular density and distribution in neocortex. *NeuroImage* 197:792–805.
- Sergent J, Ohta S, MacDonald B (1992) Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain* 115:15–36.
- Serre T (2019) Deep learning: the good, the bad, and the ugly. *Annu Rev Vis Sci* 5:399–426.
- Serre T, Wolf L, Poggio T (2005) Object recognition with features inspired by visual cortex. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 2, pp 994–1000.
- Silver MA, Ress D, Heeger DJ (2007) Neural correlates of sustained spatial attention in human early visual cortex. *J Neurophysiol* 97:229–237.
- Smith AT, Kossillo P, Williams AL (2011) The confounding effect of response amplitude on MVPA performance measures. *NeuroImage* 56:525–530.
- Tong F, Pratte MS (2012) Decoding patterns of human brain activity. *Annu Rev Psychol* 63:483–509.
- Tootell RBH, Mendola JD, Hadjikhani NK, Liu AK, Dale AM (1998) The representation of the ipsilateral visual field in human cerebral cortex. *Proc Natl Acad Sci U S A* 95:818–824.

- Tsao DY, Moeller S, Freiwald WA (2008) Comparing face patch systems in macaques and humans. *Proc Natl Acad Sci U S A* 105:19514–19519.
- Vinke LN, Ling S (2020) Luminance potentiates human visuocortical responses. *J Neurophysiol* 123:473–483.
- Wang G, Tanaka K, Tanifuji M (1996) Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* 272:1665–1668.
- Wang G, Tanifuji M, Tanaka K (1998) Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. *Neurosci Res* 32:33–46.
- Weibert K, Flack TR, Young AW, Andrews TJ (2018) Patterns of neural response in face regions are predicted by low-level image properties. *Cortex* 103:199–210.
- Wichmann FA, Geirhos R (2023) Are deep neural networks adequate behavioral models of human visual perception? *Annu Rev Vis Sci* 9:501–524.
- Wickens TD (2014) *The geometry of multivariate statistics*. New York: Psychology Press.
- Xu Y, Vaziri-Pashkam M (2021) Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nat Commun* 12:2065.
- Yamada K, Miyawaki Y, Kamitani Y (2015) Inter-subject neural code converter for visual image representation. *NeuroImage* 113:289–297.
- Yue X, Cassidy BS, Devaney KJ, Holt DJ, Tootell RBH (2011) Lower-level stimulus features strongly influence responses in the fusiform face area. *Cereb Cortex* 21:35–47.