INTERFACE FOCUS

rsfs.royalsocietypublishing.org

Review

Cite this article: Grill-Spector K, Weiner KS, Gomez J, Stigliani A, Natu VS. 2018 The functional neuroanatomy of face perception: from brain measurements to deep neural networks. *Interface Focus* **8**: 20180013. http://dx.doi.org/10.1098/rsfs.2018.0013

Accepted: 8 May 2018

One contribution of 12 to a theme issue 'Understanding images in biological and computer vision'.

Subject Areas:
biomathematics

Keywords:
fMRI, human ventral visual stream, population receptive field

Author for correspondence:
Kalanit Grill-Spector
e-mail: kalanit@stanford.edu

THE ROYAL SOCIETY
PUBLISHING

The functional neuroanatomy of face perception: from brain measurements to deep neural networks

Kalanit Grill-Spector[1,2], Kevin S. Weiner[4,5], Jesse Gomez[3], Anthony Stigliani[1] and Vaidehi S. Natu[1]

[1]Department of Psychology, [2]Stanford Neurosciences Institute, and [3]Stanford Neurosciences Program, School of Medicine, Stanford University, Stanford, CA 94305, USA
[4]Department of Psychology, University of California Berkeley, and [5]Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, CA 94720, USA

KG-S, 0000-0002-5404-9606

A central goal in neuroscience is to understand how processing within the ventral visual stream enables rapid and robust perception and recognition. Recent neuroscientific discoveries have significantly advanced understanding of the function, structure and computations along the ventral visual stream that serve as the infrastructure supporting this behaviour. In parallel, significant advances in computational models, such as hierarchical deep neural networks (DNNs), have brought machine performance to a level that is commensurate with human performance. Here, we propose a new framework using the ventral face network as a model system to illustrate how increasing the neural accuracy of present DNNs may allow researchers to test the computational benefits of the functional architecture of the human brain. Thus, the review (i) considers specific neural implementational features of the ventral face network, (ii) describes similarities and differences between the functional architecture of the brain and DNNs, and (iii) provides a hypothesis for the computational value of implementational features within the brain that may improve DNN performance. Importantly, this new framework promotes the incorporation of neuroscientific findings into DNNs in order to test the computational benefits of fundamental organizational features of the visual system.

1. Introduction

A central goal in cognitive and computational neuroscience is to understand how processing within the ventral visual stream enables rapid and robust recognition and classification of the visual input. Visual recognition is thought to be mediated by a series of serial computations that form a processing stream referred to as the ventral visual processing stream [1,2]. The ventral visual processing stream emerges in V1—the first cortical visual area that resides in the calcarine sulcus [3]—through a series of occipital visual areas, and ends in high-level visual regions in ventral temporal cortex (VTC), whose activation predicts visual perception and recognition [4–8].

Recent neuroscientific discoveries have significantly advanced understanding of the function, structure and computations along the ventral stream processing hierarchy, revealing rich detail about their anatomical implementation, representations and computations (see reviews [9–13]). By anatomical implementation, we mean the physical features of the cortical tissue that act as the substrates performing the computation that produces accurate behaviour. Two important insights have emerged from neuroscience research: (i) the functional organization of the ventral visual stream is structured and (ii) it is reliable across individuals. That is, functional regions are consistently organized with respect to the cortical folding not only in V1 [3], but across the ventral stream more generally [14–17]. For example, the locations of retinotopic areas that contain maps of the visual field (V1-VO1, figure 1*a*–*c*) and face-selective regions (IOG-faces, pFus-faces, mFus-faces, figure 1*c*) are consistently arranged relative to the cortical folding and relative to each other [16,19,20]. These types of findings have led researchers to ask new questions such as (i) how do structural factors such as the underlying microarchitecture and white matter connections constrain the functional organization of the ventral stream? (ii) What is the computational purpose of this functional neural architecture?

In parallel, significant advances in computational models including hierarchical deep neural networks (DNNs) and technological advances that enable training DNNs using large and labelled image sets [21] have brought machine performance in recognition and classification of visual images to a level that rivals human performance [18,22–24]. This computational work has led to two important insights: (i) neurally inspired architectures trained with millions of images can produce optimal, human-like performance [22,23] and (ii) DNNs that learn by optimizing a behaviourally relevant cost function—such as categorization—better predict neural responses and representations in the primate and human brain, respectively, compared to other DNNs [18,25,26].

Because of these exciting recent advancements, this is an excellent time for the field of computational neuroscience to leverage advances in DNNs and to use them as a tool to probe the human visual system [27]. This will allow for a more mechanistic understanding of particular computations at different stages of the processing hierarchy and will provide crucial insights to the computational benefits of specific neural implementational features. Furthermore, perturbing aspects of the computational architecture will enable probing the necessity and sufficiency of specific neural implementational features for particular behaviours. Together, this can lead not only to foundational knowledge, but also to new approaches that could build predictions from computational models that may help rectify deficiencies and maldevelopment of the visual system.

To achieve these important goals, it is necessary for the field to implement and test neurally accurate computational models of the human visual system rather than models that are loosely 'neurally inspired'. Therefore, the goal of this review is to use a model system within the ventral steam—the ventral face network—to illustrate how this goal can be achieved. We chose to focus on the ventral face network for several reasons: (i) it is a well-understood and studied system in both human [10,11,28–45] and non-human primates [46–56], (ii) functional regions in VTC which are causally involved in face recognition can be identified within each individual using functional magnetic resonance imaging (fMRI) [19,20,28,30], and (iii) the output computation of this system can be well defined in several levels of specificity ranging from categorizing a stimulus as a face to identifying a particular person (e.g. 'this is Angela Merkel'). Thus, this review begins with a brief overview of the face recognition system in the human brain. The rest of the review is arranged in sections that describe specific neural implementational features of the ventral face network. For each feature, we consider similarities and differences between the functional architecture of the brain and DNNs, as well as provide a hypothesis for the computational value of this feature.

2. The ventral face network

To identify face-selective regions in the brain, participants are scanned in an fMRI scanner as they view faces and a variety of other stimuli such as body parts, objects, places and printed characters. In each subject, voxels in the ventral aspects of occipital and temporal cortex that respond significantly more strongly to faces than other stimuli are identified as face-selective. As shown in an example subject's inflated cortical surface (figure 1*c*), there are three face-selective clusters in the ventral visual stream, found bilaterally. One cluster is located in the inferior occipital gyrus (IOG) and is called IOG-faces (also referred to as the occipital face area [57]). A second cluster is located on the posterior-lateral aspect of the fusiform gyrus and is called pFus-faces [19]. A third patch is located on the lateral fusiform gyrus, about 1–1.5 cm anterior to pFus-faces, and tends to overlap the anterior tip of the mid-fusiform sulcus (MFS). This patch is referred to as mFus-faces [19]. In fact, in the right hemisphere, a 1 cm disc aligned with the anterior tip of the right MFS identifies approximately 80% of the face-selective voxels in the right mFus-faces [16]. pFus-faces and mFus-faces are often lumped together and referred to as the fusiform face area (FFA [28]). A characteristic of these ventral face-selective regions is that they respond to faces significantly more strongly compared to other stimuli [28,30], and this preference for faces is maintained across formats [29,58–61]. That is, both photographs and line drawings of faces evoke higher responses than photographs and line drawings of common objects. Selectivity to faces is also maintained when low-level features of the visual input are matched across faces and control stimuli (e.g. face silhouettes generate higher responses than shape silhouettes that are matched in contrast and area).

Ventral face-selective regions are thought to receive inputs from earlier retinotopic areas V1, V2, V3 and hV4 [62–64]. These earlier areas are labelled by their order in the visual processing hierarchy [62]. Each of these visual areas contains a map of the visual field (where the left hemifield is represented in the right hemisphere and vice versa). Retinotopic visual areas are thought to be connected to each other and also to the ventral face regions via axons [62,63]. Long-range axonal connections tend to be myelinated and form white matter tracts. Thus, some of the inputs from earlier visual areas to face-selective regions include portions of the inferior longitudinal fasciculus [65–67] (a large tract that connects the occipital lobe to the inferior aspect of the temporal lobe [68]). Additionally, ventral face-selective regions also have white matter connections to visual regions in the parietal cortex through vertical fasciculi such as the vertical occipital fasciculus (VOF [69–71]) and posterior arcuate fasciculus [70]. These vertical connections are thought to facilitate top-down modulations from the parietal-attention network to ventral regions [72]. However, in this review, we will concentrate on the feed-forward connections of the ventral face network.

Understanding this organization is useful for generating a tentative schematic of the processing hierarchy of the ventral face network (figure 1*b*). However, this is not often how the ventral stream processing hierarchy is portrayed in 'neurally inspired' DNNs. A typical DNN of the ventral stream based on the macaque visual system (shown in figure 1*a*) is portrayed as a feed-forward architecture progressing from V1

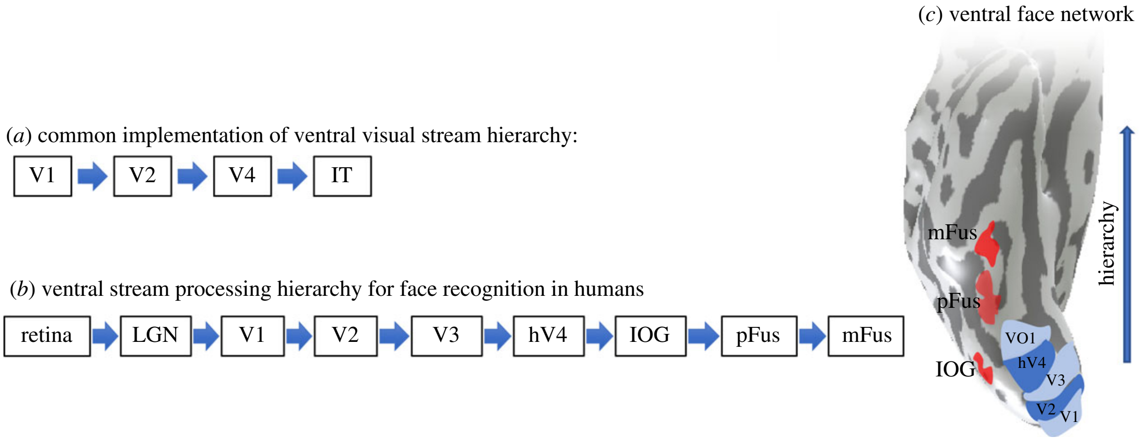

(a) common implementation of ventral visual stream hierarchy:
V1
V2
V4
IT
(b) ventral stream processing hierarchy for face recognition in humans
retina
LGN
V1
V2
V3
hV4
IOG
pFus
mFus
(c) ventral face network
mFus
pFus
IOG
VO1
hV4
V3
V2
V1
hierarchy

Figure 1. Ventral stream processing hierarchy for face recognition in humans. (*a*) A common ventral visual stream hierarchy based on the macaque visual system, implemented or referred to in the DNN literature. This hierarchy is adapted from [18], though some models begin in the retina [13]. (*b*) The ventral stream visual hierarchy of the human ventral face network. In the manuscript, we will only describe cortical regions starting from V1. This is a tentative suggestion based on present understanding of visual areas in the human brain (see 1*c*), but could be refined in future research when new knowledge (such as understanding the full connectivity pattern including feedback connections and bypass routes) will update this schematic. (*c*) Visualization of the ventral face network on an inflated cortical surface of an example participant showing the ventral aspect of occipito-temporal cortex (sulci in dark grey, gyri in light grey). Retinotopic areas are shown in shades of blue and labelled V1 to VO1. Face-selective regions are shown in shades of red and include IOG-faces (on the inferior occipital gyrus), pFus-faces (on the posterior fusiform gyrus) and mFus-faces (on the mid-fusiform gyrus). (Online version in colour.)

to V2 to V4 to IT (IT, or infero-temporal in the macaque, is thought to be homologous to human VTC). However, there are two main differences between the commonly implemented DNN and the human ventral stream. First, V3 is missing. This omission may be due to the fact that in macaque, V3 is substantially smaller than either V2 or V4 and there are direct white matter connections from V2 to V4. However, in the human brain, V3 is both equivalent in size to V2 [73,74] and larger than hV4 (figure 1*c*). Second, IT is often represented in DNN schematics as a single area. In the macaque, IT contains multiple subdivisions [55,75–80], and in humans, VTC is divided into several cytoarchitectonic areas [16,81–84], which contain more than 10 visual regions including: (i) two face-selective regions, pFus and mFus, figure 1*c*, (ii) additional domain-specific regions selective for places [85,86], bodies [87,88], objects [89] and characters/words [90], and (iii) several retinotopic areas: VO1/2 [91]; PHC1/2 [92]. Thus, we propose that the first step in building a neurally accurate feed-forward DNN for the human face recognition system is to include all the relevant areas in the human brain. Consistent with this idea, in the present manuscript, we will consider the following ventral face network: V1 ! V2 ! V3 ! hV4 ! IOG ! pFus ! mFus (figure 1*b*).

Why are we focusing only on the feed-forward aspect of this network? There are several reasons. First, humans can classify a stimulus as a face in less than 100 ms and recognize the identity of the face in approximately 150 ms [93,94]. This fast processing has prompted researches to suggest that face recognition does not necessitate top-down information and can be accomplished with fast, feed-forward processing. Second, face-selective responses in the fusiform gyrus emerge within 100–170 ms [38,95–98]. Third, as standard DNNs have a feed-forward architecture, we first compare them to the feed-forward components of the human visual system. Once these are well-understood, subsequent analyses will elucidate the role of non-hierarchical connections including the modulatory role of top-down connections from the parietal lobe [69,70,72] to the ventral stream, as well as the role of bypass connections [64].

As illustrated in table 1*a*,*b* and figure 1, there are some commonalities in the basic neural implementation of the ventral face network and DNNs. Critically, both types of networks enable hierarchical and feed-forward processing, which are thought to support two important computational benefits. First, the universal approximation theorem [99] has shown that these types of architectures can approximate any complex continuous function relating the input (here, the visual input) to the output (here, face recognition). Second, feed-forward processing with simple linear–nonlinear operations (which we will elaborate below) allows fast computations and, consequently, rapid performance (in our case, face recognition). Now that we have a foundation regarding the architecture of the ventral face network, we next turn to the computations that this structure produces.

3. Basic computational unit in the visual system: receptive fields

In the human visual system, the basic computation is performed by receptive fields. A receptive field (RF) is the region in visual space that is processed by a neuron. Since neurons with similar RFs are spatially clustered, with fMRI we can measure the population receptive field (pRF)—the region in the visual field that is processed by the population of neurons in a voxel. RFs are often modelled by spatial filters that have linear–nonlinear operations. Example receptive fields that have been used to model responses in the visual system include Gaussians, difference of Gaussians and Gabor filter banks. These filtering operations are often followed by a nonlinearity such as a normalization, rectification or a compressive exponential nonlinearity [100–102].

These types of RF models have inspired the implementation of filters within DNNs. Indeed, each layer of a DNN contains a series of linear filter banks. Filters in each layer are applied

Table 1. Comparison between several major characteristics of human ventral face network and deep neural networks.

property	human brain	deep neural network	hypothesized utility
a. hierarchical processing	√	√	enables computing of complex functions
b. feed-forward processing	√	√	speed
c. local computations	√	√	parallel processing
d. pRF/filter size increases along hierarchy	√	√	extraction of useful features
e. pRF/filter size increases with eccentricity	√	✗7	solution to limited brain size
f. adjustable pRFs/filters	√	✗7	task-optimized processing
g. learned pRFs/filters	√	√	flexibility; optimization for task and natural statistics
h. spatio-temporal pRFs/filters	√	✗7	capture dynamics of natural environment

uniformly on the input (image or output of prior layer) using a convolution operation. The output of the convolution can be followed by several mathematical operations to mimic neural responses: a thresholding nonlinearity (e.g. rectification or sigmoid), then pooling and, finally, normalization. Thus, filters in DNNs perform local operations on the image akin to those of receptive field models. The computations by pRFs/filters enable local, parallel processing of the image, which, in turn, increases computational efficiency (table 1*c*).

PRFs in the human brain have four fundamental characteristics that are interesting to consider when comparing to filters in DNNs. First, pRFs in the right hemisphere are centred in the left visual field, and those in the left hemisphere are centred in the right visual field. This is referred to as processing of the *contralateral* visual field. In other words, to increase parallel processing, the brain splits the visual input into two halves, each processed in a different hemisphere. DNNs typically process the entire image, though some implementations split processing across more than one graphics processing unit [22].

Second, mean pRF size increases across the hierarchy of the ventral face network (figure 2*a*). The smallest pRFs are in V1 and the largest pRFs are in face-selective regions. For example, pRFs in face-selective regions are on average four times larger than those in V1 (figure 2*a*). This characteristic is also present in DNNs due to both the pooling operation and the repeated use of local convolutional filters. This results in a systematic increase in the extent of the visual image processed by filters as one ascends stages of the DNN. This increase in pRF/filter size is hypothesized to allow neurons/filters in higher stages to process information across several features, and perhaps even the entire object, rather than just local features as is the case for processing in lower stages of the network.

To give the reader an intuition of how mean pRF sizes in the ventral face network (figure 2*a*) relate to a real-life example, let us consider an example in which a face is viewed from a typical viewing distance (approx. 1 m away) and determine what facial features are processed by pRFs in different visual regions of the ventral face network. In this example, illustrated in figure 2*c*, a V1 pRF processes only the corner of the eye, a hV4 pRF processes the eye and the top of the nose, and a mFus-faces pRF processes the entire face. This example shows that the increase in pRF/filter size across the ventral visual hierarchy allows higher stages of the hierarchy to process more useful features for recognition (table 1*d*).

Third, in both the human and non-human primate visual system, RF size and consequently pRF size, increase with eccentricity [102–104] (figure 2*c*). That is, starting from the retina, and continuing throughout the entire processing hierarchy, RF size is not constant in a given region. Rather, both RFs and pRFs are smallest near fixation (centre of gaze) and increase roughly linearly with eccentricity (figure 2*b*). By contrast, filter size in DNNs is constant across each layer of the network. One reason why pRF size scales with eccentricity in the human and primate brain, but not in DNNs, may be limited resources. That is, the brain may need to optimize visual resolution given limited physical space as well as limited metabolic resources. The brain's solution to these limitations is to provide more resolution (smaller RFs) at the centre of gaze at the expense of less resolution (larger RFs) in the periphery (table 1*e*).

Fourth, in the human brain, pRFs in face-selective regions have a foveal bias. In face-selective regions, like in earlier visual areas, pRF centres are in the contralateral visual field (e.g. pRFs in the left hemisphere are centred in the right visual field, figure 3*a*). However, in face-selective regions, almost all of these pRFs overlap the fovea (figure 3*a*). We refer to this phenomenon as foveal bias. Given that pRFs in face-selective regions are large and overlap the fovea, this enables them to process information across both visual fields. Additionally, as one ascends from face-selective IOG, to pFus, to mFus, the foveal bias increases as pRF centres become more concentrated around fixation. Consequently, in face-selective regions, the centre of the visual field is more densely covered by pRFs than the periphery of the visual field [36,106–108].

It is appealing to hypothesize how this tiling of the visual field by pRFs in face-selective regions may relate to behaviour. One interesting behaviour is how people look at faces. A large literature indicates that during recognition, people tend to fixate on the centre of the face [109–113], as shown for the example in figure 3*b* (but see [114,115]). This fixation behaviour places pRFs in face-selective regions on the part of the face that has the most informative features for recognition [116–118]—that is, the eyes and nose.

4. PRFs in face-selective regions are modulated by the task

One interesting question is whether pRFs in the visual system are fixed or are modulated by task and behavioural goals. Several results show that attention and task may modulate pRF properties and this modulation seems to increase across the visual processing hierarchy [36,119,120]. Namely,

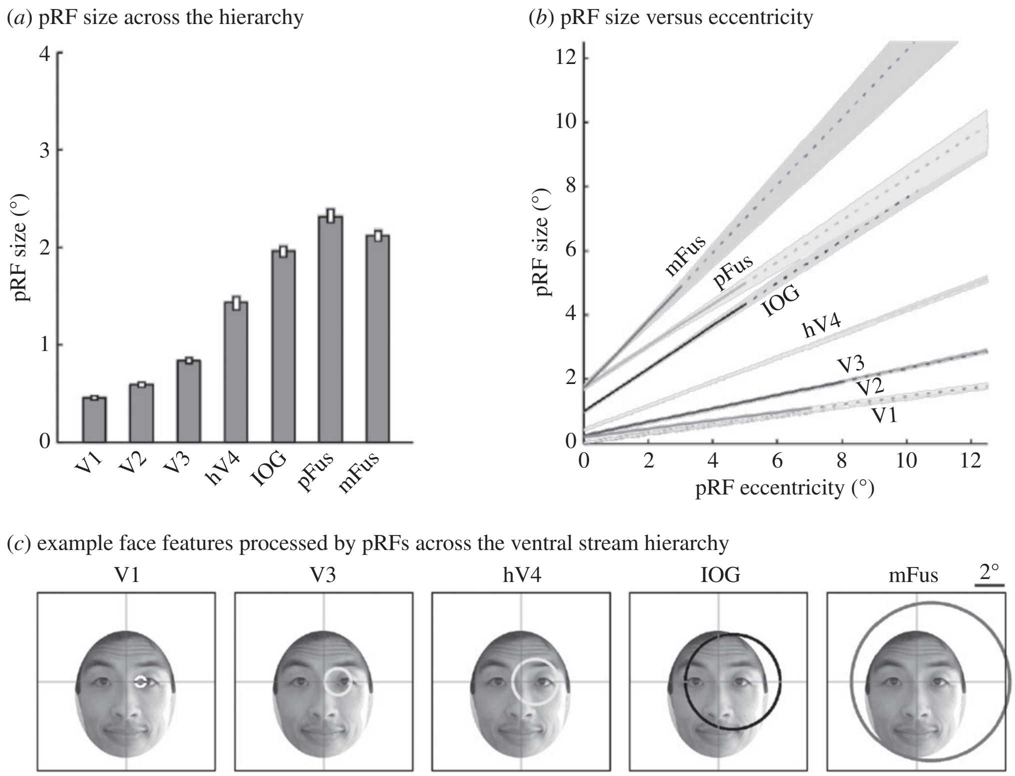


Figure 2. pRF properties across the ventral face network hierarchy. (*a*) Mean pRF size measured across the central 7° of each visual area. (*b*) There is a linear relationship between pRF size and pRF eccentricity across the ventral face network hierarchy. The slopes of lines relating pRF size and eccentricity increase across the processing hierarchy. (*c*) Example pRFs from the ventral face network. In each region, we illustrate a pRF centred at a 2° eccentricity on a face that is at typical viewing distance (approx. 1 m). The crosshair indicates the fixation point. Figure is adapted from [34].

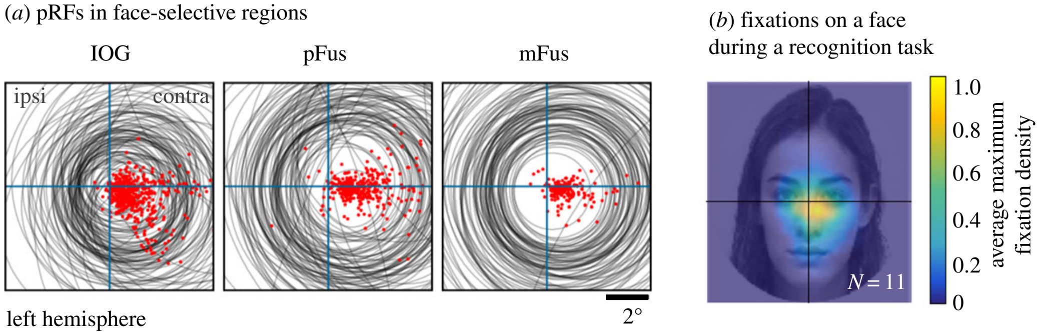


Figure 3. pRF properties in face-selective regions may affect the way people look at and fixate on faces. (*a*) Tiling of the visual field by pRFs in face-selective regions. pRFs are indicated by the grey circles, and their centres by the red dots. Ascending from face-selective IOG, to pFus, to mFus, pRFs become larger and become more concentrated on the centre of gaze. Adapted from [36]. (*b*) Fixation density on an example face during a face recognition task. Data are averaged across 11 adults. Colourbar indicates average maximum fixation density. Adults tend to fixate on the centre of the face when performing face recognition tasks. This behaviour puts the combined visual field coverage of pRFs in face-selective regions on informative facial features. Adapted from [105].

attention has a more profound effect on pRFs in higher levels than lower levels of the hierarchy.

In our experiments, we tested if pRFs in the ventral face network are modulated by the task [36]. To do so, we measured pRFs by showing faces randomly in 25 locations while subjects centrally fixated on a stream of digits under two tasks: a digit task and a face task. In the digit task, participants indicated via a button press if two successively presented digits were the same, and in the face task, participants indicated if two successively presented images were of the same person.

Our results revealed three findings. First, attention to peripheral faces relative to central fixation increased pRF eccentricity in face-selective regions, but not early visual areas. That is, during the face task, pRFs in face-selective regions were further from fixation than during the digit task. In contrast, there were no changes to pRF eccentricity across tasks in early visual areas (V1–V3). Second, attention to faces increased pRF size in face-selective regions, but not early visual areas. In face-selective regions, pRF sizes were substantially larger during the face task than the digit task. For example, in mFus-faces, median pRF size increased from 1.8° in the digit task to 3.4° in the face task. Third, pRF gain in face-selective regions was larger in the face than digit task, but this was not apparent in early visual areas.

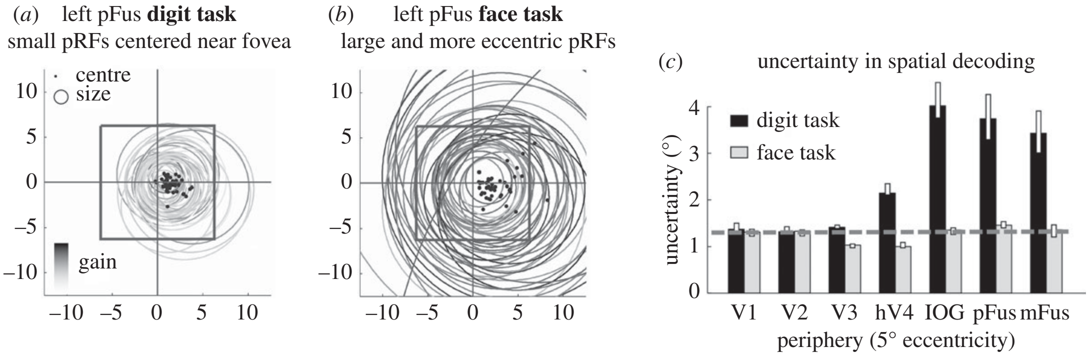


Figure 4. Attention to faces enhances representation and spatial precision in the periphery. (*a*) pRFs of left pFus-faces under the digit task, (*b*) pRFs of left pFus-faces under the face task. In *a* and *b*, pRFs are indicated by the circles, their centres are indicated by black dots, and their gain is indicated by the grey-level intensity (see colourbar). The black square indicates the size of a 5° image. (*c*) Spatial uncertainty in decoding the location of a face compared to an anchor face placed at 5° eccentricity based on the collection of pRFs in each task. Spatial uncertainty is lower during the face task (grey) than the digit task (black). Adapted from [36]. (Online version in colour.)

The combined effects of task on pRF size and eccentricity have a profound impact on the spatial representation of visual space by the collection of pRFs spanning each of the face-selective regions. This effect is illustrated in figure 4*a*,*b*: figure 4*a* illustrates the visual field coverage by pRFs of pFus-faces under the digit task, and figure 4*b* shows the pRFs of the same voxels during the face task. Notably, during the face task, pRFs are more scattered and extend further into the periphery than during the digit task. Thus, the consequence of attention to faces is enhanced representation of the periphery by pRFs of face-selective regions.

To quantify the effect of task on spatial acuity of the neural representation, we used a model-based decoding approach to quantify the spatial uncertainty obtained by pRFs measured under the different tasks. We found a significant four-fold reduction in spatial uncertainty in the periphery (5° eccentricity) in face-selective regions during the face task compared to the digit task (figure 4*c*). In contrast, spatial uncertainty obtained by pRFs in early visual areas remained stable across tasks. Interestingly, the spatial uncertainty obtained by pRFs in face-selective regions in the face task was no greater than that of V1 even though pRFs were substantially larger (figure 4*c*).

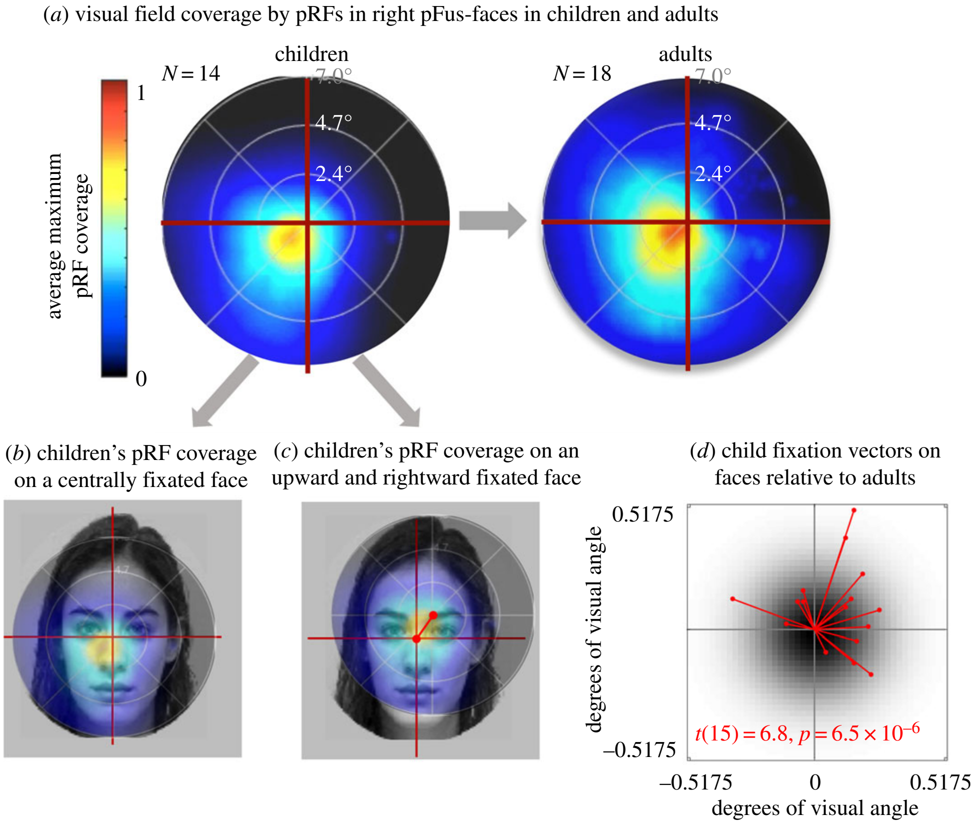
Thus, another difference between the human brain and DNNs is the finding of task-adjustable pRFs in higher stages of the hierarchy (table 1*f*). We speculate that this implementational feature allows the brain to adjust pRFs according to task demands and to enable more effective task-relevant processing. This task-based modulation is likely implemented in the brain via top-down connections. One candidate pathway that may facilitate such task-based modulation is the VOF. This white matter tract connects regions in the IPS that are involved in attentional gating with ventral stream regions, such as pFus-faces, thereby modulating responses in the ventral stream [72]. In addition to task-based modulations, experience and development also modify pRFs, which we address in the next section.

5. Both cortical and artificial networks are shaped by experience

One of the big contributions of the DNN literature for understanding biological visual systems is elucidating what types of filters are learned under different tasks. For example, in their seminal paper, Krizhevsky *et al.* [22] showed that training a DNN to categorize natural images generated V1-like oriented and colour-opponent filters in the first stage of their neural network. In other words, training the network to perform a categorization task using real images during training (ImageNet [21]) generated filters in the first convolutional layer that had similar properties to V1 receptive fields (RFs). Likewise, a large body of literature has examined the role of experience in shaping RF properties in V1 in species other than humans [121–124]. While the general retinotopic preference is present in infancy, likely due to wiring, experience is thought to be necessary to fine-tune RF properties of V1 neurons to obtain the adult-like specificity of their size, position and orientation tuning. This ability of DNNs and of the human brain to learn is key, as it gives the system considerable flexibility to learn the natural statistics of the visual world as well as to optimize the filters for extracting task-relevant properties (table 1*g*).

Presently, most DNNs use supervised learning (e.g. by labelling the category of training images) and algorithms such as back-propagation [125], which optimize a task-relevant cost function to learn relevant information. While humans may receive some supervised learning (e.g. a mother may name objects as they speak to their babies), it is thought that neurons in the brain can also fine tune their response properties via unsupervised learning from the natural statistics. Thus, a goal for computational modelling would be to develop a family of DNNs that learns from unsupervised training to better model biological visual systems.

Notably, recent evidence suggests that the development of pRFs in higher visual areas, such as face-selective regions, continues well past infancy and during childhood [105] even as pRFs in V1 and other early visual areas are adult-like by age 5 [105,126,127]. In a recent study, we measured pRF properties and the visual field coverage of pRFs in face-selective regions of school-age children and adults [105]. We found substantial developmental changes in the visual field coverage in face-selective regions from childhood to adulthood. As illustrated in figure 5*a*, the right pFus-faces of children shows a foveal bias (higher density of the visual field coverage around the centre of gaze), and a coverage of the left, lower visual field. In adults, right pFus-faces also shows a foveal bias. However, compared to children, the visual field coverage


(*a*) visual field coverage by pRFs in right pFus-faces in children and adults

(*b*) children's pRF coverage on a centrally fixated face

(*c*) children's pRF coverage on an upward and rightward fixated face

(*d*) child fixation vectors on faces relative to adults

Figure 5. Development of visual field coverage in face-selective regions correlates with fixation patterns on faces. Adapted from [105]. (*a*) Visual field coverage by pRFs in right pFus-faces averaged across 14 children (left) and 18 adults (right). Colour indicates the average maximum pRF coverage in the central 7°. Crosshairs indicate fixation. (*b*) Placing the visual field coverage of right pFus-faces in children on the centre of the face would place pRF resources in a region without informative features. (*c*) Moving fixation upwards and rightwards (indicated by the red vector) places the visual field coverage of children's pRFs on the region of the face containing informative features. (*d*) Child fixation patterns on 16 faces compared to adults. Fixations are significantly shifted rightwards and upwards.

in adults' right pFus-faces (i) expands to the upper and right (ipsilateral) visual field and (ii) the foveal bias increases. These data show that pRF properties in face-selective regions continue to develop after age 5.

What are the implications of the development of visual field coverage by pRFs? One prediction from our findings is that face viewing behaviour should differ across age groups. In other words, we predict that if pRFs in face-selective regions guide viewing behaviour, then the differing visual field coverage of pFus-faces across age groups would result in differing fixation patterns on faces across age groups. To illustrate this point, consider figure 5*b*, which shows the pRF coverage of children's right pFus-faces superimposed on an example face. Central fixation, as performed by a typical adult, will put the visual field coverage of the child's pFus-faces on the edge of the nose and cheeks, which do not contain useful information for face recognition. In other words, a child presented with the example face should not fixate on the centre of the face as it will place the visual field coverage of pFus-faces outside the region with useful features. Instead, the child should shift their fixation upwards and rightwards (figure 5*c*), as this fixation behaviour will place the visual field coverage of right pFus-faces on informative features for face recognition. It turns out that this is precisely what children do. Comparison of fixation patterns on faces in children and adults indicate that children's fixations on faces are indeed consistently shifted upwards and rightwards compared to adults (figure 5*c*), thus putting the pRFs of face-selective regions on the informative features. A second implication from our results is that fixation patterns on faces, as well as pRFs in face-selective regions, may be shaped by lifelong experience and consequently, may vary across cultures with different stereotypical viewing of faces (e.g. [115]. Future research comparing pRFs across cultures with distinct face viewing norms can address this question.

6. Neural sensitivity to face identify develops from childhood to adulthood

While development of pRFs in face-selective regions is related to face viewing patterns, this development does not explain why face recognition performance in adults is better than in children. We hypothesized that another facet of functional development may be increased neural sensitivity to face identity. Increased neural sensitivity may lead to increased perceptual sensitivity and consequently, better face recognition performance.

To test if neural sensitivity to face identity develops from childhood to adulthood, in a different study [128], we used a parametric fMRI-adaptation (fMRI-A [89,129]) experiment. In adults, responses to repetitions of the same face are lower than responses to different faces, due to neural adaptation [89,129]. Importantly, the level of fMRI-A is dependent on the level of face similarity [130–132]. That is, the more similar

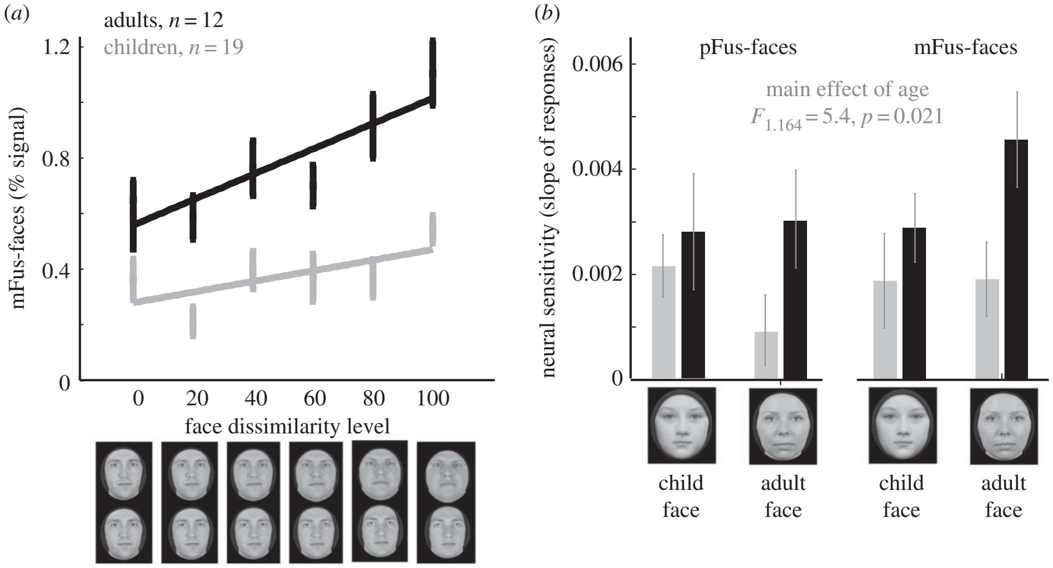


Figure 6. Sensitivity to face identity develops from childhood to adulthood. (*a*) Average response in mFus-faces across 12 adults (19–34 years old, black) and 19 children (5–12 years old, grey) to faces that vary in their level of dissimilarity. The slope of this line indicates sensitivity to face identity. The *x*-axis indicates the dissimilarity between faces in a trial starting from 0 (identical) to 100 (different real-world individuals) in increments of 20%. In order to systematically vary dissimilarity among faces, Natu *et al*. [128] morphed a target face to six different identities and varied the weighting of the source and target faces. In each 4-s trial, subjects viewed six faces from these morphs. In different trials, subjects viewed male and female faces as well as adult and child faces. (b) Slope of the line relating amplitude of response to face-dissimilarity in children (grey) and adults (black) as they viewed adult and child faces. Data in this figure are adapted from [128]; *Error bars*: standard error of the mean.

the faces are, the larger the fMRI-A. Therefore, we designed an experiment in which we systemically varied face similarity and tested if the slope of the function relating neural responses to face dissimilarity (defined as neural sensitivity) varies across age groups [128]. We predicted that if neural sensitivity to faces develops, the slope of this line will be steeper in adults than children. Indeed, that is precisely what we found. Interestingly, this development was specific to the face-selective regions of the ventral face network (figure 6*a*). Further analyses indicated that the neural sensitivity to face identity is also influenced by recent experience and the social salience of faces. In pFus-faces, children had higher neural sensitivity to child than adult faces, and in mFus-faces, adults had higher neural sensitivity to adult faces than child faces (figure 6*b*). Notably, the degree of neural sensitivity was correlated to perceptual discriminability of face identity. That is, subjects with higher neural sensitivity to faces in pFus- and mFus-faces had higher perceptual sensitivity. Together, these data show that both pRFs and the neural sensitivity to face identity develop from childhood to adulthood. Furthermore, this development was coupled with improved perceptual discriminability.

7. Receptive fields in the visual system process changes across both space and time

Finally, another key difference between processing by filters in the brain and filters in DNNs emulating the ventral stream is their temporal sensitivity. Typical DNNs for recognition, categorization and face identification contain temporally-static filters. In contrast, the visual system has dynamic RFs (table 1*h*). For example, electrophysiological recordings in macaque V1 have found that V1 RFs are best understood as spatio-temporal filters [133–137] in which RFs process changes in the visual input across both space and time.

Electrophysiology studies commonly report two types of temporal filters in V1: monophasic and biphasic filters [138–140]. Monophasic temporal filters compute the ongoing sustained visual response—that is, they produce elevated firing when a visual stimulus is present. In contrast, biphasic temporal filters compute the temporal derivative of the visual input, indicating when there is a change in the visual stimulus. Thus, spatio-temporal filters compute time-varying aspects of the visual stimulus. For instance, in V1 they process changes in contrast and/or orientation over time (figure 7).

While initial research on spatio-temporal filters [133,137,138] was focused on understanding properties of neurons that code the direction of visual motion (which are found in V1 and MT), recent evidence suggests that such transient and sustained temporal channels are found not only in V1, but also across the visual system [101,141] including the ventral stream [141]. This finding is somewhat surprising because recognition can be done from brief, static images [93,94,142] and visual motion does not strongly modulate responses in ventral face-selective regions [143]. The combination of this recent evidence leads to the following intriguing question: What is the computational purpose of spatio-temporal filters in the ventral face network and the ventral visual stream more broadly?

We speculate that spatio-temporal filters may serve several computational goals. First, in contrast to artificial DNNs in which the visual input is introduced one image at a time, the visual input in the natural worlds is continuous, except for discontinuities introduced by eye movements. Therefore, spatio-temporal filters may parse the visual input. For example, biphasic temporal filters may be useful

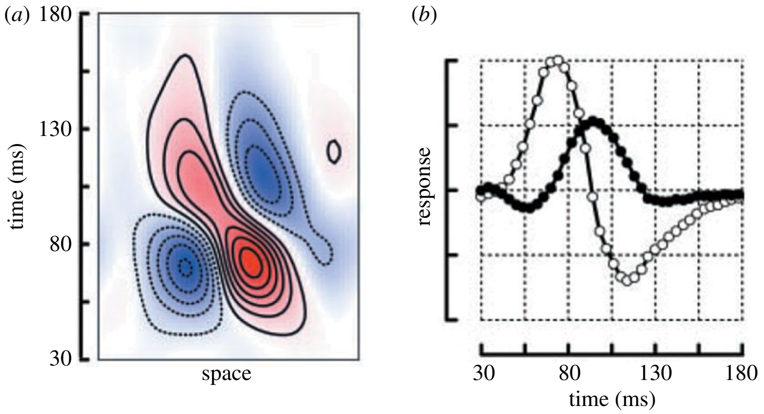


Figure 7. Example spatio-temporal receptive fields (RF) in macaque V1. (*a*) Example spatio-temporal receptive field recorded in macaque V1. This filter has both spatial (*x*-axis) and temporal (*y*-axis) tuning. (*b*) Example temporal characteristic of a monophasic (black) and biphasic (grey) temporal RF in macaque V1. Adapted from [138]. (Online version in colour.)

for detecting novel stimuli (e.g. a new face) and monophasic temporal filters may code sustained aspects of the visual input [141]. Second, spatio-temporal filters may compute correlations across space and time from the visual input that may function to bind incident two-dimensional views of the same object together [144,145] (e.g. linking among different face views belonging to the same individual), which is a process that may be particularly useful for unsupervised learning [145–147]. Third, some items in the world, such as bodies and animate beings, are non-rigid [148]. Thus, spatio-temporal filters may aid in computing dynamic features, which may be particularly useful for recognition of non-rigid stimuli. Therefore, a productive avenue for future DNN research would be to implement dynamic spatio-temporal filters within the DNN architecture to test these hypotheses and to determine the added value of dynamic compared to static filters.

8. Using deep neural networks to test the computational utility of implementational features of the neural architecture

Throughout this review, we described important implementational features of the human ventral face network, compared these features with present DNN architectures, and proposed hypotheses for the computational utilities of various implementational features. These ideas are summarized in table 1. We are hopeful that these neural features will be incorporated into modern DNNs to generate a new class of neurally accurate computational models of the ventral stream and specifically of the face network. To make DNNs neurally accurate, there is a need to implement neural features that are presently absent including: (i) filters that sample the visual field in a non-uniform manner, (ii) filters that can be adjusted to accommodate varying task demands, (iii) temporally dynamic filters, (iv) a correct number of processing stages, and (v) recurrent and top-down connections. Adding these features into DNNs may (i) enhance understanding of the computations along the ventral stream, (ii) likely improve the predicted brain responses to a variety of stimuli, and (iii) provide important insights to the hypothesized utility of various architectural features of the human brain. As the interplay between neuroscience and computer science increases, it is important to consider that comparisons between DNNs and the human brain can be done at many levels. For example, DNNs can be used to predict responses of single neurons or fMRI voxels. Alternatively, one can compare the types of representational spaces emerging in DNNs compared to the brain, or examine if the spatial layouts of these representations are similar to the spatial layouts across the cortical sheet [18,25,26]. We believe that each of these different comparison levels (as well as others that we have not considered) are useful, because they will provide important insights to cortical computations, as well as anatomical and functional constraints that serve as the infrastructure for these computations.

Critically, if these neurally accurate DNNs prove to be better models of brain responses as well as human behaviour compared to standard DNNs, we can use these computational models to test the role of specific implementational features on both brain responses and recognition behaviour. For example, we have shown that pRFs in face-selective regions have a foveal bias and that adults tend to fixate on the centre of the face during recognition. We hypothesized that this viewing behaviour places pRFs of face-selective regions on the informative features for recognition. This hypothesis can be tested by a neurally accurate DNN in which lower layers have filters that scale with eccentricity and higher layers have foveally biased filters. For example, using such a network trained on face recognition, we can test if better recognition occurs when an input image of a face is presented either (a) centrally, at the network's 'fovea' or (b) off-centre.

Another enigma that can be resolved with neurally accurate DNNs is why there are three face-selective regions in the ventral face network and what computational goal they may serve. To investigate this question, one can generate a family of DNNs in which the number of higher layers vary (even as lower layers are held constant). Using this framework, researchers could directly test what features emerge in higher layers, as well as how the number of layers may affect (i) performance, (ii) the efficiency of computations or (iii) the speed and accuracy of learning. Nonetheless, we acknowledge that this comparison will be complex, as there may not be a 1-to-1 correspondence between layers in a DNN to stages (or brain areas) spanning the ventral visual hierarchy.

In sum, neuroimaging research has advanced our understanding regarding the functional architecture of the human ventral face network. Importantly, incorporating these recent findings in up-to-date computational DNNs will further advance the field by providing enhanced understanding of the computational benefits of specific implementational features of the human brain.

Authors' contributions. This review is based on research done in the Vision and Perception Lab at Stanford University directed by K.G.S. K.W. conducted the experiments that determined the anatomical and topological locations of the ventral face network as well as pRF mapping experiments with faces; J.G. and V.N. conducted developmental studies including pRF measurements in children and fMRI-adaptation experiments of face sensitivity. A.S. developed temporal channel models of the visual system. K.G.S. contributed to all these studies. All authors contributed to writing of this review.

Data accessibility. Code related to figure 1 which contains the experiment used to define face-selective regions by contrasting responses to faces versus other images can be found here: https://github.com/VPNL/fLoc. Data and code related to figures 2–4 can be found here: http://kendrickkay.net/vtcdata/. Code related to figure 5 can be found here: https://github.com/VPNL/pRF_Development.

Competing interests. We declare we have no competing interests.

Funding. This research was funded by NIH grant nos. 1ROI1EY02231801A1, 1RO1EY02391501A1 to K.G.S., training grant no. 5T32EY020485 supporting V.N., and the NSF Graduate Research Development Program grant no. DGE-114747 as well as Ruth L. Kirschstein National Research Service Award grant no. F31EY027201 supporting J.G.

Acknowledgements. We thank Michael Barnett, Brianna Jeska and Kendrick Kay who contributed to fMRI experiments described in this review.

References

1. Ungerleider LG, Mishkin M. 1982 Two cortical visual systems. In *Analysis of visual behavior* (eds DJ Ingle, MA Goodale, RJW Mansfield), pp. 549–586. Cambridge, MA: MIT Press.
2. Milner AD, Goodale MA. 1993 Visual pathways to perception and action. *Prog. Brain Res.* **95**, 317–337. (doi:10.1016/S0079-6123(08)60379-9)
3. Benson NC, Butt OH, Datta R, Radoeva PD, Brainard DH, Aguirre GK. 2012 The retinotopic organization of striate cortex is well predicted by surface topology. *Curr. Biol.* **22**, 2081–2085. (doi:10.1016/j.cub.2012.09.014)
4. Parvizi J, Jacques C, Foster BL, Withoft N, Rangarajan V, Weiner KS, Grill-Spector K. 2012 Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci.* **32**, 14 915–14 920. (doi:10.1523/JNEUROSCI.2609-12.2012)
5. Rangarajan V, Hermes D, Foster BL, Weiner KS, Jacques C, Grill-Spector K, Parvizi J. 2014 Electrical stimulation of the left and right human fusiform gyrus causes different effects in conscious face perception. *J. Neurosci.* **34**, 12 828–12 836. (doi:10.1523/JNEUROSCI.0527-14.2014)
6. Tong F, Nakayama K, Vaughan JT, Kanwisher N. 1998 Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* **21**, 753–759. (doi:10.1016/S0896-6273(00)80592-9)
7. Grill-Spector K, Knouf N, Kanwisher N. 2004 The fusiform face area subserves face perception, not generic within-category identification. *Nat. Neurosci.* **7**, 555–562. (doi:10.1038/nn1224)
8. Moutoussis K, Zeki S. 2002 The relationship between cortical activation and perception investigated with invisible stimuli. *Proc. Natl Acad. Sci. USA* **99**, 9527–9532. (doi:10.1073/pnas.142305699)
9. Grill-Spector K, Weiner KS. 2014 The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* **15**, 536–548. (doi:10.1038/nrn3747)
10. Duchaine B, Yovel G. 2015 A revised neural framework for face processing. *Annu. Rev. Vis. Sci.* **1**, 393–416. (doi:10.1146/annurev-vision-082114-035518)
11. Freiwald W, Duchaine B, Yovel G. 2016 Face processing systems: from neurons to real-world social perception. *Annu. Rev. Neurosci.* **39**, 325–346. (doi:10.1146/annurev-neuro-070815-013934)
12. Hong H, Yamins DL, Majaj NJ, DiCarlo JJ. 2016 Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat. Neurosci.* **19**, 613–622. (doi:10.1038/nn.4247)
13. Yamins DL, DiCarlo JJ. 2016 Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* **19**, 356–365. (doi:10.1038/nn.4244)
14. Benson NC, Butt OH, Brainard DH, Aguirre GK. 2014 Correction of distortion in flattened representations of the cortical surface allows prediction of V1-V3 functional organization from anatomy. *PLoS Comput. Biol.* **10**, e1003538. (doi:10.1371/journal.pcbi.1003538)
15. Witthoft N, Nguyen M, Golarai G, LaRocque KF, Liberman A, Smith ME, Grill-Spector K. 2014 Where is human V4? Predicting the location of hV4 and VO1 from cortical folding. *Cereb. Cortex* **24**, 2401–2408. (doi:10.1093/cercor/bht092)
16. Weiner KS, Golarai G, Caspers J, Chuapoco MR, Mohlberg H, Zilles K, Amunts K, Grill-Spector K. 2014 The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *Neuroimage* **84**, 453–465. (doi:10.1016/j.neuroimage.2013.08.068)
17. Weiner KS *et al.* 2017 Defining the most probable location of the parahippocampal place area using cortex-based alignment and cross-validation. *Neuroimage* **70**, 373–384. (doi:10.1016/j.neuroimage.2017.04.040)
18. Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ. 2014 Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl Acad. Sci. USA* **111**, 8619–8624. (doi:10.1073/pnas.1403112111)
19. Weiner KS, Grill-Spector K. 2010 Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage* **52**, 1559–1573. (doi:10.1016/j.neuroimage.2010.04.262)
20. Weiner KS, Grill-Spector K. 2012 The improbable simplicity of the fusiform face area. *Trends Cogn. Sci.* **16**, 251–254. (doi:10.1016/j.tics.2012.03.003)
21. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. 2009 *ImageNet: a large-scale hierarchical image database*. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, *Miami, FL*, *20–25 June*, pp. 248–255. New York, NY: IEEE.
22. Krizhevsky A, Sutskever I, Hinton GE. 2012 Imagenet classification with deep convolutional neural networks. In *Neural information processing systems* (*NIPS*) *2012* (eds F Pereira, CJC Burges, L Bottou, KQ Weinberger), *Lake Tahoe, CA*, *3–8 December*. Neural Information Processing Systems Foundation.
23. Taigman Y, Yang M, Ranzato M, Wolf L. 2014 *DeepFace: closing the gap to human-level performance in face verification*. In *The IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), *Columbus, OH*, *23–28 June*, pp. 1701–1708. New York, NY: IEEE.
24. Cadieu CF, Hong H, Yamins DL, Pinto N, Ardila D, Solomon EA, Majaj NJ, DiCarlo JJ. 2014 Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Comput. Biol.* **10**, e1003963. (doi:10.1371/journal.pcbi.1003963)
25. Khaligh-Razavi SM, Kriegeskorte N. 2014 Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput. Biol.* **10**, e1003915. (doi:10.1371/journal.pcbi.1003915)
26. Güçlü U, van Gerven MAJ. 2015 Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J. Neurosci.* **35**, 10 005–10 014. (doi:10.1523/JNEUROSCI.5023-14.2015)
27. Poggio T, Ullman S. 2013 Vision: are models of object recognition catching up with the brain? *Ann. NY Acad. Sci.* 1305, 72–82. (Cracking the Neural Code: Third Annual Aspen Brain Forum):1–11.
28. 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. (doi:10.1523/JNEUROSCI.17-11-04302.1997)

29. Tong F, Nakayama K, Moscovitch M, Weinrib O, Kanwisher N. 2000 Response properties of the human fusiform face area. *Cogn. Neuropsychol.* **17**, 257–280. (doi:10.1080/026432900380607)
30. Kanwisher N. 2017 The quest for the FFA and where it led. *J. Neurosci.* **37**, 1056–1061. (doi:10.1523/JNEUROSCI.1706-16.2016)
31. Yovel G, Wilmer JB, Duchaine B. 2014 What can individual differences reveal about face processing? *Front. Hum. Neurosci.* **8**, 562.
32. Behrmann M, Avidan G, Thomas C, Nishimura M. 2011 Impairments in face perception. In *Oxford handbook of face perception*. (eds A Calder, G Rhodes, M Johnson, JV Haxby), pp. 799–820. Oxford, UK: Oxford University Press.
33. Avidan G, Tanzer M, Hadj-Bouziane F, Liu N, Ungerleider LG, Behrmann M. 2013 Selective dissociation between core and extended regions of the face processing network in congenital prosopagnosia. *Cereb. Cortex* **24**, 1565–1578. (doi:10.1093/cercor/bht007)
34. Grill-Spector K, Weiner KS, Kay K, Gomez J. 2017 The functional neuroanatomy of human face perception. *Annu. Rev. Vis. Sci.* **3**, 167–196. (doi:10.1146/annurev-vision-102016-061214)
35. Gomez J, Pestilli F, Witthoft N, Golarai G, Liberman A, Poltoratski S, Yoon J, Grill-Spector K. 2015 Functionally defined white matter reveals segregated pathways in human ventral temporal cortex associated with category-specific processing. *Neuron* **85**, 216–227. (doi:10.1016/j.neuron.2014.12.027)
36. Kay KN, Weiner KS, Grill-Spector K. 2015 Attention reduces spatial uncertainty in human ventral temporal cortex. *Curr. Biol.* **25**, 595–600. (doi:10.1016/j.cub.2014.12.050)
37. Dricot L, Sorger B, Schiltz C, Goebel R, Rossion B. 2008 The roles of 'face' and 'non-face' areas during individual face perception: evidence by fMRI adaptation in a brain-damaged prosopagnosic patient. *Neuroimage* **40**, 318–332. (doi:10.1016/j.neuroimage.2007.11.012)
38. Jonas J, Jacques C, Liu-Shuang J, Brissart H, Colnat-Coulbois S, Maillard L, Rossion B. 2016 A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. *Proc. Natl Acad. Sci. USA* **113**, E4088–E4097. (doi:10.1073/pnas.1522033113)
39. Schiltz C, Rossion B. 2006 Faces are represented holistically in the human occipito-temporal cortex. *Neuroimage* **32**, 1385–1394. (doi:10.1016/j.neuroimage.2006.05.037)
40. Barton JJ. 2008 Prosopagnosia associated with a left occipitotemporal lesion. *Neuropsychologia* **46**, 2214–2224. (doi:10.1016/j.neuropsychologia.2008.02.014)
41. Andrews TJ, Davies-Thompson J, Kingstone A, Young AW. 2010 Internal and external features of the face are represented holistically in face-selective regions of visual cortex. *J. Neurosci.* **30**, 3544–3552. (doi:10.1523/JNEUROSCI.4863-09.2010)
42. Kietzmann TC, Gert AL, Tong F, König P. 2017 Representational dynamics of facial viewpoint encoding. *J. Cogn. Neurosci.* **29**, 637–651. (doi/10.1162/jocn_a_01070)
43. Pyles JA, Verstynen TD, Schneider W, Tarr MJ. 2013 Explicating the face perception network with white matter connectivity. *PLoS ONE* **8**, e61611. (doi:10.1371/journal.pone.0061611)
44. Gschwind M, Pourtois G, Schwartz S, Van De Ville D, Vuilleumier P. 2012 White-matter connectivity between face-responsive regions in the human brain. *Cereb. Cortex* **22**, 1564–1576. (doi:10.1093/cercor/bhr226)
45. Cukur T, Huth AG, Nishimoto S, Gallant JL. 2013 Functional subdomains within human FFA. *J. Neurosci.* **33**, 16 748–16 766. (doi:10.1523/JNEUROSCI.1259-13.2013)
46. Tsao DY, Freiwald WA, Tootell RB, Livingstone MS. 2006 A cortical region consisting entirely of face-selective cells. *Science* **311**, 670–674. (doi:10.1126/science.1119983)
47. Tsao DY, Livingstone MS. 2008 Mechanisms of face perception. *Annu. Rev. Neurosci.* **31**, 411–437. (doi:10.1146/annurev.neuro.30.051606.094238)
48. Freiwald WA, Tsao DY, Livingstone MS. 2009 A face feature space in the macaque temporal lobe. *Nat. Neurosci.* **12**, 1187–1196. (doi:10.1038/nn.2363)
49. Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB. 2003 Faces and objects in macaque cerebral cortex. *Nat. Neurosci.* **6**, 989–995. (doi:10.1038/nn1111)
50. Pinsk MA, Arcaro M, Weiner KS, Kalkus JF, Inati SJ, Gross CG, Kastner S. 2009 Neural representations of faces and body parts in macaque and human cortex: a comparative FMRI study. *J. Neurophysiol.* **101**, 2581–2600. (doi:10.1152/jn.91198.2008)
51. 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. (doi:10.1126/science.1157436)
52. Freiwald WA, Tsao DY. 2010 Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* **330**, 845–851. (doi:10.1126/science.1194908)
53. Livingstone MS, Vincent JL, Arcaro MJ, Srihasam K, Schade PF, Savage T. 2017 Development of the macaque face-patch system. *Nat. Commun.* **8**, 14897. (doi:10.1038/ncomms14897)
54. Arcaro MJ, Schade PF, Vincent JL, Ponce CR, Livingstone MS. 2017 Seeing faces is necessary for face-domain formation. *Nat. Neurosci.* **20**, 1404–1412. (doi:10.1038/nn.4635)
55. Janssens T, Zhu Q, Popivanov ID, Vanduffel W. 2014 Probabilistic and single-subject retinotopic maps reveal the topographic organization of face patches in the macaque cortex. *J. Neurosci.* **34**, 10 156–10 167. (doi:10.1523/JNEUROSCI.2914-13.2013)
56. Rajimehr R, Bilenko NY, Vanduffel W, Tootell RBH. 2014 Retinotopy versus face selectivity in macaque visual cortex. *J. Cogn. Neurosci.* **26**, 2691–2700. (doi:10.1162/jocn_a_00672)
57. Gauthier I, Skudlarski P, Gore JC, Anderson AW. 2000 Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci.* **3**, 191–197. (doi:10.1038/72140)
58. Ishai A, Ungerleider LG, Martin A, Haxby JV. 2000 The representation of objects in the human occipital and temporal cortex. *J. Cogn. Neurosci.* **12**(Suppl 2), 35–51. (doi:10.1162/089892900564055)
59. Kanwisher N, Tong F, Nakayama K. 1998 The effect of face inversion on the human fusiform face area. *Cognition* **68**, B1–B11. (doi:10.1016/S0010-0277(98)00035-3)
60. Davidenko N, Remus DA, Grill-Spector K. 2012 Face-likeness and image variability drive responses in human face-selective ventral regions. *Hum. Brain Mapp.* **33**, 2234–2249. (doi:10.1002/hbm.21367)
61. Farivar R, Blanke O, Chaudhuri A. 2009 Dorsal-ventral integration in the recognition of motion-defined unfamiliar faces. *J. Neurosci.* **29**, 5336–5342. (doi:10.1523/JNEUROSCI.4978-08.2009)
62. Felleman DJ, Van Essen DC. 1991 Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47. (doi:10.1093/cercor/1.1.1)
63. Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. 2013 The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* **17**, 26–49. (doi:10.1016/j.tics.2012.10.011)
64. Weiner KS *et al.* 2016 The face-processing network is resilient to focal resection of human visual cortex. *J. Neurosci.* **36**, 8425–8440. (doi:10.1523/JNEUROSCI.4509-15.2016)
65. Thomas C, Avidan G, Humphreys K, Jung KJ, Gao F, Behrmann M. 2009 Reduced structural connectivity in ventral visual cortex in congenital prosopagnosia. *Nat. Neurosci.* **12**, 29–31. (doi:10.1038/nn.2224)
66. Tavor I, Yablonski M, Mezer A, Rom S, Assaf Y, Yovel G. 2014 Separate parts of occipito-temporal white matter fibers are associated with recognition of faces and places. *Neuroimage* **86**, 123–130.
67. Plaut DC, Behrmann M. 2013 Response to Susilo and Duchaine: beyond neuropsychological dissociations in understanding face and word representations. *Trends Cogn. Sci.* **17**, 546. (doi:10.1016/j.tics.2013.09.010)
68. Catani M, Howard RJ, Pajevic S, Jones DK. 2002 Virtual *in vivo* interactive dissection of white matter fasciculi in the human brain. *Neuroimage* **17**, 77–94. (doi:10.1006/nimg.2002.1136)
69. Yeatman JD, Weiner KS, Pestilli F, Rokem A, Mezer A, Wandell BA. 2014 The vertical occipital fasciculus: a century of controversy resolved by in vivo measurements. *Proc. Natl Acad. Sci. USA* **111**, E5214–E5223. (doi:10.1073/pnas.1418503111)
70. Weiner KS, Yeatman JD, Wandell BA. 2016 The posterior arcuate fasciculus and the vertical occipital fasciculus. *Cortex* **20**, S0010–S9452.
71. Takemura H, Rokem A, Winawer J, Yeatman JD, Wandell BA, Pestilli F. 2016 A major human white matter pathway between dorsal and ventral visual

cortex. *Cereb. Cortex* **26**, 2205–2214. (doi:10.1093/cercor/bhv064)

72. Kay KN, Yeatman JD. 2016 Bottom-up and top-down computations in high-level visual cortex. May 16th 2. BioRxiv.

73. Tootell RB, Dale AM, Sereno MI, Malach R. 1996 New images from human visual cortex. *Trends Neurosci.* **19**, 481–489. (doi:10.1016/S0166-2236(96)10053-9)

74. Dougherty RF, Koch VM, Brewer AA, Fischer B, Modersitzki J, Wandell BA. 2003 Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *J. Vis.* **3**, 586–598.

75. Boussaoud D, Desimone R, Ungerleider LG. 1991 Visual topography of area TEO in the macaque. *J. Comp. Neurol.* **306**, 554–575. (doi:10.1002/cne.903060403)

76. Nakamura H, Gattass R, Desimone R, Ungerleider LG. 1993 The modular organization of projections from areas V1 and V2 to areas V4 and TEO in macaques. *J. Neurosci.* **13**, 3681–3691. (doi:10.1523/JNEUROSCI.13-09-03681.1993)

77. Bell AH, Malecek NJ, Morin EL, Hadj-Bouziane F, Tootell RB, Ungerleider LG. 2011 Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. *J. Neurosci.* **31**, 12 229–12 240. (doi:10.1523/JNEUROSCI.5865-10.2011)

78. Van Essen DC, Glasser MF, Dierker DL, Harwell J, Coalson T. 2012 Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cereb. Cortex* **22**, 2241–2262. (doi:10.1093/cercor/bhr291)

79. Tootell RB, Tsao D, Vanduffel W. 2003 Neuroimaging weighs in: humans meet macaques in ‘primate’ visual cortex. *J. Neurosci.* **23**, 3981–3989. (doi:10.1523/JNEUROSCI.23-10-03981.2003)

80. Orban GA, Van Essen D, Vanduffel W. 2004 Comparative mapping of higher visual areas in monkeys and humans. *Trends Cogn. Sci.* **8**, 315–324. (doi:10.1016/j.tics.2004.05.009)

81. Caspers J, Zilles K, Eickhoff SB, Schleicher A, Mohlberg H, Amunts K. 2013 Cytoarchitectonical analysis and probabilistic mapping of two extrastriate areas of the human posterior fusiform gyrus. *Brain Struct. Funct.* **218**, 511–526. (doi:10.1007/s00429-012-0411-8)

82. Lorenz S *et al.* 2015 Two new cytoarchitectonic areas on the human mid-fusiform gyrus. *Cereb. Cortex* **27**, 373–385. (doi:10.1093/cercor/bhv225)

83. Rosenke M, Weiner KS, Barnett MA, Zilles K, Amunts K, Goebel R, Grill-Spector K. 2017 A cross-validated cytoarchitectonic atlas of the human ventral visual stream. *Neuroimage*.

84. Weiner KS, Barnett MA, Lorenz S, Caspers J, Stigliani A, Amunts K, Zilles K, Fischl B, Grill-Spector K. 2017 The Cytoarchitecture of domain-specific regions in human high-level visual cortex. *Cereb. Cortex* **27**, 146–161. (doi:10.1093/cercor/bhw361)

85. Epstein R, Kanwisher N. 1998 A cortical representation of the local visual environment. *Nature* **392**, 598–601. (doi:10.1038/33402)

86. Aguirre GK, Zarahn E, D’Esposito M. 1998 An area within human ventral cortex sensitive to ‘building’ stimuli: evidence and implications. *Neuron* **21**, 373–383. (doi:10.1016/S0896-6273(00)80546-2)

87. Schwarzlose RF, Baker CI, Kanwisher N. 2005 Separate face and body selectivity on the fusiform gyrus. *J. Neurosci.* **25**, 11 055–11 059. (doi:10.1523/JNEUROSCI.2621-05.2005)

88. Peelen MV, Downing PE. 2005 Selectivity for the human body in the fusiform gyrus. *J. Neurophysiol.* **93**, 603–608. (doi:10.1152/jn.00513.2004)

89. Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R. 1999 Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* **24**, 187–203. (doi:10.1016/S0896-6273(00)80832-6)

90. Cohen L, Dehaene S, Naccache L, Lehericy S, Dehaene-Lambertz G, Henaff MA, Michel F. 2000 The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* **123**, 291–307. (doi:10.1093/brain/123.2.291)

91. Brewer AA, Liu J, Wade AR, Wandell BA. 2005 Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nat. Neurosci.* **8**, 1102–1109. (doi:10.1038/nn1507)

92. Arcaro MJ, McMains SA, Singer BD, Kastner S. 2009 Retinotopic organization of human ventral visual cortex. *J. Neurosci.* **29**, 10 638–10 652. (doi:10.1523/JNEUROSCI.2807-09.2009)

93. Grill-Spector K, Kanwisher N. 2005 Visual recognition: as soon as you know it is there, you know what it is. *Psychol. Sci.* **16**, 152–160. (doi:10.1111/j.0956-7976.2005.00796.x)

94. Thorpe S, Fize D, Marlot C. 1996 Speed of processing in the human visual system. *Nature* **381**, 520–522. (doi:10.1038/381520a0)

95. Jacques C, Witthoft N, Weiner KS, Foster BL, Rangarajan V, Hermes D, Miller KJ, Parvizi J, Grill-Spector K. 2016 Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. *Neuropsychologia* **83**, 14–28. (doi:10.1016/j.neuropsychologia.2015.07.024)

96. Jacques C, Rossion B. 2009 The initial representation of individual faces in the right occipito-temporal cortex is holistic: electrophysiological evidence from the composite face illusion. *J. Vis.* **9**, 8.

97. Liu H, Agam Y, Madsen JR, Kreiman G. 2009 Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* **62**, 281–290. (doi:10.1016/j.neuron.2009.02.025)

98. McCarthy G, Puce A, Belger A, Allison T. 1999 Electrophysiological studies of human face perception. II: response properties of face-specific potentials generated in occipitotemporal cortex. *Cereb. Cortex* **9**, 431–444. (doi:10.1093/cercor/9.5.431)

99. Hornik K, Stinchcombe M, White H. 1989 Multilayer feedforward networks are universal approximators. *Neural Netw.* **2**, 359–366. (doi:10.1016/0893-6080(89)90020-8)

100. Heeger DJ. 2017 Theory of cortical function. *Proc. Natl Acad. Sci. USA* **114**, 1773–1782. (doi:10.1073/pnas.1619788114)

101. Zhou J, Benson NC, Kay K, Winawer J. 2017 Compressive temporal summation in human visual cortex abbreviated title: compressive temporal summation.

102. Kay KN, Winawer J, Mezer A, Wandell BA. 2013 Compressive spatial summation in human visual cortex. *J. Neurophysiol.* **110**, 481–494. (doi:10.1152/jn.00105.2013)

103. Wandell BA, Winawer J. 2015 Computational neuroimaging and population receptive fields. *Trends Cogn. Sci.* **19**, 349–357. (doi:10.1016/j.tics.2015.03.009)

104. Dumoulin SO, Wandell BA. 2008 Population receptive field estimates in human visual cortex. *Neuroimage* **39**, 647–660. (doi:10.1016/j.neuroimage.2007.09.034)

105. Gomez J, Natu VS, Jeska B, Barnett MA, Grill-Spector K. 2018 Development differentially sculpts receptive fields across human visual cortex. *Nat. Commun.* **9**, 788. (doi:10.1038/s41467-018-03166-3)

106. Levy I, Hasson U, Avidan G, Hendler T, Malach R. 2001 Center-periphery organization of human object areas. *Nat. Neurosci.* **4**, 533–539. (doi:10.1038/87490)

107. Malach R, Levy I, Hasson U. 2002 The topography of high-order human object areas. *Trends Cogn. Sci.* **6**, 176–184. (doi:10.1016/S1364-6613(02)01870-3)

108. Witthoft N, Poltoratski S, Nguyen M, Golarai G, Liberman A, LaRocque KF, Smith ME, Grill-Spector K. 2016 Developmental prosopagnosia is associated with reduced spatial integration in the ventral visual cortex. *bioRxiv*.

109. Van Belle G, De Graef P, Verfaillie K, Busigny T, Rossion B. 2010 Whole not hole: expert face recognition requires holistic perception. *Neuropsychologia* **48**, 2620–2629. (doi:10.1016/j.neuropsychologia.2010.04.034)

110. Van Belle G, Busigny T, Lefevre P, Joubert S, Felician O, Gentile F, Rossion B. 2011 Impairment of holistic face perception following right occipito-temporal damage in prosopagnosia: converging evidence from gaze-contingency. *Neuropsychologia* **49**, 3145–3150. (doi:10.1016/j.neuropsychologia.2011.07.010)

111. Busigny T, Joubert S, Felician O, Ceccaldi M, Rossion B. 2010 Holistic perception of the individual face is specific and necessary: evidence from an extensive case study of acquired prosopagnosia. *Neuropsychologia* **48**, 4057–4092. (doi:10.1016/j.neuropsychologia.2010.09.017)

112. de Xivry JJ O, Ramon M, Lefevre P, Rossion B. 2008 Reduced fixation on the upper area of personally familiar faces following acquired prosopagnosia. *J. Neuropsychol.* **2**, 245–268.

113. Pelphrey KA, Sasson NJ, Reznick JS, Paul G, Goldman BD, Piven J. 2002 Visual scanning of faces in autism. *J. Autism Dev. Disord.* **32**, 249–261. (doi:10.1023/A:1016374617369)

114. Mehoudar E, Arizpe J, Baker CI, Yovel G. 2014 Faces in the eye of the beholder: unique and stable eye scanning patterns of individual observers. *J. Vis.* **14**, 6. (doi:10.1167/14.7.6)

115. Caldara R, Zhou X, Miellet S. 2010 Putting culture under the *spotlight* reveals universal information use for face recognition. *PLoS ONE* **5**, e9708. (doi:10.1371/journal.pone.0009708)

116. Caldara R, Schyns P, Mayer E, Smith ML, Gosselin F, Rossion B. 2005 Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage. *J. Cogn. Neurosci.* **17**, 1652–1666. (doi:10.1162/089892905774597254)

117. Schyns PG, Bonnar L, Gosselin F. 2002 Show me the features! Understanding recognition from the use of visual information. *Psychol. Sci.* **13**, 402–409. (doi:10.1111/1467-9280.00472)

118. Loftus GR, Harley EM. 2005 Why is it easier to identify someone close than far away? *Psychon. Bull. Rev.* **12**, 43–65. (doi:10.3758/BF03196348)

119. Sprague TC, Serences JT. 2013 Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nat. Neurosci.* **16**, 1879–1887. (doi:10.1038/nn.3574)

120. Klein BP, Harvey BM, Dumoulin SO. 2014 Attraction of position preference by spatial attention throughout human visual cortex. *Neuron* **84**, 227–237. (doi:10.1016/j.neuron.2014.08.047)

121. Ackman JB, Crair MC. 2014 ScienceDirect Role of emergent neural activity in visual map development. *Curr. Opin. Neurobiol.* **24**, 166–175. (doi:10.1016/j.conb.2013.11.011)

122. Huberman AD, Feller MB, Chapman B. 2008 Mechanisms underlying development of visual maps and receptive fields. *Annu. Rev. Neurosci.* **31**, 479–509. (doi:10.1146/annurev.neuro.31.060407.125533)

123. Shatz CJ, Stryker MP. 1978 Ocular dominance in layer IV of the cat's visual cortex and the effects of monocular deprivation. *J. Physiol.* **281**, 267–283. (doi:10.1113/jphysiol.1978.sp012421)

124. Levay S, Stryker MP, Shatz CJ. 1978 Ocular dominance columns and their development in layer IV of the cat's visual cortex: a quantitative study. *J. Comp. Neurol.* **179**, 223–244. (doi:10.1002/cne.901790113)

125. Rumelhart DE, Hinton GE, Williams RJ. 1986 Learning representations by back-propagating errors. *Nature* **323**, 533–536. (doi:10.1038/323533a0)

126. Conner IP, Sharma S, Lemieux SK, Mendola JD. 2004 Retinotopic organization in children measured with fMRI. *J. Vis.* **4**, 509–523.

127. Dekker TM, Schwarzkopf DS, de Haas B, Nardini M, Sereno MI. 2017 Population receptive field tuning properties of visual cortex during childhood. *bioRxiv* 213108. See https://www.biorxiv.org/content/early/2017/11/02/213108.

128. Natu VS, Barnett MA, Hartley J, Gomez J, Stigliani A, Grill-Spector K. 2016 Development of neural sensitivity to face identity correlates with perceptual discriminability. *J. Neurosci.* **36**, 10 893–10 907. (doi:10.1523/JNEUROSCI.1886-16.2016)

129. Grill-Spector K, Henson R, Martin A. 2006 Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* **10**, 14–23. (doi:10.1016/j.tics.2005.11.006)

130. Jiang X, Rosen E, Zeffiro T, Vanmeter J, Blanz V, Riesenhuber M. 2006 Evaluation of a shape-based model of human face discrimination using FMRI and behavioral techniques. *Neuron* **50**, 159–172. (doi:10.1016/j.neuron.2006.03.012)

131. Gilaie-Dotan S, Malach R. 2007 Sub-exemplar shape tuning in human face-related areas. *Cereb. Cortex* **17**, 325–338. (doi:10.1093/cercor/bhj150)

132. Gilaie-Dotan S, Gelbard-Sagiv H, Malach R. 2010 Perceptual shape sensitivity to upright and inverted faces is reflected in neuronal adaptation. *Neuroimage* **50**, 383–395. (doi:10.1016/j.neuroimage.2009.12.077)

133. De Valois RL, Cottaris NP, Mahon LE, Elfar SD, Wilson JA. 2000 Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. *Vision Res.* **40**, 3685–3702. (doi:10.1016/S0042-6989(00)00210-8)

134. De Valois KK, Tootell RB. 1983 Spatial-frequency-specific inhibition in cat striate cortex cells. *J. Physiol.* **336**, 359–376. (doi:10.1113/jphysiol.1983.sp014586)

135. Mazer JA, Vinje WE, McDermott J, Schiller PH, Gallant JL. 2002 Spatial frequency and orientation tuning dynamics in area V1. *Proc. Natl Acad. Sci. USA* **99**, 1645–1650. (doi:10.1073/pnas.022638499)

136. Conway BR, Livingstone MS. 2006 Spatial and temporal properties of cone signals in alert macaque primary visual cortex. *J. Neurosci.* **26**, 10 826–10 846. (doi:10.1523/JNEUROSCI.2091-06.2006)

137. Conway BR, Livingstone MS. 2003 Space-time maps and two-bar interactions of different classes of direction-selective cells in macaque V-1. *J. Neurophysiol.* **89**, 2726–2742. (doi:10.1152/jn.00550.2002)

138. De Valois RL, Cottaris NP. 1998 Inputs to directionally selective simple cells in macaque striate cortex. *Proc. Natl Acad. Sci. USA* **95**, 14 488–14 493. (doi:10.1073/pnas.95.24.14488)

139. Horiguchi H, Nakadomari S, Misaki M, Wandell BA. 2009 Two temporal channels in human V1 identified using fMRI. *Neuroimage* **47**, 273–280. (doi:10.1016/j.neuroimage.2009.03.078)

140. Watson AB. 1986 Temporal sensitivity. In *Handbook of perception and human performance* (eds K Boff, L Kaufman, J Thomas), pp. 6.1–6.43. New York, NY: Wiley.

141. Stigliani A, Jeska B, Grill-Spector K. 2017 Encoding model of temporal processing in human visual cortex. *Proc. Natl Acad. Sci. USA* **114**, E11047–E11056. (doi:10.1073/pnas.1704877114)

142. Biederman I. 1995 Visual object recognition. In *Visual cognition* (eds SM Kosslyn, DN Osherson), pp. 121–166. Cambridge, UK: MIT Press.

143. Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N. 2011 Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage* **56**, 2356–2363. (doi:10.1016/j.neuroimage.2011.03.067)

144. Wallis G, Bülthoff H. 1999 Learning to recognize objects. *Trends Cogn. Sci.* **3**, 22–31. (doi:10.1016/S1364-6613(98)01261-3)

145. Wallis G, Bülthoff HH. 2001 Effects of temporal association on recognition memory. *Proc. Natl. Acad. Sci. USA* **98**, 4800–4804. (doi:10.1073/pnas.071028598)

146. Tian M, Grill-Spector K. 2015 Spatiotemporal information during unsupervised learning enhances viewpoint invariant object recognition. *J. Vis.* **15**, 7. (doi:10.1167/15.6.7)

147. Tian M, Yamins D, Grill-Spector K. 2016 Learning the 3-D structure of objects from 2-D views depends on shape, not format. *J. Vis.* **16**, 7. (doi:10.1167/16.7.7)

148. Ullman S, Harari D, Dorfman N. 2012 From simple innate biases to complex visual concepts. *Proc. Natl Acad. Sci. USA* **109**, 18 215–18 220. (doi:10.1073/pnas.1207690109)