ANNUAL REVIEWS

Annual Review of Vision Science

Population Models, Not Analyses, of Human Neuroscience Measurements

Justin L. Gardner[1] and Elisha P. Merriam[2]

[1]Department of Psychology, Stanford University, Stanford, California 94305, USA; email: jlg@stanford.edu

[2]Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, USA; email: elisha.merriam@nih.gov

Annu. Rev. Vis. Sci. 2021. 7:225–55

First published as a Review in Advance on July 20, 2021

The *Annual Review of Vision Science* is online at vision.annualreviews.org

https://doi.org/10.1146/annurev-vision-093019-111124

Keywords

functional magnetic resonance imaging, visual cortex, orientation selectivity, computational models, decoding, linear systems analysis

Abstract

Selectivity for many basic properties of visual stimuli, such as orientation, is thought to be organized at the scale of cortical columns, making it difficult or impossible to measure directly with noninvasive human neuroscience measurement. However, computational analyses of neuroimaging data have shown that selectivity for orientation can be recovered by considering the pattern of response across a region of cortex. This suggests that computational analyses can reveal representation encoded at a finer spatial scale than is implied by the spatial resolution limits of measurement techniques. This potentially opens up the possibility to study a much wider range of neural phenomena that are otherwise inaccessible through noninvasive measurement. However, as we review in this article, a large body of evidence suggests an alternative hypothesis to this superresolution account: that orientation information is available at the spatial scale of cortical maps and thus easily measurable at the spatial resolution of standard techniques. In fact, a population model shows that this orientation information need not even come from single-unit selectivity for orientation tuning, but instead can result from population selectivity for spatial frequency. Thus, a categorical error of interpretation can result whereby orientation selectivity can be confused with spatial frequency selectivity. This is similarly problematic for the interpretation of results from numerous studies of more complex representations and cognitive functions that have built upon the computational

techniques used to reveal stimulus orientation. We suggest in this review that these interpretational ambiguities can be avoided by treating computational analyses as models of the neural processes that give rise to measurement. Building upon the modeling tradition in vision science using considerations of whether population models meet a set of core criteria is important for creating the foundation for a cumulative and replicable approach to making valid inferences from human neuroscience measurements.

1. INTRODUCTION

Sophisticated computational analyses of human neuroscience measurements have seemingly provided a means of peering past spatial resolution limits to provide a superresolution view of cortical function. Noninvasive measurement of cortical function using blood oxygen level–dependent (BOLD) imaging (Ogawa et al. 1990) is a mainstay of human neuroscience. For vision scientists, it is inarguably the best tool that we have to measure fundamental properties of human cortical function such as retinotopy (Engel et al. 1994) and visual field (Wandell & Winawer 2011) and categorical (Cohen et al. 2000, Downing et al. 2001, Epstein & Kanwisher 1998, Grill-Spector & Weiner 2014, Kanwisher et al. 1997) representation on the scale of millimeters to centimeters. Nonetheless, extending the spatial resolution below the scale of millimeters has been an enduring goal (Dumoulin et al. 2018, Lawrence et al. 2019, Martino et al. 2018, Ugurbil 2016), as this would allow for measurement of cortical columns (de No & Fulton 1938, Mountcastle 1957), long thought to be fundamental units of cortical computation (but see Horton & Adams 2005). While some success has been achieved in this endeavor (Cheng et al. 2001; Sun et al. 2007; Yacoub et al. 2007, 2008), routine and replicable measurement of cortical columns in humans has proven elusive. Instead, a remarkable development in using linear classification analyses to decode orientation of a visual stimulus from the human visual cortex (Kamitani & Tong 2005) suggested that computational analyses could overcome spatial resolution limits without the hard work required to directly measure cortical columns.

However, nothing magical need be attributed to such computational analyses; instead, extensions of traditional single-unit models of visual receptive fields to population responses can account for orientation decoding at a spatial scale matched to BOLD measurement. Models of orientation tuning are foundational to cortical visual neurophysiology; Hubel & Wiesel (1962) initially described orientation tuning in computational terms. They proposed that simple cells have receptive fields composed of excitatory and inhibitory regions that exhibit properties of summation and antagonism; thereby, the responses of cells to any stimulus can be predicted by a set of simple computational rules. This formulation of a linear receptive field triggered an entire field of study with the aim of formalizing, refining, and extending computational models of cortical visual function (Heeger et al. 1996). Such models of single-unit receptive fields are now so ingrained in the field that it is tempting to apply them directly to population measures of human brain activity. However, models of population activity can behave in unexpectedly different ways than those of single units (Hara et al. 2014, Mante & Carandini 2005). In this article, we review work that has argued against a superresolution view of orientation decoding. We describe evidence indicating that, when an appropriate model of population, rather than single-unit, receptive fields is applied, orientation decoding at the spatial scale of cortical maps is explained. Notably, these population models need not even be composed of any single units with orientation selectivity to exhibit cortical signals that can be used to decode orientation.

Orientation decoding has been used extensively throughout human neuroscience, so proper interpretation has consequences for a broad array of results. Decoding of orientation and other

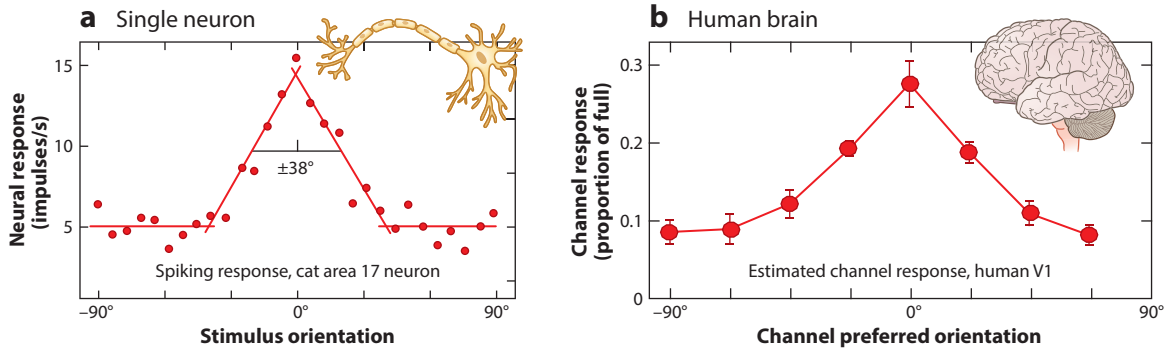


Figure 1

Tuning curves for neurons are not the same as channel response functions. (*a*) Measurement of the response of a simple cell from the cat striate cortex to grating stimuli of different orientations, plotting spike rate as a function of stimulus orientation. Panel adapted with permission from Campbell et al. (1968). (*b*) Channel response function of an inverted encoding model recovered from blood oxygen level–dependent (BOLD) imaging data from the human primary visual cortex. While the image in panel *b* may appear visually similar to that in panel *a*, it is not a tuning function because it does not plot response as a function of variations of the stimulus. Instead, it plots inferred model responses as a function of their stimulus preference, where the recovered shape of the function is an arbitrary assumption of the model (Gardner & Liu 2019). Panel adapted with permission from Liu et al. (2018).

basic visual properties has spurred countless applications to the study of a wide array of cognitive function, from perception to attention, memory, decision making, and beyond (Tong & Pratte 2012). Moreover, computational techniques routinely introduce many analysis steps in between measurement and data presentation, which can cause interpretational problems. For example, analyses of human visual cortex activity can produce a result (**Figure 1*b***) that visually resembles a classic tuning function (**Figure 1*a***). One might be forgiven for ascribing properties of the function, such as its width, to properties of the population response, such as selectivity, or for ascribing changes in the function to particular changes in memory representation or effects of attention. However, as we review in this article, the function in **Figure 1*b*** is not a tuning function, and such interpretations are not warranted (Gardner & Liu 2019). Indeed, population models suggest that these functions can result from populations of neurons that are not even tuned for orientation.

As machine learning propagates as a powerful tool to make sense of increasingly large and complex neuroscience data, its application to orientation decoding suggests broad implications for the development of models, rather than blind application of computer and data science analyses to understanding neural measurement. Data are not simply numbers in the rows and columns of a matrix as seen through the lens of a data science analysis. They are generated by a set of processes, and proper inferences can be made by treating the computational steps applied to data as models of one's best hypotheses about those processes. Thus, we describe in this review a consensus model for orientation decoding that can be used to test general hypotheses about the source of the orientation signal, as well as being applied to the study of cognitive function. The way out of the interpretational problems that have arisen around orientation decoding is through building and sharing such models. More generally, the approach that we advocate for in this review can be widely applied across neuroscience measurements that capture population responses in humans and animals. Building and testing such population models form the foundation of a replicable and cumulative science and provide means of drawing inferences from human neuroscience measurements.

2. DECODING AND ORIENTATION COLUMNS

Is the mismatch between the scale of BOLD measurements and the smaller spatial scale of cortical columnar computation an insurmountable limit? Diverse patterns of neural selectivity within a voxel might be expected to partially or fully cancel each other out, resulting in BOLD responses that exhibit diminished feature selectivity. The problem of heterogeneous neural populations within a voxel is demonstrated by the case of orientation selectivity in the primary visual cortex. The orientations of visual features are represented in an orderly pinwheel-like progression within each hypercolumn across the cortical surface (Das & Gilbert 1997, Grinvald et al. 1986, Ohki et al. 2006). Based on measurements of ocular dominance columns in humans (Adams et al. 2007, Horton & Hedley-Whyte 1984), orientation columns are likely to be at a submillimeter scale when viewed from the surface of the cortex. This well-established neural architecture is considerably smaller than the spatial point spread function for BOLD measurements at conventional resolution (Engel et al. 1997, Parkes et al. 2005). At first glance, it seems totally infeasible to study orientation selectivity with BOLD imaging. Indeed, the spatial scale of BOLD measurements seems inappropriate for studying a wide range of neural computations that are instantiated by cortical circuits smaller than the point spread function.

One approach to accessing fine-spatial-scale patterns of neural selectivity is a straightforward attempt to match the scale of the measurement to the scale of the underlying neural architecture. Advances along these lines using ultrahigh-field-strength scanners and pulse sequences that are sensitive to signals originating in small capillaries, which are more spatially colocalized with changes in neural activity (Kay et al. 2019, Markuerkiaga et al. 2016), have led to impressive demonstrations characterizing responses in cortical columns (Cheng et al. 2001; Kim & Fukuda 2008; Sun et al. 2007; Yacoub et al. 2007, 2008) and layers (Finn et al. 2019, Huber et al. 2017, Yu et al. 2019). However, achieving the necessary resolution is technically challenging, and potential sources of confounds and artifacts, such as head motion and spatial distortions, remain a hindrance. Moreover, the long scanning sessions needed to achieve the necessary signal-to-noise ratio make achieving columnar resolution difficult for experiments that require routine and replicable measurements across experimental conditions.

An alternative approach is to pair conventional, low-resolution BOLD measurements with paradigms and data analysis strategies that attempt to peer past spatial resolution limits. An early attempt at such a strategy relied on adaptation. In an adaptation protocol, subjects are presented with an adaptor stimulus followed by a probe stimulus. If the probe stimulus differs from the adaptor along a particular stimulus dimension and fails to produce a reduced response, then this is interpreted to mean that a different nonadapted population of neurons has responded to the probe. If, however, changing the features of the probe stimulus results in the same amount of adaptation as repeating the adaptor stimulus, then this is interpreted to mean that the underlying neural population has homogeneous selectivity (Grill-Spector et al. 1999). Adaptation analysis has been used as a tool to investigate neural selectivity in a wide range of domains in many brain regions, including early visual cortical areas (Cavina-Pratesi et al. 2010, Fang et al. 2005, Gardner et al. 2005, Hallum et al. 2011, Larsson et al. 2006, Lingnau et al. 2009, Sapountzis et al. 2010), higher-order visual areas (Henson 2016, Weiner et al. 2010, Winston et al. 2004), and motor systems (Chong et al. 2008, Dinstein et al. 2007). However, using adaptation as a tool to evaluate selectivity is subject to several interpretational ambiguities. Adaptation studies typically focus on the reduced response to the probe stimulus, but neural responses to adaptor stimuli are frequently enhanced due to the increased salience of novel stimuli (Summerfield et al. 2008), so it is necessary to develop additional controls to distinguish an enhanced adaptor from an adapted probe (Larsson & Smith 2012). Observing an adapted response does not necessarily indicate which neurons

exhibit stimulus selectivity, as the adaptation could have occurred at an earlier stage of processing and then been carried forward (Larsson & Harrison 2015). The reduced response to the probe stimulus can also reflect response learning, rather than a change in the stimulus representation (Dobbins et al. 2004, Schacter et al. 2007). Thus, the adaptation protocol does not offer a simple and direct means of studying neural selectivity (Grill-Spector et al. 2006, Larsson et al. 2016).

Decoding analyses have been widely heralded as a solution for studying selectivity (Haynes & Rees 2006). Decoding refers to a broad class of statistical procedures borrowed from machine learning. Unlike the adaptation protocol, which relies on the average response within a particular brain area, decoding exploits the distributed pattern of response.

In a landmark study, Kamitani & Tong (2005) demonstrated that it is possible to use a linear classifier to decode the orientation of a grating presented to the subject in an individual trial. This was a major conceptual breakthrough and suggested that it is possible to directly study neural representations in the human brain that are instantiated at a finer spatial scale than a typical BOLD voxel measurement. The reasoning behind this is illustrated in **Figure 2**. Subjects were presented with a set of oriented grating stimuli, each of which evoked responses from orientation-tuned neurons in the primary visual cortex. Each voxel (at a conventional resolution of 2.5 × 2.5 × 2.5 mm) samples from a patch of cortical tissue containing roughly 750,000 neurons and a similar number of glia. If neural selectivity for orientation were randomly organized, forming

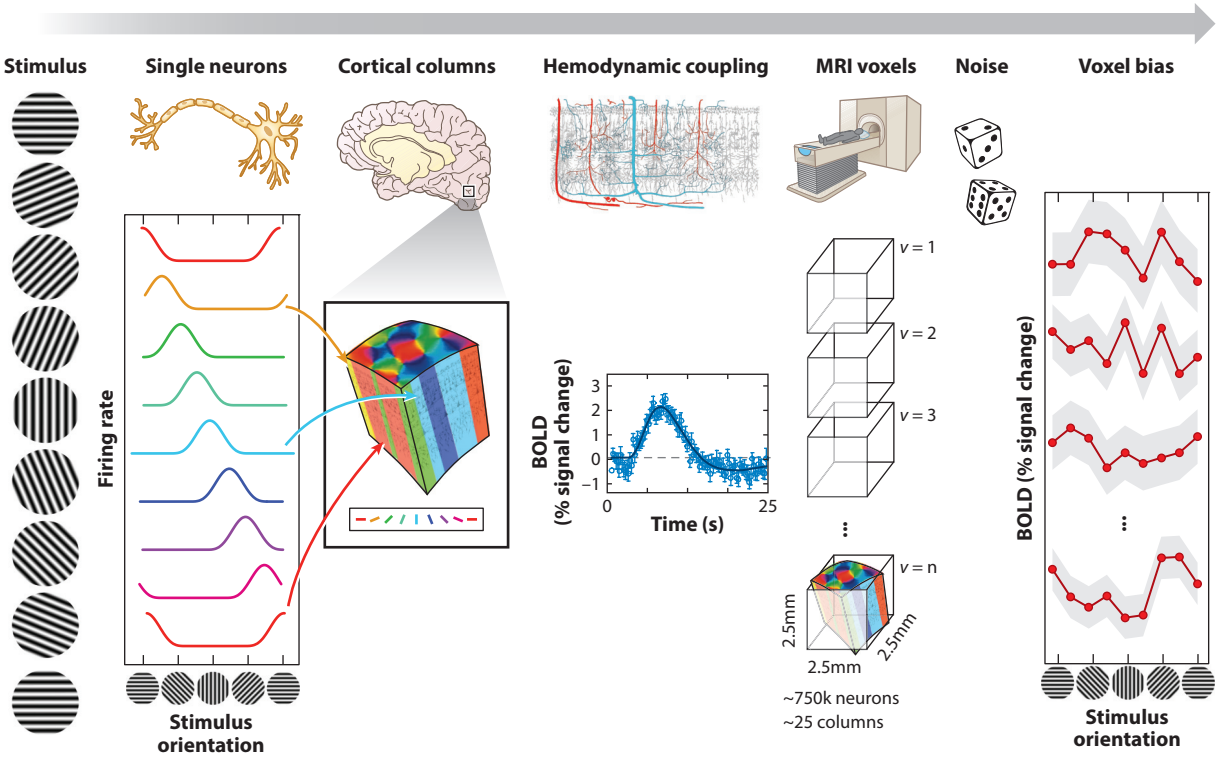


Figure 2

Superresolution account of orientation decoding. Stimuli of different orientations (*left*) evoke responses in single neurons that are tuned for orientations and organized into cortical columns. The BOLD signal in each voxel is then hypothesized to be biased for orientation (*right*) due to random sampling of columns within each voxel. Abbreviations: BOLD, blood oxygen level–dependent; MRI, magnetic resonance imaging.

a salt-and-pepper pattern in the cortex, then the response of that large population of neurons would cancel out, and the net response of the voxel would not be orientation tuned. However, the distribution of orientation-selective V1 neurons is not random. In many species (presumably including humans), orientation-tuned neurons are organized into columns along the cortical surface, and a voxel likely samples from approximately 25 columns (Gardner 2010). Again, if the pattern of orientation-selective columns formed a perfectly organized repeating motif across the cortical surface in register with the voxel grid, then pooling responses across columns might also cancel out, as all orientations would be sampled by a voxel. However, columnar patterns are irregular. They vary in shape and size across the cortical surface (Das & Gilbert 1997, Grinvald et al. 1986, Ohki et al. 2006). It is unlikely that the columnar pattern would be in alignment with the voxel grid. Any given voxel likely samples an uneven distribution of columns, with some orientations being represented more frequently than others. This uneven sampling could, in principle, produce a small bias in the voxel's response for a particular orientation.

The conjecture that orientation preferences arise from random spatial irregularities in the fine-spatial-scale columnar architecture is attractive for several reasons. An analysis method provides a window into subvoxel columnar structure—a sort of superresolution account—without the immense technical challenges associated with ultrahigh-resolution magnetic resonance imaging (MRI). This notion places the burden of technical development on the data analysis pipeline, rather than on the MRI data acquisition. It also opens up the possibility of applying these analysis methods to other brain areas outside of the visual cortex. Indeed, the logic of superresolution rests on the assumption of a columnar organization. Following this logic, does successful application of decoding for a stimulus feature or task imply the existence of columnar organization for that feature? As we review below, an abundance of evidence argues against this superresolution view.

3. DECODING AND TOPOGRAPHIC MAP-LIKE ORIENTATION BIASES

An alternative account to superresolution has emerged over the years, demonstrating that orientation decoding depends on a topographic map-like representation of orientation, which is distinct from the more familiar fine-spatial-scale columnar neural architecture (Freeman et al. 2011, Sasaki et al. 2006). Studies in human subjects have shown that each voxel in V1 exhibits an orientation preference that depends on the region of space that it represents (Freeman et al. 2011). This map-like structure is most pronounced as a radial bias in the peripheral representation of V1 that is closely matched to the angular component of the retinotopic map (compare **Figure 3*a*** and **Figure 3*b***). Voxels that respond to peripheral locations near the vertical meridian tend to respond most strongly to vertical orientations, and voxels along the peripheral horizontal meridian respond most strongly to horizontal orientations; the same is true for oblique orientations (**Figure 3*c***).

The relationship between preferred orientation and receptive field location varies with position in the visual field. In addition to the radial bias in the periphery, there is a near-vertical bias closer to the fovea (Freeman et al. 2013). This near-vertical bias is evident for voxels that respond to stimuli at either the vertical or horizontal meridians (Freeman et al. 2013); however, others have reported a mix of vertical and horizontal biases (Sun et al. 2013), perhaps corresponding to earlier studies of the oblique effect (Furmanski & Engel 2000). The map-like orientation bias has gone mostly unnoticed by the field until recently, despite decades of research on orientation selectivity with electrophysiology and optical imaging. The spatial coverage of BOLD imaging makes it ideally suited to comparing preferred orientation and receptive field position throughout the visual cortex. Indeed, early studies in the rodent visual cortex reported only a salt-and-pepper organization for orientation selectivity (Ohki et al. 2005). However, a wide-field calcium imaging

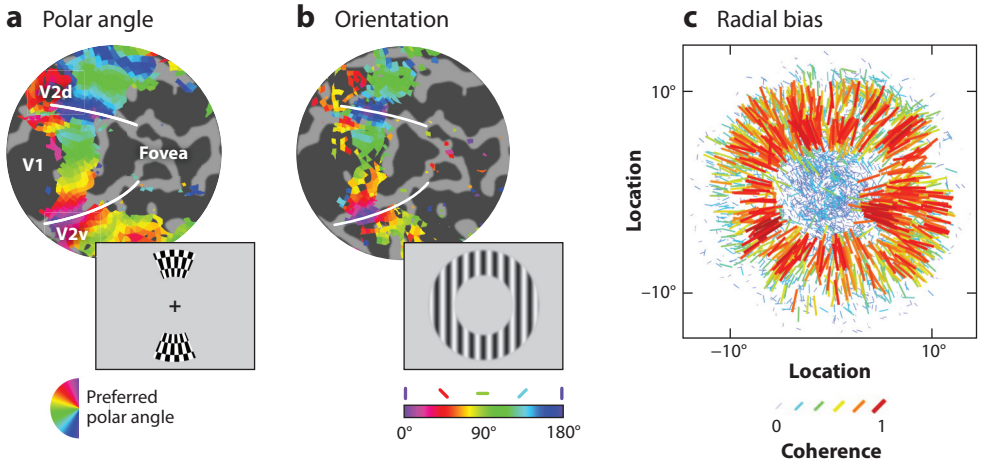


Figure 3

Map-like topography for orientation. (*a*) Retinotopic responses to a rotating wedge from a single subject, illustrating the angular component of the retinotopic map. Colors indicate preferred angular position. Gray scale represents a flattened representation of a patch of the occipital cortex. (*b*) Responses to oriented gratings restricted to an annular aperture. Colors indicate preferred orientation. (*c*) Radial bias. Each voxel is represented by a line indicating its population receptive field (pRF) location, preferred orientation, and response amplitude (color and size). Inward-pointing lines indicate a radial bias. Figure adapted with permission from Freeman et al. (2011).

study in mice, which affords wide coverage at single-cell resolution, has now reported what may be a map-like representation for orientation (Fahey et al. 2019) analogous to that found in humans. This highlights the need for the scale of the imaging modality to match the scale of the underlying neural architecture.

While the map-like pattern of orientation selectivity in humans and mice does not confirm the presence (or absence) of orientation columns in either species, these results do offer an alternative mechanism by which orientation information may be decoded. When orientation decoding was first introduced, it was natural to assume that the decoded signal was somehow related to columnar structure (Boynton 2005, Haynes & Rees 2005, Kamitani & Tong 2005, Peelen et al. 2006). An early indication that this was not the case came from a study showing that spatially smoothing (i.e., blurring) the BOLD data had no impact on decoding performance (de Beeck 2010). This observation runs contrary to the intuition that smoothing should degrade information arising from columns. However, this notion was soon challenged by others who took issue with the particular way in which spatial smoothing was implemented, noting that smoothing simply spreads information spatially across voxels, rather than actually removing information from the signal (Kamitani & Sawahata 2010). This was followed by another study that estimated the bandwidth of information used by decoding by filtering the data with low-pass and high-pass spatial filters with different cutoff frequencies (Swisher et al. 2010). The problem with all of these approaches is that any activation pattern in BOLD imaging is spatially broadband, with prominent high- and low-spatial-frequency components. The same critical bandwidth observed for orientation is also observed for retinotopy (Freeman et al. 2011), which is known to be mapped topographically across the cortex. In conclusion, these smoothing analyses do not support the superresolution account.

The map-like orientation bias has a strong radial component that is especially prominent in the periphery. Removing this radial component from the BOLD signal through projection makes orientation decoding impossible. This observation suggests that the orientation map is necessary

for decoding and argues against superresolution (Freeman et al. 2011). However, these results have been challenged by Pratte et al. (2015), who repeated the same analyses described above but with a slightly different set of assumptions regarding the shape and time course of the hemodynamic response. In their analysis, orientation decoding was mostly, but not completely, explained by the radial bias, leaving open the possibility that some columnar-scale signals remain.

If orientation is mapped in the cortex as a radial bias, then it should not be possible to decode the sense of spiral stimuli because local orientation information in spirals with respect to the radial direction are balanced across clockwise and counterclockwise spirals (Mannion et al. 2009). However, it is indeed possible to decode spiral sense (Alink et al. 2013, 2017; Clifford & Mannion 2014; Mannion & Clifford 2011; Mannion et al. 2009). Is this evidence for superresolution (Carlson & Wardle 2015, Maloney 2015)? The map-like topography for orientation is not perfectly radial in all parts of the map. As discussed above, responses are strongest for near-vertical orientations at locations closer to the fovea (Freeman et al. 2011). Such a near-vertical bias does enable the decoding of spiral sense (Freeman et al. 2013), suggesting that the ability to decode spiral stimuli does not require a superresolution account.

Perhaps the most compelling argument against superresolution comes from experiments in which the position of slice acquisition was shifted by half of a voxel (1 mm) in the slice acquisition direction on half of the runs (Freeman et al. 2013). Such a shift would maximally disrupt the relationship between the voxel grid and the underlying columnar architecture. Yet shifting slices between training and testing data sets has no discernible impact on orientation decoding accuracy. However, even this result has been challenged by others, who have found that translating (i.e., spatially interpolating) the slices in postprocessing results in a significant decrement in decoding accuracy (Alink et al. 2017, Vizioli et al. 2020).

Another coarse-spatial-scale signal for orientation could come from the vasculature if it is organized in such a way as to pool orientation-specific signals (Gardner 2010, Kriegeskorte et al. 2010). Indeed, at a large spatial scale on the order of centimeters, some evidence suggests that vasculature boundaries can follow functional areas (Harrison et al. 2002). Moreover, sinuses have been suggested to distort large-spatial-scale representations such as topographic maps (Winawer et al. 2010). If vasculature is organized around functional boundaries such as cortical columns, then it could pool orientation-dependent signals, giving rise to an orientation bias at a larger scale. However, more recent evidence suggests that vasculature is not organized around ocular dominance columns (Adams et al. 2015) or barrels in the rodent somatosensory cortex (Blinder et al. 2013).

The debate on whether orientation decoding is due to a superresolution or cortical map-like representation has produced a dizzying, and seemingly contradictory, array of results. Why has this debate attracted so much attention? If orientation decoding depends entirely on a map-like bias, and not at all on a fine-spatial-scale columnar bias, then this would imply that superresolution decoding is not possible. If superresolution decoding does not work in the visual cortex for orientation, then it is not likely to work for other brain areas and functions for which columnar structure is less well understood, limiting the use of decoding. While this back-and-forth in the field has certainly been productive, we argue that resolving these issues and building a stronger foundation for the use of decoding requires treating decoding as a model of the neural processes that give rise to measurement, and not simply as an analysis of data.

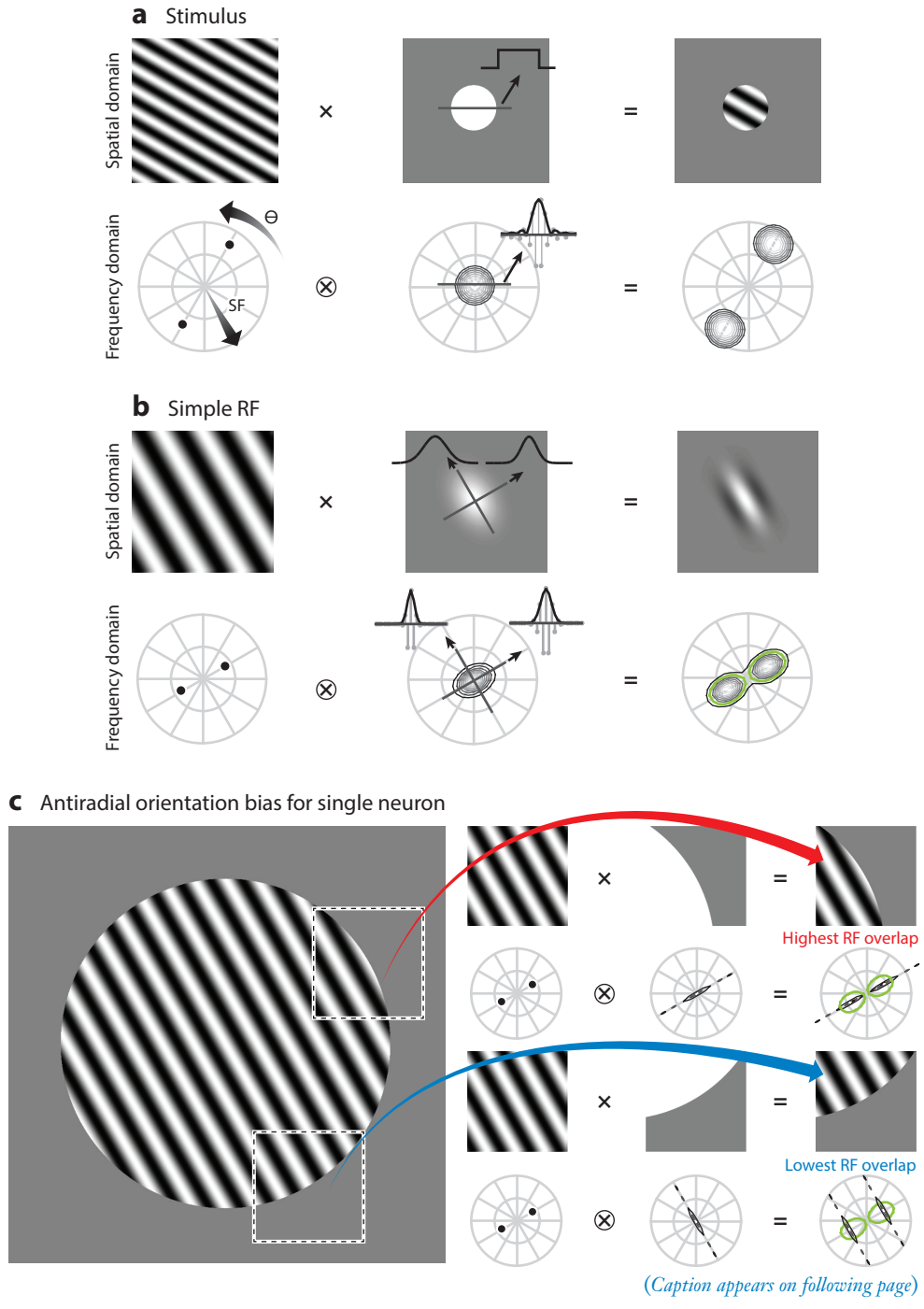
4. POPULATION MODEL OF V1 EXPLAINS MAP-LIKE TOPOGRAPHY FOR ORIENTATION

Perhaps the most fundamental and widely used model in visual neuroscience is that of a linear time-invariant system, which is simple and completely predictable but provides an unexpected

answer to where map-like orientation biases might actually come from. A linear time-invariant system is one in which the relationship between stimulus and response does not change over time and obeys the superposition principle (response to the sum of inputs is equal to the sum of the responses to each). The receptive field of a neuron or population that is linear time invariant can be characterized with incredible efficiency, since responses to combinations of visual inputs need not be directly measured, but instead can be predicted from the response to those inputs presented alone. Of course, neural (or any) systems are rarely, if ever, completely linear time invariant. Neural responses break time invariance when they are found to be subject to adaptation effects (Cavina-Pratesi et al. 2010, Fang et al. 2005, Gardner et al. 2005, Hallum et al. 2011, Larsson et al. 2006, Lingnau et al. 2009), and a long tradition within visual neuroscience has explored violations of superposition (Albrecht & Geisler 1991, Bonds 1989, Movshon et al. 1978, Reid et al. 1987) and how to fix them with simple forms of static nonlinearities such as exponentiation and through normalization (Carandini & Heeger 2012; Carandini et al. 1997; Heeger 1992, 1993). Despite being an oversimplification that is demonstrably wrong, linear time-invariant models still can provide incredibly useful insight because they make concrete predictions that are not always immediately obvious. One such counterintuitive prediction is that a set of linear receptive fields whose orientation preference is completely isotropic at any given location of the visual field will give rise to radial orientation bias when a sinusoidal-oriented stimulus is presented in a circular aperture (Carlson 2014).

For single units, map-like orientation biases result from considerations of how the stimulus border that is in the receptive field of a neuron depends on the retinotopic location of the receptive field. For a linear time-invariant receptive field (**Figure 4*b***), the response to a stimulus (**Figure 4*a***) can be appreciated either in the space domain or in the frequency domain by determining the projection of the stimulus onto the receptive field. For these oriented stimuli, the frequency domain representation is convenient, as the response (ignoring the stimulus phase) can be visually appreciated as the amount of overlap of the frequency domain representations of the stimulus and the receptive field. When the stimulus is very large (**Figure 4*c***) compared to the receptive field of a neuron, as is typical for human neuroscience experiments, depending on the retinotopic location of each neuron's receptive field, different borders of the stimulus will overlap the neuron's receptive field. For example, a neuron whose receptive field is selective for the grating orientation and is at the top-right portion of the stimulus would see a vignette nearly aligned in orientation with the underlying grating (**Figure 4*c***, *top right*). The vignette would then have a frequency domain representation with energy largely along the orientation of the grating stimulus. Thus, the frequency domain representation of the stimulus at this location (being the convolution of the grating and the vignette) would have a profile that smears stimulus energy best into the receptive field of a neuron with an elongated receptive field (compare the receptive field profile and the stimulus energy in **Figure 4*c***, *top right*). This can equivalently be thought of as giving the strongest response because both the grating and the vignette are oriented in the preferred orientation for the neuron. Thus, receptive fields selective for the orientation of the stimulus with a receptive field location at exactly the location where the vignette aligns with the orientation of the stimulus will respond the most, and all other locations will respond less. Importantly, this is a consequence of the elongated shape of the receptive field, a property known from physiological measurements (Gardner et al. 1999, Jones & Palmer 1987, Ringach 2002, Watkins & Berkley 1974).

While this single-unit model gives an explanation for map-like orientation bias, the bias predicted is exactly opposite to what has been observed in humans. Because of the receptive field's elongated shape, a single unit should respond most strongly to the antiradial orientation (i.e., the tangential orientation parallel to the vignette edge) (**Figure 4*c***, *top right*). Thus, the single-unit model prediction runs counter to the maps of radial orientation bias (Freeman et al. 2011,



(Caption appears on following page)

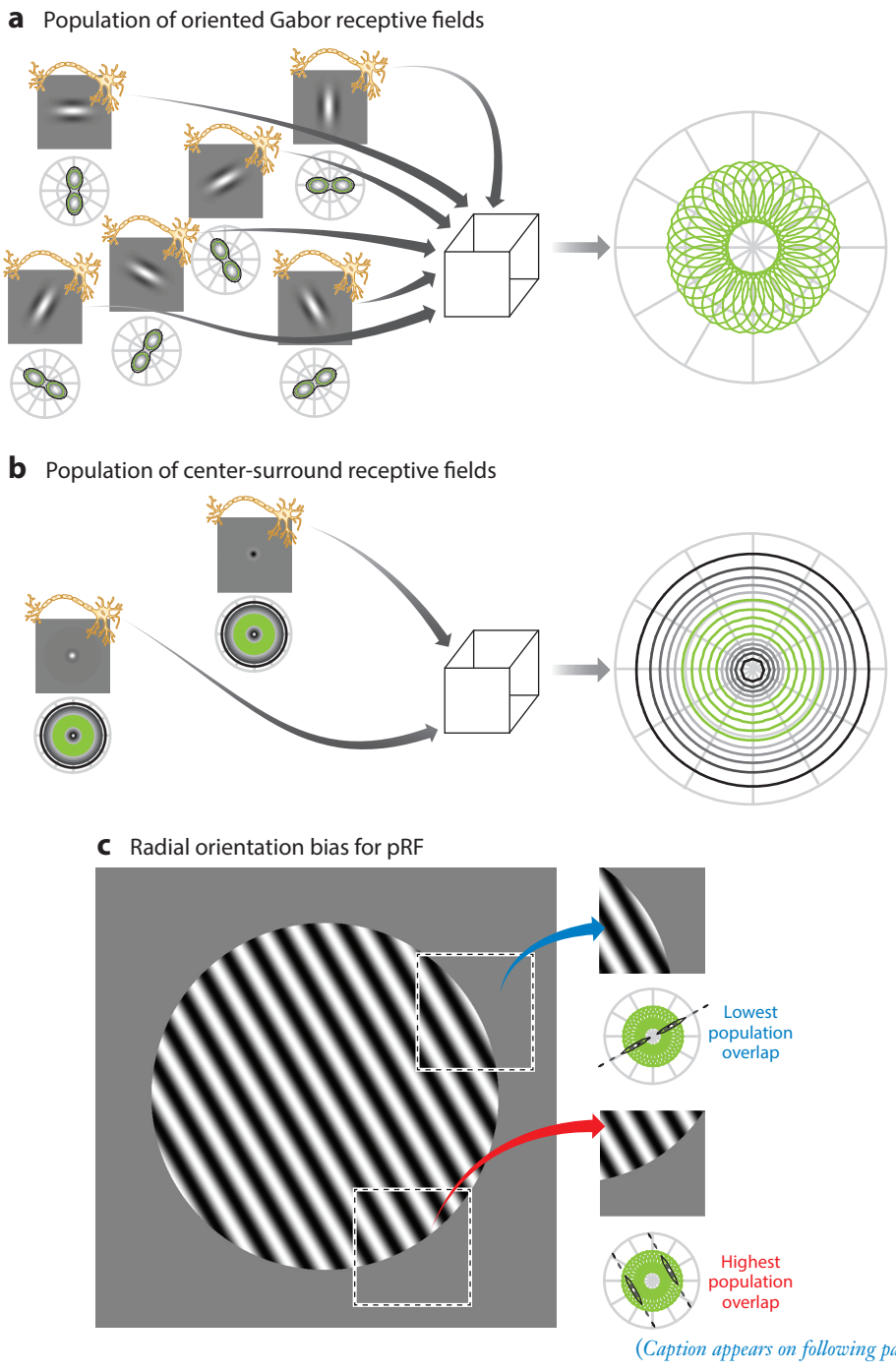
Figure 4 (*Figure appears on preceding page*)

Stimulus vignetting and responses of an individual V1 neuron. (*a*) The stimulus in a typical experiment consists of an infinite grating multiplied by an aperture (*top*). In the frequency domain, the grating is represented as a pair of delta functions at the corresponding orientation (polar angle) and spatial frequency (radius). Vignetting in the frequency domain corresponds to convolution in the frequency domain, which simply creates copies of the Fourier transform of the vignette (*bottom middle*) centered at the orientation and spatial frequency of the grating stimulus (*bottom right*). (*b*) The RF of a V1 simple cell can be represented by a Gabor filter created by multiplying an infinite grating and an elongated two-dimensional Gaussian. The long axis of the Gaussian is typically aligned with the preferred orientation. In the frequency domain, this elongation produces an RF that is elongated along the orientation axis, making it selective for orientation at multiple spatial frequencies. (*c*) Bias in orientation preference results when the RF overlaps the edge of the stimulus aperture. The RF depicted (*green contour*, *right*) has a preferred orientation that matches the stimulus. The amount of orientation energy in the stimulus (*black contours*) that overlaps the RF is highest when the stimulus vignette (*top right*, *middle*) is approximately aligned with the orientation of the stimulus causing the RF to have the strongest response. Thus, this cell responds most strongly to the vignette depicted in the upper right panel, and least strongly to the vignette in the lower right panel, creating antiradial bias (i.e., largest response to the orientation tangential to the RF location relative to the fovea). Abbreviations: RF, receptive field; SF, spatial frequency.

Sasaki et al. 2006). What is wrong with this logic? Perhaps the assumption about the aspect ratio of the receptive fields of V1 neurons, namely, that they are elongated along their orientation axis, is incorrect. Indeed, if receptive fields had the opposite aspect ratio (fatter in the direction along the bars), then this would give rise to the expected radial orientation bias. However, single-unit measurements suggest that this is not the case (Gardner et al. 1999, Jones & Palmer 1987, Ringach 2002, Watkins & Berkley 1974). Moreover, for this to be true, human V1 neurons would be expected to have broad spatial frequency tuning, as their receptive field in the frequency domain would be elongated along spatial frequency instead of orientation (opposite of what is shown in **Figure 4*b***, *bottom right*), which does not appear to be the case (Aghajari et al. 2020, Broderick et al. 2018, Keliris et al. 2019).

An explanation of the apparent contradiction between the single-unit prediction and the BOLD results comes from remembering that the BOLD measurement reflects a neural population. Thus, a single-unit model is not appropriate. Each voxel samples from a population of neurons with different orientation selectivities. An appropriate population model will thus have a population of receptive fields, which together form an annulus in the frequency domain (**Figure 5*a***). The population receptive field will respond equally to all orientations, but only to a range of spatial frequencies. Following the same logic outlined above for individual neurons (**Figure 4*c***), this annulus intersects with the spread in Fourier energy produced by the stimulus vignette (**Figure 5*c***) to produce the observed radial orientation bias. The smear of stimulus energy for the radial orientation (**Figure 5*c***, *bottom right*) has the highest overlap with the population receptive field, and the least overlap for the antiradial orientation (**Figure 5*c***, *top right*). Thus, consideration of the population, rather than the single-unit, receptive field explains the radial orientation bias. If this account is true, then the population model should be able to generate predictions of responses to novel stimuli. For example, stimuli with the same underlying orientation could be shown through different vignettes tailor-made to reverse orientation selectivity along the map. Indeed, in an experiment using a similar population receptive field model to that of Simoncelli & Freeman (1995), these predictions were borne out (Roth et al. 2018).

This population model suggests that the apparent orientation selectivity of BOLD measurements comes about not because of orientation selectivity, but because of spatial frequency selectivity. Perfectly circularly symmetric center-surround receptive fields such as those of retinal ganglion cells and the lateral geniculate nucleus (LGN) (Barlow 1953, Kuffler 1953, Rodieck 1965)



(Caption appears on following page)

Figure 5 (*Figure appears on preceding page*)

Stimulus vignetting and responses of populations of neurons. (*a*) Neurons in V1 with a range of orientation preferences (*left*) contribute to the activity of a magnetic resonance imaging (MRI) voxel in V1, creating a population receptive field (pRF) (*right*). The orientation tuning of the pRF is different from that of any individual neuron in the population (*green contours*, *right*). The Fourier representation of the pRF has a characteristic ring due to the fact that neurons contributing to the pRF have a range of orientation preferences but similar spatial frequency preferences. (*b*) pRF created by a population of circularly symmetric center-surround lateral geniculate nucleus (LGN) cells. Note that the population response creates a ring in the Fourier domain that is similar to the ring created by the population of V1 cells, even though the LGN neurons differ from V1 neurons in that they are not orientation selective. (*c*) Bias in orientation preference of the pRF created by different combinations of aperture edge and stimulus, following the same conventions as in **Figure 4**. While individual V1 neurons may have an antiradial bias, and individual LGN neurons may have no orientation preference, the pRFs of both populations exhibit a radial bias.

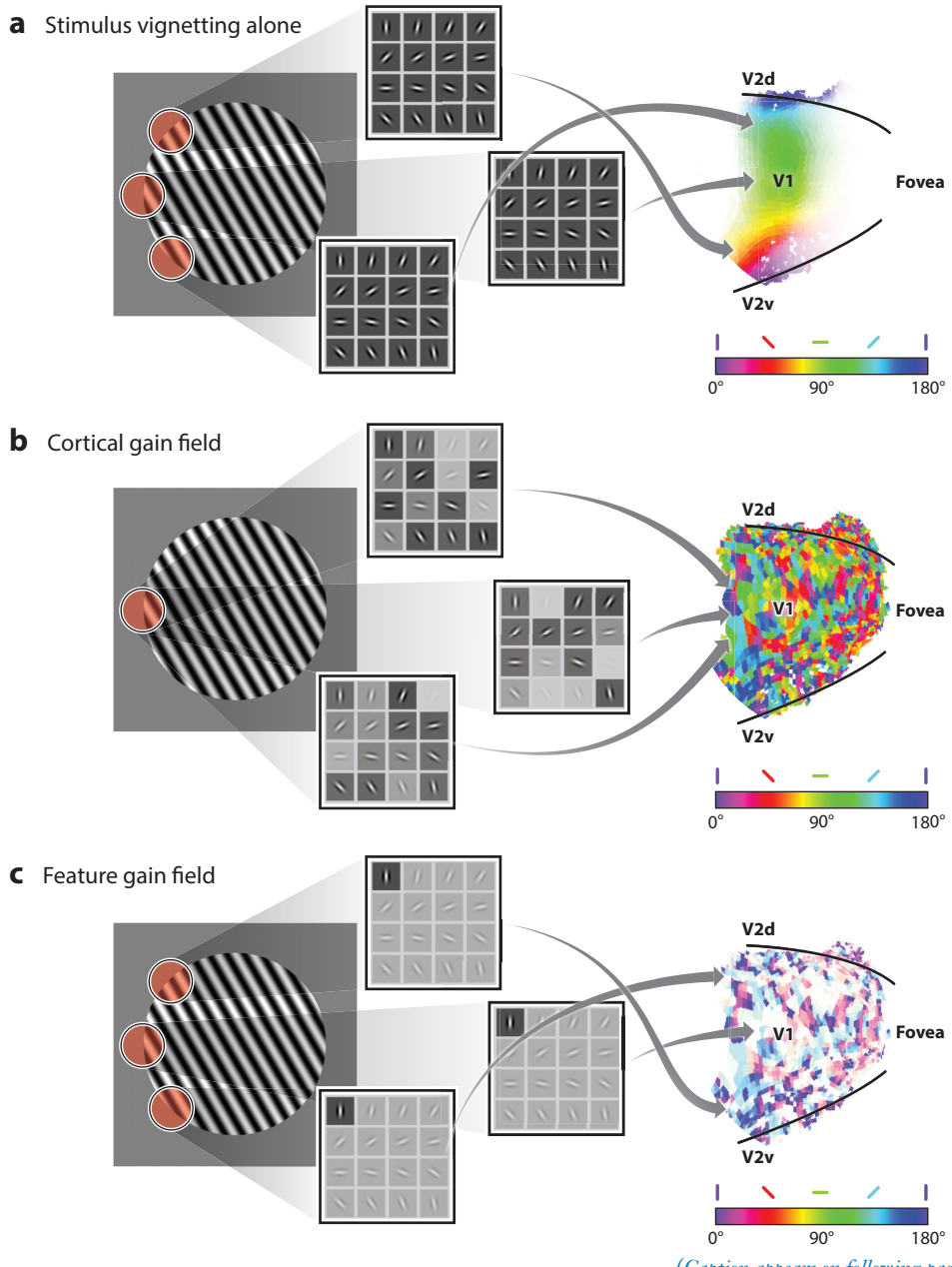
would have a population response with the same required annular shape in the frequency domain (**Figure 5*b***), which would lead to the ability to decode orientation from areas that do not exhibit orientation selectivity in single units. Indeed, the orientation of a visual stimulus can be decoded from the LGN (Ling et al. 2015). Thus, according to the population model, ascribing orientation selectivity to the ability to decode orientation from population responses is a categorical error. That is, decoding orientations is evidence of spatial frequency tuning, not orientation tuning.

5. A CONSENSUS POPULATION MODEL

While a simple population model and considerations of the effect of stimulus vignetting can account for the ability to decode orientation without invoking a superresolution account, this is not to say that other sources of orientation information are not at play. Certainly, there is the possibility that columnar organization could contribute to BOLD measurements at high spatial resolution. Biases in organization from earlier stages of the early visual system, including the retina and LGN (Levick & Thibos 1980, 1982; Ling et al. 2015; Ringach 2007; Rodieck et al. 1985; Schall et al. 1986; Shou et al. 1986; Smith et al. 1990), might contribute to cortical organization. An over-representation of cardinal orientations (Sun et al. 2013) that might match the statistical distribution of local orientation in natural images (Girshick et al. 2011) may also contribute. Attention (Ling et al. 2015), working memory (Harrison & Tong 2009), and other cognitive factors may change the gain of response or leave a trace of previously encountered stimuli. A model like the one in **Figure 5**, which does not include these factors, offers no possibility to test for these types of effects. We need to generate a consensus model that incorporates multiple sources of orientation bias, from fine spatial scale to coarse spatial scale, and permits inferences regarding the relative contribution of each.

Beyond stimulus vignetting, what are the most likely contributions to orientation selectivity in population measures of V1? One possibility is that there are differences in orientation tuning that depend on position in the cortex. The impact of cortical position on orientation selectivity can be included in the population model by applying a gain to the orientation-selective filters. Specifically, the output of the orientation-selective filters is multiplied by a cortical gain field in which the amplitude of the gain depends on the location in the cortex (**Figure 6*b***). The columnar architecture is itself an example of a cortical gain field. However, there could be many factors that result in an interaction between orientation selectivity and position in the cortex, such as differences across cortical layers or even between cell types. Another possible source of orientation selectivity, other than stimulus vignetting, is related to coarse-scale anisotropies in preferred orientation across the population, for example, a more prominent representation of vertical and/or horizontal orientations. Such effects can be included in the model we propose by incorporating a feature gain field

in which each simulated neuron's response is multiplied by an orientation-dependent modulator, which could produce, for example, larger responses to vertical and/or horizontal orientations, resulting in an oblique effect across the population (**Figure 6*c***). Of course, it is also likely the case that each of these potential sources of orientation selectivity—stimulus vignetting, cortical gain field, and feature gain field—all contribute to some degree to orientation bias and that any population measurement reflects a combined contribution of each. Our goal in this review is not to



(*Caption appears on following page*)

Figure 6 (*Figure appears on preceding page*)

Consensus model of orientation selectivity. Several distinct factors in addition to stimulus vignetting can, in principle, contribute to orientation preferences in population measurements. The consensus model provides a flexible, modular platform for testing predictions of each of these factors. (*a*) Stimulus vignetting, in the absence of any additional factors, creates a radial bias most pronounced in voxels with population receptive fields overlapping the aperture edge. Colors indicate the preferred orientation of each voxel, as in **Figure 3**, and opacity indicates the strength of the response. (*b*) Cortical location gain depends on position in the cortex, creating a bias in the population measurement. The most notable source of cortical gain arises from the columnar architecture. (*c*) Feature-specific gain, such as an enhanced response to vertical orientations, creates a map with an over-representation of vertical orientations. Responses to orientations other than vertical are shown with decreased opacity.

provide an exhaustive list of factors that could give rise to an orientation bias, but rather to provide the framework for building a flexible and modular model in which features can be added or taken away, fit to experimental data with novel stimuli across a range of measurement modalities, and refined over time to produce a more complete model of the visual cortex.

A consensus model provides a framework for testing the impact of each of the computations that link the stimulus with the measured response (**Figure 7**). A consensus model is abstracted from physiological and anatomical implementation, but those considerations can be built in as

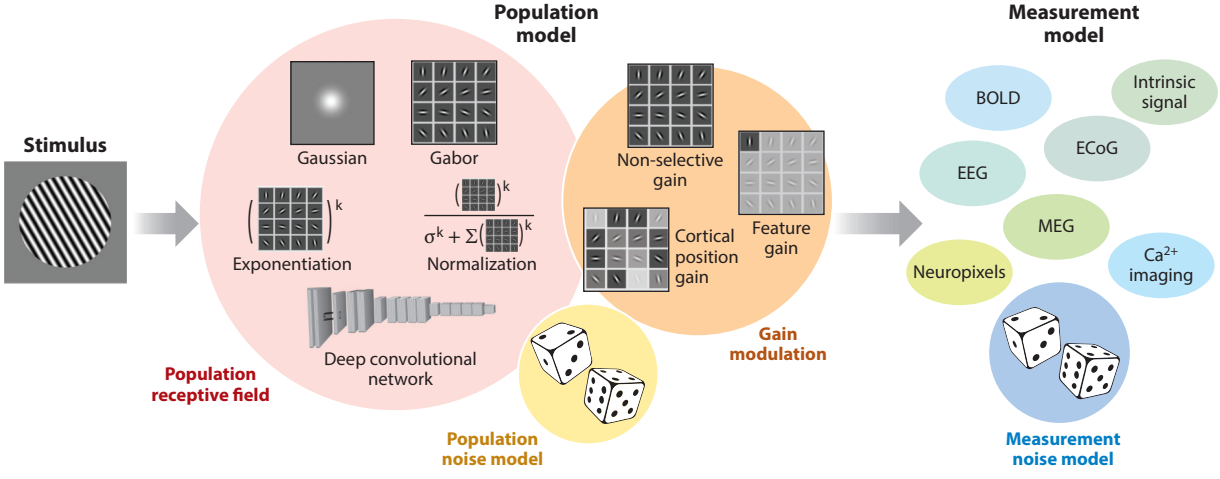


Figure 7

Consensus model comparison approach. The consensus model provides a framework for testing the impact of each of the computations that link the stimulus (*left*) with the measured response. The population model (*center*) represents different hypotheses about neural function that will give rise to different measurements, as predicted through measurement models (*right*). Population models gain experimental support by making testable and falsifiable predictions for different measurement modalities. For example, illustrated are receptive field models with varying degrees of complexity, ranging from a simple Gaussian model to a deep convolutional neural network. The consensus model incorporates various forms of gain field modulation, which can account for different sources of bias such as columnar architecture (cortical position gain) or cardinal orientation bias (feature gain). Sources of population noise also should be explicitly modeled. The responses of the population models will manifest in different ways for different measurements, for example, when measured with different functional MRI contrast mechanisms (BOLD or CBV) or via different physiological signals such as Ca^{2+} or electrophysiological signals (EEG, MEG, ECoG, or neuropixels). Thus, measurement models that also incorporate assumptions about different sources of measurement noise are required to make explicit predictions for each population model. The goal of bridging across measurement modalities is met by a population model with gain modulation, which can best account for responses measured across a wide array of different measurement techniques. Abbreviations: BOLD, blood oxygen level–dependent; CBV, cerebral blood volume; ECoG, electrocorticography; EEG, electroencephalogram; MEG, magnetoencephalography; MRI, magnetic resonance imaging.

well. For example, a feature gain field could be instantiated in the cortex in several ways, including a bias in the strength of feedforward stimulus drive inherited from earlier stages or fed back from higher-order cortical areas (Nasr & Tootell 2012). A feature gain field could also be intrinsic to V1, instantiated by differences in the proportion of cells preferring a particular stimulus orientation, differences in response amplitude, or differences in tuning bandwidth. In addition, existing models of the optics of the eye and retinal processing (Cottaris et al. 2019) could be leveraged as a first stage of processing the stimulus, before transforming through receptive field models, and such front-end mechanisms could themselves lead to orientation biases.

The consensus model (**Figure 7**) is also agnostic to the particular receptive field model used, with possibilities ranging from very simple (elongated Gaussian V1-like receptive fields and circular center-surround LGN-like receptive fields) to much more complex models. More complex receptive field models may be useful for characterizing the impact of computations such as divisive normalization (Carandini & Heeger 2012) and asymmetric surround suppression (Cavanaugh et al. 2002a,b; Tanaka & Ohzawa 2009; Walker et al. 1999), both of which could themselves contribute to an orientation bias across the population.

6. WHY DOES THIS MATTER? AN EXAMPLE FROM VISUAL ATTENTION

Why does all of this matter? If one can decode orientation from imaging measurements, and if we know that the visual cortex is orientation selective, then does the underlying mechanism really matter? Even if it is imperfect, why not use decoding to discover new principles of cognitive functions? Indeed, orientation decoding has been used as a means to investigate a variety of cognitive processes such as attention (Ester et al. 2016, Garcia et al. 2013, Jehee et al. 2011, Ling et al. 2015, Scolari & Serences 2009, Scolari et al. 2012) and working memory (Bettencourt & Xu 2016; Ester et al. 2013, 2015a,b; Harrison & Tong 2009; Lorenc et al. 2018; Yu & Shim 2017), among many others (Tong & Pratte 2012). However, without clearly identifying how the stimulus is being decoded, what might seem like innocuous changes to an experimental design could give unexpected results. For example, asking subjects to match orientations but changing among circular, oval, and square gratings will give rise to different decoded cortical signals unrelated to orientation. Manipulations in working memory might lead to subtle changes in the spatial frequency of the represented stimulus that could wreak havoc on the interpretation of results.

An even more fundamental issue concerns what inferences about neural mechanisms can be made from orientation decoding. Consider, for example, the question of whether sensory responses change their selectivity or their gain with selective attention (Carrasco 2011). These are fundamentally different ways in which sensory codes could adapt to behavioral demands, as changes in selectivity imply that the population code for sensory stimuli has changed, while gain changes are akin to turning up the volume knob on the most informative neural populations. While at first glance it might appear that an analysis that retrieves a curve such as the one shown in **Figure 1*b*** would be perfectly suited to address this question for human population responses, it is not.

The curve in **Figure 1*b***, which visually shares the features of an orientation tuning curve measured in single units, is the result of an inverted encoding model analysis (Sprague et al. 2018). A forward encoding model can encode visual stimuli in a lower-dimensional representation (Brouwer & Heeger 2009), for example, by taking an image and encoding it as the output of a small number of orientation-tuned filters. Population responses are then fit as weighted sums of these filter outputs. This approach has been used for basic features such as color (Brouwer & Heeger 2009, Yu & Shim 2017), orientation (Brouwer & Heeger 2011; Byers & Serences 2014;

Chong et al. 2016; Ester et al. 2013, 2015b, 2016; Garcia et al. 2013; Ho et al. 2012; Liu et al. 2018; Lorenc et al. 2018; Scolari et al. 2012; Yu & Shim 2017), and direction of motion (Chen et al. 2015, Saproo & Serences 2014) and is an example of the type of population model that we advocate for above. An inverted encoding model goes one step further by using matrix inversion of the weights on a left-out set of data to try to infer how the model behaves.

Such an inverted encoding model representation should not be confused for a tuning function, as it cannot be interpreted as a unique representation of population selectivity (Gardner & Liu 2019). Indeed, this representation may not be due to orientation selectivity at all. The inverted encoding model analysis is simply a linear regression problem that is designed to produce whatever shape of function was assumed in the creation of the model. That is, the forward model encoding functions form a basis set for the responses of voxels, and inversion recovers this basis set. Any rotation of the basis set axes will not change the ability of the analysis to account for variance in the data. Thus, rather than an emergent property of the analysis that discovers the orientation selectivity of the population, the shape of the basis set functions is a basic assumption baked into the analysis. The shape of the resulting function, be it unimodal, bimodal, skinny, skewed, or wide, is a consequence of that analysis choice, and the degree to which that functional shape is recovered is a reflection of the signal-to-noise ratio of the data. Examination of the recovered function thus cannot tell us what stimuli give what population response, the way that a tuning function does (**Figure 8*a***). Even more problematic is the fact that, if orientation selectivity is a consequence of stimulus vignetting and the spatial frequency tuning of the population, then

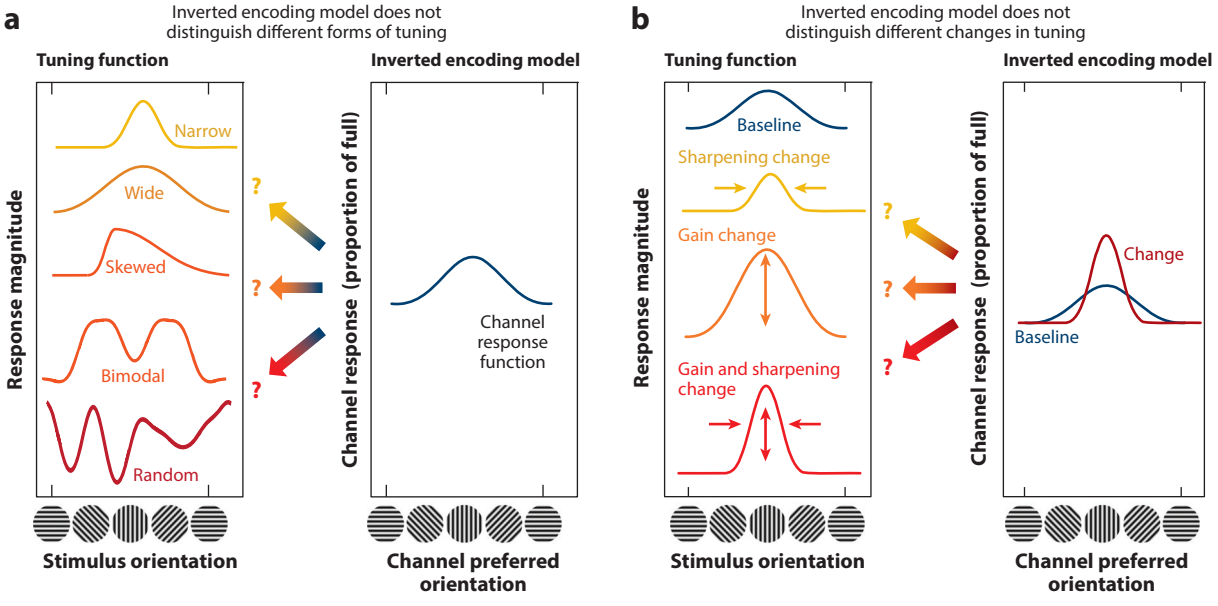


Figure 8

Interpretational ambiguities of inverted encoding model. (*a*) The inverted encoding model approach does not distinguish between different neuronal tuning functions. Tuning of single units or populations of units could take any of the shapes depicted at the left and still give rise to a channel response function that looks perfectly unimodal because the channel response function is simply a reconstruction of the assumed form of model tuning and not reflective of any tuning function in the neural system (Gardner & Liu 2019). (*b*) Changes in tuning functions that might be hypothesized to underlie change with attention, for example, gain or sharpening of tuning functions (*left*), cannot be distinguished through examining channel response functions, as they will all manifest in a better ability to reconstruct the assumed form of the model tuning function (*right*).

the recovered function, though nominally described in terms of orientation, is not even due to orientation selectivity itself.

Even if the inverted encoding model were due to orientation selectivity, sensory gain or selectivity change would not be distinguishable (Liu et al. 2018). That is, imagine a case in which one obtains a change in inverted encoding model output with attention (**Figure 8*b***, *right*). Visually, this invites the interpretation of a tuning function with changed selectivity. However, this interpretation is not warranted, as the same result would have been obtained if there was a gain change. Moreover, changes in population selectivity, for example, if a subset of the population of neurons were inhibited to achieve stronger selectivity for the attended orientation, would have similar results. These possibilities and many others would result in the same outcome because they all change signal-to-noise ratio. A more appropriate representation of the analysis is provided by stimulus-, rather than model-, referred representation (van Bergen et al. 2015), which displays what one could infer about the stimulus given the responses, rather than the state of an arbitrary model (Gardner & Liu 2019).

The way out of this ambiguous state of affairs is to make model comparison central to the approach. If one wants to determine whether a change in gain or a change in selectivity occurred, then one should model those effects and see which best explains the data. Then, model comparison statistics can determine if there is a discriminable difference in ability to predict the data. If there is, then one has evidence of one mechanism over another. If there is not, then one cannot make that conclusion. Sometimes, subtleties of the analysis, such as how one uses cross-validation to build and test models, can make a difference in what one can conclude. For example, fitting a model to two conditions that only vary in signal-to-noise ratio and then testing each condition separately may result in an average model that reduces the effect of signal-to-noise differences, whereas fitting two different models might accentuate those effects. However, model comparison again comes to the rescue, as it makes clear what can and cannot be differentiated from the whole analysis.

7. BEYOND ORIENTATION

To what extent is the need to have the right population model and proper consideration of the stimulus applicable to other stimulus features beyond orientation? Decoding of other basic visual stimulus features, such as direction of motion (Beckett et al. 2012, Kamitani & Tong 2006, Sterzer et al. 2006, Wang et al. 2014), speed and temporal frequency (Hammett et al. 2013, Nishimoto et al. 2011, Vintch & Gardner 2014), color (Bannert & Bartels 2018, Brouwer & Heeger 2009), spatial frequency (Kay et al. 2008), and eye of origin (Larsson et al. 2017), has been found; are these results evidence for superresolution decoding? Several results strongly suggest against a superresolution account.

For example, motion stimuli presented with an aperture appear to induce larger responses at the leading edge of the motion (Beckett et al. 2012, Wang et al. 2014), an effect that has been described as consistent with predictive coding (Schellekens et al. 2016) and that could be the basis for slight changes in retinotopic mapping (Whitney et al. 2003) with moving stimuli. Evidence exists that there are retinotopic differences in the distribution of color-selective responses in the cortex (Vanni et al. 2006), which could arise from differences in the density of short-wave cones (Curcio et al. 1991) and intrinsically sensitive retinal ganglion cells (Horiguchi et al. 2013) between the foveal and peripheral retina. Chromatic aberration effects from the lens could also distort the retinotopic position of different wavelengths. Spatial frequency selectivity (Aghajari et al. 2020) covaries with receptive field size, and thus, at more eccentric locations where receptive fields are larger (Broderick et al. 2018), spatial frequency selectivity should shift toward lower

spatial frequencies (Kay et al. 2008). Indeed, spatial frequency sensitivity has been used to infer receptive field size (Keliris et al. 2019). The temporal frequency response of two channels can be determined across visual areas (Horiguchi et al. 2009, Stigliani et al. 2017); thus, temporal frequency selectivity can be determined at a coarse spatial scale. Differences in distribution of magnocellular and parvocellular types in the foveal compared to the peripheral retina could give rise to cortical retinotopic distribution of spatial and temporal frequency selectivity. Astigmatism in some subjects could cause different amounts of spatial blur at different places in the retina. Eye of origin can be decoded not only from the primary visual cortex, but also from higher-order visual areas that are not thought to have strongly monocular responses (Larsson et al. 2017). Features such as image contrast (Avidan et al. 2002, Birman & Gardner 2018, Kastner & Pinsk 2004, Logothetis et al. 2001, Tootell et al. 1995) and motion coherence (Birman & Gardner 2018, Costagli et al. 2014), which affect the visibility of stimuli, give rise to monotonic responses that can be measured in the average response within a cortical area. All of these effects suggest that, rather than superresolution decoding, basic visual features are available to BOLD measurement through coarse-spatial-scale representations in the cortex.

Categorical representation has long been known to be encoded on a large spatial scale in humans, and while decoding analyses show that more can be learned by examining the fine pattern of responses, these results do not require superresolution explanation. Selectivities for faces (Kanwisher et al. 1997), places (Epstein & Kanwisher 1998), body parts (Downing et al. 2001), visual word forms (Cohen et al. 2000), and other visual categories have been found and are grouped into cortical areas along the ventral temporal cortex. Larger-scale groups are separated by the midfusiform sulcus into representations for animate and inanimate and other proposed large-scale categories (Chao et al. 1999, Kriegeskorte et al. 2008, Grill-Spector & Weiner 2014). Examining more than just the category that elicits the strongest response shows that category can be decoded from distributed response patterns (Haxby et al. 2001). However, this does not require superresolution. It can simply result from reliable differences in response to the nonpreferred category. Classification analysis can also decode position from the ventral visual cortex (Carlson et al. 2011, Schwarzlose et al. 2008), which is thought to be position invariant (Ito et al. 1995, Rust & DiCarlo 2010). While these results show that more fine-grained visual information can be inferred from examination of fine-scaled patterns of responses, they do not necessarily require a superresolution account. Indeed, a foveal bias for faces compared to other stimuli has long been known (Levy et al. 2001), and topographic response preference within areas at a coarse spatial scale could account for decoding results. Moreover, it has been suggested that representation of some complex properties of visual stimuli, such as numerosity (Eger et al. 2009), is mapped along the cortical surface (Harvey et al. 2013) at a scale easily accessible through imaging.

8. BROADER CONCLUSIONS ON COMPUTATIONAL MODELING FOR HUMAN NEUROSCIENCE

What general lessons can be learned for human neuroscience from this effort to understand the basis of the orientation signal for decoding? The goal of making models that can explain measurements of the human visual cortex should not be minimized. After all, visual neuroscience has long sought better models of the properties of single neurons. Models of single units, such as contrast or motion energy models (Adelson & Bergen 1985, Watson & Ahumada 1985), have been refined and expanded to include computational motifs such as static output nonlinearities and normalization that can account for a wide range of phenomena. The ubiquity of the computations that have been found has led to the idea that some computations such as divisive normalization should be considered canonical because they occur in many different systems (Carandini & Heeger 2012).

PROPERTIES OF A POPULATION MODEL

- Explicitly referenced: Is it clear what the population model is a model of?
- Generality: Does the model generalize easily to a new situation?
- Goodness-of-fit: Does the population model explain a sizable amount of the explainable variance of the data?
- Reasonable alternatives: Has the population model been compared to reasonable alternatives?
- Validation: Has the population model been validated through simulation?
- Interpretable: Does the population model provide insight into the phenomenon that it is modeling?
- Robust: Is the population model robust to small perturbations of parameters?
- Specific: Is the population model matched in complexity to the phenomenon that it is modeling?
- Bridging: Can the model bridge across different measurement modalities and species?

Extending and developing new models that can account for population responses rather than just single units is important, as the predictions can be substantially different (Hara et al. 2014, Mante & Carandini 2005), and principles of computation viewed from the perspective of populations can demonstrate that properties of the world need not be encoded in each individual neuron. Indeed, significant advances have been made in understanding how mixed selectivity (Rigotti et al. 2013) and dynamically changing activity across populations of neurons (Mante et al. 2013) can encode aspects of the world. Building explicit models of populations should be a goal in and of itself to further understand the human visual cortex.

As with any modeling endeavor, there are a few key ingredients that make for a rigorous approach (see sidebar titled Properties of a Population Model); many of them stem from basic modeling considerations, but they come up frequently and thus deserve attention.

8.1. Explicitly Referenced

Perhaps the most salient problems for population models occur when they are not explicitly referenced to the phenomenon that they are modeling. This is fundamentally the difference between an analysis and a model. Consider the difference between a support vector machine (SVM) analysis and the population model in **Figure 7**. Both are mathematical models of the data, but the SVM does not reference any properties of the visual system. Instead, it treats the data as generic matrices and obtains an analytic result. As with any analytic approach, there is danger in blindly trying algorithms until one obtains a desired result (Botvinik-Nezer et al. 2020). Explicitly referencing the model to the processes that one is trying to explain provides guidance as to the choice of model and makes for cumulative science. That is, one chooses to model the phenomenon, for example, the receptive field structure, based on the cumulative knowledge of visual responses and not as an arbitrary choice, subject to the lack of specificity for the phenomenon. Building explicitly referenced population models, rather than analyzing, is a mindset. While an SVM is technically a model of the data, are any biological processes well modeled by the assumptions of an SVM?

8.2. Generality

Making visual models general by making them image computable can easily resolve issues that might otherwise engender endless debate. Consider the methods of mitigating the impact of the edge of the aperture to evaluate actual orientation selectivity with population measurements. One might consider smoothing or blurring the edge of the aperture (Warren et al. 2014); however, this simply spreads the aperture across a larger region of space, causing a more pronounced vignetting

effect (Carlson 2014). Alternatively, one might use a larger stimulus aperture and characterize orientation selectivity in voxels with receptive fields farther away from the aperture edge (Wardle et al. 2017). However, population receptive fields are large, and identifying responses that are truly not affected by the aperture is difficult. Indeed, with sufficient averaging and improved signal-to-noise ratio, many more voxels are modulated by stimuli and tasks than is typically revealed by statistical thresholding (Gonzalez-Castillo et al. 2012). Moreover, vignetting effects are also expected near the fovea, and because of cortical magnification, such effects may be evident in many voxels. Rather than arguing about whether any of these methods of removing vignetting effects reveals true orientation selectivity, one can use a concrete population model, which makes concrete predictions. Passing a proposed stimulus through an image-computable population model enables one to make explicit predictions, rather than proposing an endless array of ad hoc changes to stimuli that may or may not have the desired effect.

8.3. Validation

Ground-truth simulation should be performed before a conclusion is drawn about the outcome of a model. Model fitting will sometimes be unable to distinguish certain parameters because of the ambiguity of the fit. For example, the size or aspect ratio of population receptive field models may not be well determined (Lerma-Usabiaga et al. 2020a, Silson et al. 2018). These model parameters cannot then be meaningfully interpreted. The solution is to simulate data with realistic amounts of noise to see if the model can recover ground-truth (Lerma-Usabiaga et al. 2020b). If the model cannot recover simulated ground-truth, or if different models that one wishes to compare fit the data equally well, then testing on real data is hopeless.

8.4. Goodness-of-Fit

A particularly important issue for population models is whether they can account for a substantial amount of variance. Goodness-of-fit can be evaluated by using cross-validation and reporting the amount of variance that can be accounted for in data that the model was not built to predict. Typically, one examines the amount of explainable variance by considering how well the data can predict itself through splitting up the data and computing the correlation between different portions of the data, sometimes scaling this value. While this is a sensible metric, it can be abused if the amount of explainable variance is small. Often, the values are not reported or are unscaled, and it is impossible to know how repeatable response patterns are. Building inferences about how models explain data, for example, finding evidence for spatiotopic representations for responses that are not well fit by any model (Gardner et al. 2008), can lead to unwarranted conclusions.

8.5. Robustness

Some models can be sensitive to the choice of parameters, and thus, their predictions are not robust. If a model is brittle in this way, then the inferences may not be strongly constrained. For example, if a slight change in the parameters of a model can cause predictions to qualitatively flip (Alink et al. 2018, Ramírez & Merriam 2020), then the model may not be meaningful interpreted. In such cases, it is important that these parameters be constrained by empirical measurements. For example, a model may make an interesting prediction, but only if the signal-to-noise ratio in the simulation is set to levels that cannot be achieved experimentally (Ramírez & Merriam 2020). Robustness is also important when comparing models. One model may account for more variance than another, assuming a given set of parameters. However, it makes little difference if the parameters required to differentiate models are not biologically plausible.

8.6. Reasonable Alternatives

Even when a population model can account for a significant amount of variance, this does not imply that the model is a good model (Pitt & Myung 2002). Significance is typically determined by testing against a null model, for example, by permuting labels on the data and determining whether the model captures more variance than is expected by chance. However, this approach does not guarantee that the model is a good model. Instead, one should provide alternative, plausible models and compare them using model comparison statistics. For example, a complex population model with orientation, spatial frequency, and spatial tuning, like a Gabor wavelet pyramid, can account for a statistically significant proportion of variance. However, if removing specificity for orientation and spatial frequency yields a model that performs just as well, given the number of fit parameters needed, then the data may only need to account for the position and size of the receptive field.

8.7. Interpretability

A model that explains more of the data variance is not necessarily one that has the most explanatory power. While complex models like deep convolutional networks may provide a better ability to squeeze out a few more percentage points of explained variance, they may do so with little added interpretability. Indeed, the effort to understand how deep learning models arrive at their predictions has become a growing subfield called explainable AI (Krishnapuram et al. 2016, Lundberg & Lee 2017, Molnar 2020, Owen & Prieur 2017). A related issue is that highly parameterized models may not have one-to-one correspondence between model features and neural responses; a deep neural network model for a single neuron may require a linear combination of many convolutional filters to reach high explainable variance (Yamins et al. 2014). Seeking models with one-to-one correspondence (Higgins et al. 2020) between model features and neural responses can provide models that are more interpretable. Interpretability may be at odds with some of the other criteria listed in this section, such as goodness-of-fit. A more complex model may fit the data better, but interpretability is itself a goal, and in many cases, we should choose the simpler model.

8.8. Bridging

A population model should be adaptable to a range of neuroimaging modalities, including coarse-scale methods like BOLD MRI and magnetoencephalography (Hermes et al. 2019, Kupers et al. 2020), as well as intrinsic signal optical imaging, calcium imaging, and neuropixels (**Figure 7**), thus providing a mechanism for bridging different measurement modalities and even species.

8.9. Conclusion

Not all population models need to possess all of these properties. For example, an engineering application for brain machine interfaces may be less interested in interpretability and more interested in generality, i.e., being able to predict the correct output. Nonetheless, even for such applications, how well-behaved the models will be in new situations is heavily constrained by how well we understand the system that we are modeling, and thus, it is perilous to ignore other considerations of good population modeling. The best model might not have the highest goodness-of-fit, for example, but might offer some other of these criteria (such as bridging across measurement modalities) that are particularly important for addressing the scientific question of interest.

9. CONCLUSIONS

The discovery that linear classification analysis could decode basic visual stimulus properties (Kamitani & Tong 2005, 2006) without painstaking high-spatial-resolution imaging is clearly a

landmark finding that propelled the analysis of patterns of activity (Haxby et al. 2001) into a powerful way to learn about the human brain. While some may consider the subsequent effort investigating whether the analysis retrieves superresolution information about cortical columns an affront to this achievement, it was not. In fact, this effort has led to a better foundation for our understanding of how pattern analysis reveals the orientation of a stimulus by showing that map-like topographic signals at a spatial scale easily measurable with BOLD imaging can be formed by population responses that need only be selective for spatial frequency. This provides a framework for building population models (Dumoulin & Wandell 2008, Kay et al. 2008) that form a foundation for the many uses that pattern analysis has been put to in the study of the human brain (Kriegeskorte et al. 2008, Tong & Pratte 2012). Consensus models of the type that we propose in this review (**Figure 7**) can be used to quantitatively adjudicate between different proposed mechanisms for decoding. Use of these models builds on the strong tradition of computational modeling in visual neuroscience and thus forms a foundation for cumulative and replicable use and interpretation of the analysis of patterns of activity in the human brain.

SUMMARY POINTS

1. Decoding of orientation is explained by a population model that shows that stimulus vignetting can give rise to stronger responses for radial orientations.
2. Computational analysis can thus capitalize on information encoded in maps, but not because they are able to retrieve superresolution information.
3. Models based on single-unit receptive fields only are not appropriate for modeling population responses; instead, the population of neurons needs to be taken into account.
4. Population models without selectivity to orientation can still give rise to an orientation-selective signal.
5. If categorical errors of interpretation can occur for orientation, then other, more complex features and cognitive responses that rely on similar techniques may be based on incorrect inferences.
6. Population models, rather than blind analyses, provide a way to make valid interpretations of human neuroscience measurements.
7. Population models should be explicitly referenced, general, validated, interpretable, robust, and specific; be compared against reasonable alternatives; and display high goodness-of-fit so that they can form the basis for a cumulative and replicable approach to making valid inferences from human neuroscience measurements.
8. Population models need to be developed to explain decoding of basic stimulus features beyond orientation, such as motion direction.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

J.L.G is supported by Research to Prevent Blindness and the Lions Clubs International Foundation. E.P.M. is supported by the Intramural Research Program of the National Institute of Mental

Health (ZIA-MH-002909). We thank David Heeger, Brian Wandell, Denis Schluppeck, Kendrick Kay, Taosheng Liu, Austin Kuo, Akshay Jagadeesh, Jun Hwan Ryu, Josh Wilson, Zvi Roth, Fernando Ramirez, Tina Liu, Tomas Knapen, and Maryam Vaziri-Pashkam.

LITERATURE CITED

Adams DL, Piserchia V, Economides JR, Horton JC. 2015. Vascular supply of the cerebral cortex is specialized for cell layers but not columns. *Cereb. Cortex* 25(10):3673–81

Adams DL, Sincich LC, Horton JC. 2007. Complete pattern of ocular dominance columns in human primary visual cortex. *J. Neurosci.* 27(39):10391–403

Adelson EH, Bergen JR. 1985. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am.* 2(2):284–99

Aghajari S, Vinke LN, Ling S. 2020. Population spatial frequency tuning in human early visual cortex. *J. Neurophysiol.* 123(2):773–85

Albrecht DG, Geisler WS. 1991. Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Vis. Neurosci.* 7(6):531–46

Alink A, Abdulrahman H, Henson RN. 2018. Forward models demonstrate that repetition suppression is best modelled by local neural scaling. *Nat. Commun.* 9:3854

Alink A, Krugliak A, Walther A, Kriegeskorte N. 2013. fMRI orientation decoding in V1 does not require global maps or globally coherent orientation stimuli. *Front. Psychol.* 4:493

Alink A, Walther A, Krugliak A, Kriegeskorte N. 2017. Local opposite orientation preferences in V1: fMRI sensitivity to fine-grained pattern information. *Sci. Rep.* 7:7128

Avidan G, Harel M, Hendler T, Ben-Bashat D, Zohary E, Malach R. 2002. Contrast sensitivity in human visual areas and its relationship to object recognition. *J. Neurophysiol.* 87(6):3102–16

Bannert MM, Bartels A. 2018. Human V4 activity patterns predict behavioral performance in imagery of object color. *J. Neurosci.* 38(15):3657–68

Barlow HB. 1953. Summation and inhibition in the frog's retina. *J. Physiol.* 119(1):69–88

Beckett A, Peirce JW, Sanchez-Panchuelo R-M, Francis S, Schluppeck D. 2012. Contribution of large scale biases in decoding of direction-of-motion from high-resolution fMRI data in human early visual cortex. *NeuroImage* 63(3):1623–32

Bettencourt KC, Xu Y. 2016. Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nat. Neurosci.* 19(1):150–57

Birman D, Gardner JL. 2018. A quantitative framework for motion visibility in human cortex. *J. Neurophysiol.* 120(4):1824–39

Blinder P, Tsai PS, Kaufhold JP, Knutsen PM, Suhl H, Kleinfeld D. 2013. The cortical angiome: an interconnected vascular network with noncolumnar patterns of blood flow. *Nat. Neurosci.* 16(7):889–97

Bonds AB. 1989. Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Vis. Neurosci.* 2(1):41–55

Botvinik-Nezer R, Holzmeister F, Camerer CF, Dreber A, Huber J, et al. 2020. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* 582(7810):84–88

Boynton GM. 2005. Imaging orientation selectivity: decoding conscious perception in V1. *Nat. Neurosci.* 8(5):541–42

Broderick W, Benson N, Simoncelli E, Winawer J. 2018. Mapping spatial frequency preferences in the human visual cortex. *J. Vis.* 18(10):253

Brouwer GJ, Heeger DJ. 2009. Decoding and reconstructing color from responses in human visual cortex. *J. Neurosci.* 29(44):13992–93

Brouwer GJ, Heeger DJ. 2011. Cross-orientation suppression in human visual cortex. *J. Neurophysiol.* 106(5):2108–19

Byers A, Serences JT. 2014. Enhanced attentional gain as a mechanism for generalized perceptual learning in human visual cortex. *J. Neurophysiol.* 112(5):1217–27

Campbell FW, Cleland BG, Cooper GF, Enroth-Cugell C. 1968. The angular selectivity of visual cortical cells to moving gratings. *J. Physiol.* 198(1):237–50

Carandini M, Heeger DJ. 2012. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13(1):51–62
Carandini M, Heeger DJ, Movshon JA. 1997. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* 17(21):8621–44
Carlson TA. 2014. Orientation decoding in human visual cortex: new insights from an unbiased perspective. *J. Neurosci.* 34(24):8373–83
Carlson TA, Hogendoorn H, Fonteijn H, Verstraten FAJ. 2011. Spatial coding and invariance in object-selective cortex. *Cortex* 47(1):14–22
Carlson TA, Wardle SG. 2015. Sensible decoding. *NeuroImage* 110:217–18
Carrasco M. 2011. Visual attention: the past 25 years. *Vis. Res.* 51(13):1484–525
Cavanaugh JR, Bair W, Movshon JA. 2002a. Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J. Neurophysiol.* 88(5):2530–46
Cavanaugh JR, Bair W, Movshon JA. 2002b. Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *J. Neurophysiol.* 88(5):2547–56
Cavina-Pratesi C, Kentridge RW, Heywood CA, Milner AD. 2010. Separate channels for processing form, texture, and color: evidence from fMRI adaptation and visual object agnosia. *Cereb. Cortex* 20(10):2319–32
Chao LL, Haxby JV, Martin A. 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat. Neurosci.* 2:913–19
Chen N, Bi T, Zhou T, Li S, Liu Z, Fang F. 2015. Sharpened cortical tuning and enhanced cortico-cortical communication contribute to the long-term neural mechanisms of visual motion perceptual learning. *NeuroImage* 115:17–29
Cheng K, Waggoner RA, Tanaka K. 2001. Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. *Neuron* 32(2):359–74
Chong E, Familiar AM, Shim WM. 2016. Reconstructing representations of dynamic visual objects in early visual cortex. *PNAS* 113(5):1453–58
Chong TT-J, Cunnington R, Williams MA, Kanwisher N, Mattingley JB. 2008. fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Curr. Biol.* 18(20):1576–80
Clifford CWG, Mannion DJ. 2014. Orientation decoding: sense in spirals? *NeuroImage* 110:219–22
Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, et al. 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(2):291–307
Costagli M, Ueno K, Sun P, Gardner JL, Wan X, et al. 2014. Functional signalers of changes in visual stimuli: cortical responses to increments and decrements in motion coherence. *Cereb. Cortex* 24(1):110–18
Cottaris NP, Jiang H, Ding X, Wandell BA, Brainard DH. 2019. A computational-observer model of spatial contrast sensitivity: effects of wave-front-based optics, cone-mosaic structure, and inference engine. *J. Vis.* 19(4):8
Curcio CA, Allen KA, Sloan KR, Lerea CL, Hurley JB, et al. 1991. Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *J. Comp. Neurol.* 312(4):610–24
Das A, Gilbert CD. 1997. Distortions of visuotopic map match orientation singularities in primary visual cortex. *Nature* 387(6633):594–98
de Beeck HPO. 2010. Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses? *NeuroImage* 49(3):1943–48
de No RL, Fulton JF. 1938. Architectonics and structure of the cerebral cortex. In *Physiology of the Nervous System*, ed. JF Fulton, pp. 291–330. Oxford, UK: Oxford Univ. Press
Dinstein I, Hasson U, Rubin N, Heeger DJ. 2007. Brain areas selective for both observed and executed movements. *J. Neurophysiol.* 98(3):1415–27
Dobbins IG, Schnyer DM, Verfaellie M, Schacter DL. 2004. Cortical activity reductions during repetition priming can result from rapid response learning. *Nature* 428(6980):316–19
Downing PE, Jiang Y, Shuman M, Kanwisher N. 2001. A cortical area selective for visual processing of the human body. *Science* 293(5539):2470–73
Dumoulin SO, Fracasso A, van der Zwaag W, Siero JCW, Petridou N. 2018. Ultra-high field MRI: advancing systems neuroscience towards mesoscopic human brain function. *NeuroImage* 168:345–57

Dumoulin SO, Wandell BA. 2008. Population receptive field estimates in human visual cortex. *NeuroImage* 39(2):647–60

Eger E, Michel V, Thirion B, Amadon A, Dehaene S, Kleinschmidt A. 2009. Deciphering cortical number coding from human brain activity patterns. *Curr. Biol.* 19(19):1608–15

Engel SA, Glover GH, Wandell BA. 1997. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* 7(2):181–92

Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, et al. 1994. fMRI of human visual cortex. *Nature* 369(6481):525

Epstein R, Kanwisher N. 1998. A cortical representation of the local visual environment. *Nature* 392(6676):598–601

Ester EF, Anderson DE, Serences JT, Awh E. 2013. A neural measure of precision in visual working memory. *J. Cogn. Neurosci.* 25(5):754–61

Ester EF, Sprague TC, Serences JT. 2015a. Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron* 87(4):893–905

Ester EF, Sprague TC, Serences JT. 2015b. Visual working memory representations are distributed throughout human cortex. *J. Vis.* 15(12):1115

Ester EF, Sutterer DW, Serences JT, Awh E. 2016. Feature-selective attentional modulations in human frontoparietal cortex. *J. Neurosci.* 36(31):8188–99

Fahey PG, Muhammad T, Smith C, Froudarakis E, Cobos E, et al. 2019. A global map of orientation tuning in mouse visual cortex. bioRxiv 745323. **https://doi.org/10.1101/745323**

Fang F, Murray SO, Kersten D, He S. 2005. Orientation-tuned fMRI adaptation in human visual cortex. *J. Neurophysiol.* 94(6):4188–95

Finn ES, Huber L, Jangraw DC, Molfese PJ, Bandettini PA. 2019. Layer-dependent activity in human prefrontal cortex during working memory. *Nat. Neurosci.* 22(10):1687–95

Freeman J, Brouwer GJ, Heeger DJ, Merriam EP. 2011. Orientation decoding depends on maps, not columns. *J. Neurosci.* 31(13):4792–804

Freeman J, Heeger DJ, Merriam EP. 2013. Coarse-scale biases for spirals and orientation in human visual cortex. *J. Neurosci.* 33(50):19695–703

Furmanski CS, Engel SA. 2000. An oblique effect in human primary visual cortex. *Nat. Neurosci.* 3(6):535–36

Garcia JO, Srinivasan R, Serences JT. 2013. Near-real-time feature-selective modulations in human cortex. *Curr. Biol.* 23(6):515–22

Gardner JL. 2010. Is cortical vasculature functionally organized? *NeuroImage* 49(3):1953–56

Gardner JL, Anzai A, Ohzawa I, Freeman RD. 1999. Linear and nonlinear contributions to orientation tuning of simple cells in the cat's striate cortex. *Vis. Neurosci.* 16(6):1115–21

Gardner JL, Liu T. 2019. Inverted encoding models reconstruct an arbitrary model response, not the stimulus. *eNeuro* 6(2):ENEURO.0363-18.2019

Gardner JL, Merriam EP, Movshon JA, Heeger DJ. 2008. Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *J. Neurosci.* 28(15):3988–99

Gardner JL, Sun P, Waggoner RA, Ueno K, Tanaka K, Cheng K. 2005. Contrast adaptation and representation in human early visual cortex. *Neuron* 47(4):607–20

Girshick AR, Landy MS, Simoncelli EP. 2011. Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* 14(7):926–32

Gonzalez-Castillo J, Saad ZS, Handwerker DA, Inati SJ, Brenowitz N, Bandettini PA. 2012. Whole-brain, time-locked activation with simple tasks revealed using massive averaging and model-free analysis. *PNAS* 109(14):5487–92

Grill-Spector K, Henson R, Martin A. 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10(1):14–23

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(1):187–203

Grill-Spector K, Weiner KS. 2014. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* 15(8):536–48

Grinvald A, Lieke E, Frostig RD, Gilbert CD, Wiesel TN. 1986. Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature* 324(6095):361–64

Hallum LE, Landy MS, Heeger DJ. 2011. Human primary visual cortex (V1) is selective for second-order spatial frequency. *J. Neurophysiol.* 105(5):2121–31
Hammett ST, Smith AT, Wall MB, Larsson J. 2013. Implicit representations of luminance and the temporal structure of moving stimuli in multiple regions of human visual cortex revealed by multivariate pattern classification analysis. *J. Neurophysiol.* 110(3):688–99
Hara Y, Pestilli F, Gardner JL. 2014. Differing effects of attention in single-units and populations are well predicted by heterogeneous tuning and the normalization model of attention. *Front. Comput. Neurosci.* 8:12
Harrison RV, Harel N, Panesar J, Mount RJ. 2002. Blood capillary distribution correlates with hemodynamic-based functional imaging in cerebral cortex. *Cereb. Cortex* 12(3):225–33
Harrison SA, Tong F. 2009. Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458(7238):632–35
Harvey BM, Klein BP, Petridou N, Dumoulin SO. 2013. Topographic representation of numerosity in the human parietal cortex. *Science* 341(6150):1123–26
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(5539):2425–30
Haynes J-D, Rees G. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat. Neurosci.* 8(5):686–91
Haynes J-D, Rees G. 2006. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7(7):523–34
Heeger DJ. 1992. Half-squaring in responses of cat striate cells. *Vis. Neurosci.* 9(5):427–43
Heeger DJ. 1993. Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *J. Neurophysiol.* 70(5):1885–98
Heeger DJ, Simoncelli EP, Movshon JA. 1996. Computational models of cortical visual processing. *PNAS* 93(2):623–27
Henson RN. 2016. Repetition suppression to faces in the fusiform face area: a personal and dynamic journey. *Cortex* 80:174–84
Hermes D, Petridou N, Kay KN, Winawer J. 2019. An image-computable model for the stimulus selectivity of gamma oscillations. *eLife* 8:e47035
Higgins I, Chang L, Langston V, Hassabis D, Summerfield C, et al. 2020. Unsupervised deep learning identifies semantic disentanglement in single inferotemporal neurons. arXiv:2006.14304 [q-bio.NC]
Ho T, Brown S, van Maanen L, Forstmann BU, Wagenmakers E-J, Serences JT. 2012. The optimality of sensory processing during the speed-accuracy tradeoff. *J. Neurosci.* 32(23):7992–8003
Horiguchi H, Nakadomari S, Misaki M, Wandell BA. 2009. Two temporal channels in human V1 identified using fMRI. *NeuroImage* 47(1):273–80
Horiguchi H, Winawer J, Dougherty RF, Wandell BA. 2013. Human trichromacy revisited. *PNAS* 110(3):E260–69
Horton JC, Adams DL. 2005. The cortical column: a structure without a function. *Phil. Trans. R. Soc. B* 360(1456):837–62
Horton JC, Hedley-Whyte ET. 1984. Mapping of cytochrome oxidase patches and ocular dominance columns in human visual cortex. *Phil. Trans. R. Soc. Lond. B* 304(1119):255–72
Hubel DH, Wiesel TN. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160:106–54
Huber L, Handwerker DA, Jangraw DC, Chen G, Hall A, et al. 2017. High-resolution CBV-fMRI allows mapping of laminar activity and connectivity of cortical input and output in human M1. *Neuron* 96(6):1253–63.e7
Ito M, Tamura H, Fujita I, Tanaka K. 1995. Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.* 73(1):218–26
Jehee JFM, Brady DK, Tong F. 2011. Attention improves encoding of task-relevant features in the human visual cortex. *J. Neurosci.* 31(22):8210–19
Jones JP, Palmer LA. 1987. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neurophysiol.* 58(6):1187–211

Kamitani Y, Sawahata Y. 2010. Spatial smoothing hurts localization but not information: pitfalls for brain mappers. *NeuroImage* 49(3):1949–52
Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8(5):679–85
Kamitani Y, Tong F. 2006. Decoding seen and attended motion directions from activity in the human visual cortex. *Curr. Biol.* 16(11):1096–102
Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17(11):4302–11
Kastner S, Pinsk MA. 2004. Visual attention as a multilevel selection process. *Cogn. Affect. Behav. Neurosci.* 4(4):483–500
Kay K, Jamison KW, Vizioli L, Zhang R, Margalit E, Ugurbil K. 2019. A critical assessment of data quality and venous effects in sub-millimeter fMRI. *NeuroImage* 189:847–69
Kay KN, Naselaris T, Prenger RJ, Gallant JL. 2008. Identifying natural images from human brain activity. *Nature* 452(7185):352–55
Keliris GA, Li Q, Papanikolaou A, Logothetis NK, Smirnakis SM. 2019. Estimating average single-neuron visual receptive field sizes by fMRI. *PNAS* 116(13):6425–34
Kim S-G, Fukuda M. 2008. Lessons from fMRI about mapping cortical columns. *Neuroscience* 14(3):287–99
Kriegeskorte N, Cusack R, Bandettini P. 2010. How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatiotemporal filter? *NeuroImage* 49(3):1965–76
Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, et al. 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60(6):1126–41
Krishnapuram B, Shah M, Smola A, Aggarwal C, Shen D, et al. 2016. Why should I trust you? Explaining the predictions of any classifier. In *KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–44. New York: ACM
Kuffler SW. 1953. Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.* 16(1):37–68
Kupers ER, Benson NC, Winawer J. 2020. A visual encoding model links magnetoencephalography signals to neural synchrony in human cortex. bioRxiv 2020.04.19.049197. **https://doi.org/10.1101/2020.04.19.049197**
Larsson J, Harrison C, Jackson J, Oh S-M, Zeringyte V. 2017. Spatial scale and distribution of neurovascular signals underlying decoding of orientation and eye of origin from fMRI data. *J. Neurophysiol.* 117(2):818–35
Larsson J, Harrison SJ. 2015. Spatial specificity and inheritance of adaptation in human visual cortex. *J. Neurophysiol.* 114(2):1211–26
Larsson J, Landy MS, Heeger DJ. 2006. Orientation-selective adaptation to first- and second-order patterns in human visual cortex. *J. Neurophysiol.* 95(2):862–81
Larsson J, Smith AT. 2012. fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cereb. Cortex* 22(3):567–76
Larsson J, Solomon SG, Kohn A. 2016. fMRI adaptation revisited. *Cortex* 80:154–60
Lawrence SJD, Formisano E, Muckli L, de Lange FP. 2019. Laminar fMRI: applications for cognitive neuroscience. *NeuroImage* 197:785–91
Lerma-Usabiaga G, Benson N, Winawer J, Wandell B. 2020a. Computational validity of neuroimaging software: the case of population receptive fields. *J. Vis.* 20(11):341
Lerma-Usabiaga G, Benson N, Winawer J, Wandell BA. 2020b. A validation framework for neuroimaging software: the case of population receptive fields. *PLOS Comput. Biol.* 16(6):e1007924
Levick WR, Thibos LN. 1980. Orientation bias of cat retinal ganglion cells. *Nature* 286(5771):389–90
Levick WR, Thibos LN. 1982. Analysis of orientation bias in cat retina. *J. Physiol.* 329(1):243–61
Levy I, Hasson U, Avidan G, Hendler T, Malach R. 2001. Center-periphery organization of human object areas. *Nat. Neurosci.* 4(5):533–39
Ling S, Pratte MS, Tong F. 2015. Attention alters orientation processing in the human lateral geniculate nucleus. *Nat. Neurosci.* 18(4):496–98
Lingnau A, Ashida H, Wall MB, Smith AT. 2009. Speed encoding in human visual cortex revealed by fMRI adaptation. *J. Vis.* 9(13):3

Liu T, Cable D, Gardner JL. 2018. Inverted encoding models of human population response conflate noise and neural tuning width. *J. Neurosci.* 38(2):398–408
Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412(6843):150–57
Lorenc ES, Sreenivasan KK, Nee DE, Vandenbroucke ARE, D'Esposito M. 2018. Flexible coding of visual working memory representations during distraction. *J. Neurosci.* 38(23):5267–76
Lundberg S, Lee S-I. 2017. A unified approach to interpreting model predictions. arXiv:1705.07874 [cs.AI]
Maloney RT. 2015. The basis of orientation decoding in human primary visual cortex: fine- or coarse-scale biases? *J. Neurophysiol.* 113(1):1–3
Mannion DJ, Clifford CWG. 2011. Cortical and behavioral sensitivity to eccentric polar form. *J. Vis.* 11(6):17
Mannion DJ, McDonald JS, Clifford CWG. 2009. Discrimination of the local orientation structure of spiral Glass patterns early in human visual cortex. *NeuroImage* 46(2):511–15
Mante V, Carandini M. 2005. Mapping of stimulus energy in primary visual cortex. *J. Neurophysiol.* 94(1):788–98
Mante V, Sussillo D, Shenoy KV, Newsome WT. 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503(7474):78–84
Markuerkiaga I, Barth M, Norris DG. 2016. A cortical vascular model for examining the specificity of the laminar BOLD signal. *NeuroImage* 132:491–98
Martino FD, Yacoub E, Kemper V, Moerel M, Uludag K, et al. 2018. The impact of ultra-high field MRI on cognitive and computational neuroimaging. *NeuroImage* 168:366–82
Molnar C. 2020. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. N.p.: Lulu.com
Mountcastle VB. 1957. Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J. Neurophysiol.* 20(4):408–34
Movshon JA, Thompson ID, Tolhurst DJ. 1978. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J. Physiol.* 283:53–77
Nasr S, Tootell RBH. 2012. A cardinal orientation bias in scene-selective visual cortex. *J. Neurosci.* 32(43):14921–26
Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL. 2011. Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* 21(19):1641–46
Ogawa S, Lee TM, Kay AR, Tank DW. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *PNAS* 87(24):9868–72
Ohki K, Chung S, Ch'ng YH, Kara P, Reid RC. 2005. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* 433(7026):597–603
Ohki K, Chung S, Kara P, Hübener M, Bonhoeffer T, Reid RC. 2006. Highly ordered arrangement of single neurons in orientation pinwheels. *Nature* 442(7105):925–28
Owen AB, Prieur C. 2017. On Shapley value for measuring importance of dependent inputs. *SIAM/ASA J. Uncertain. Quantif.* 5(1):986–1002
Parkes LM, Schwarzbach JV, Bouts AA, Deckers RHR, Pullens P, et al. 2005. Quantifying the spatial resolution of the gradient echo and spin echo BOLD response at 3 Tesla. *Magn. Reson. Med.* 54(6):1465–72
Peelen MV, Wiggett AJ, Downing PE. 2006. Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron* 49(6):815–22
Pitt MA, Myung IJ. 2002. When a good fit can be bad. *Trends Cogn. Sci.* 6(10):421–25
Pratte MS, Sy JL, Swisher JD, Tong F. 2015. Radial bias is not necessary for orientation decoding. *NeuroImage* 127:23–33
Ramírez FM, Merriam EP. 2020. Forward models of repetition suppression depend critically on assumptions of noise and granularity. *Nat. Commun.* 11:4732
Reid RC, Soodak RE, Shapley RM. 1987. Linear mechanisms of directional selectivity in simple cells of cat striate cortex. *PNAS* 84(23):8740–44
Rigotti M, Barak O, Warden MR, Wang X-J, Daw ND, et al. 2013. The importance of mixed selectivity in complex cognitive tasks. *Nature* 497(7451):585–90
Ringach DL. 2002. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J. Neurophysiol.* 88(1):455–63

Ringach DL. 2007. On the origin of the functional architecture of the cortex. *PLOS ONE* 2(2):e251
Rodieck RW. 1965. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vis. Res.* 5(12):583–601
Rodieck RW, Binmoeller KF, Dineen J. 1985. Parasol and midget ganglion cells of the human retina. *J. Comp. Neurol.* 233(1):115–32
Roth ZN, Heeger DJ, Merriam EP. 2018. Stimulus vignetting and orientation selectivity in human visual cortex. *eLife* 7:e37241
Rust NC, DiCarlo JJ. 2010. Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *J. Neurosci.* 30(39):12978–95
Sapountzis P, Schluppeck D, Bowtell R, Peirce JW. 2010. A comparison of fMRI adaptation and multivariate pattern classification analysis in visual cortex. *NeuroImage* 49:1632–40
Saproo S, Serences JT. 2014. Attention improves transfer of motion information between V1 and MT. *J. Neurosci.* 34(10):3586–96
Sasaki Y, Rajimehr R, Kim BW, Ekstrom LB, Vanduffel W, Tootell RBH. 2006. The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron* 51(5):661–70
Schacter DL, Wig GS, Stevens WD. 2007. Reductions in cortical activity during priming. *Curr. Opin. Neurobiol.* 17(2):171–76
Schall JD, Vitek DJ, Leventhal AG. 1986. Retinal constraints on orientation specificity in cat visual cortex. *J. Neurosci.* 6(3):823–36
Schellekens W, van Wezel RJA, Petridou N, Ramsey NF, Raemaekers M. 2016. Predictive coding for motion stimuli in human early visual cortex. *Brain Struct. Funct.* 221(2):879–90
Schwarzlose RF, Swisher JD, Dang S, Kanwisher N. 2008. The distribution of category and location information across object-selective regions in human visual cortex. *PNAS* 105(11):4447–52
Scolari M, Byers A, Serences JT. 2012. Optimal deployment of attentional gain during fine discriminations. *J. Neurosci.* 32(22):7723–33
Scolari M, Serences JT. 2009. Adaptive allocation of attentional gain. *J. Neurosci.* 29(38):11933–42
Shou T, Ruan D, Zhou Y. 1986. The orientation bias of LGN neurons shows topographic relation to area centralis in the cat retina. *Exp. Brain Res.* 64(1):233–36
Silson EH, Reynolds RC, Kravitz DJ, Baker CI. 2018. Differential sampling of visual space in ventral and dorsal early visual cortex. *J. Neurosci.* 38(9):2294–303
Simoncelli EP, Freeman WT. 1995. The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *Proceedings of the 2nd IEEE International Conference on Image Processing*, Vol. III, pp. 444–47. Piscataway, NJ: IEEE
Smith EL, Chino YM, Ridder WH, Kitagawa K, Langston A. 1990. Orientation bias of neurons in the lateral geniculate nucleus of macaque monkeys. *Vis. Neurosci.* 5(6):525–45
Sprague TC, Adam KCS, Foster JJ, Rahmati M, Sutterer DW, Vo VA. 2018. Inverted encoding models assay population-level stimulus representations, not single-unit neural tuning. *eNeuro* 5(3):ENEURO.0098-18.2018
Sterzer P, Haynes J-D, Rees G. 2006. Primary visual cortex activation on the path of apparent motion is mediated by feedback from hMT+/V5. *NeuroImage* 32(3):1308–16
Stigliani A, Jeska B, Grill-Spector K. 2017. Encoding model of temporal processing in human visual cortex. *PNAS* 114(51):E11047–56
Summerfield C, Trittschuh EH, Monti JM, Mesulam M-M, Egner T. 2008. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* 11(9):1004–6
Sun P, Gardner JL, Costagli M, Ueno K, Waggoner RA, et al. 2013. Demonstration of tuning to stimulus orientation in the human visual cortex: a high-resolution fMRI study with a novel continuous and periodic stimulation paradigm. *Cereb. Cortex* 23(7):1618–29
Sun P, Ueno K, Waggoner RA, Gardner JL, Tanaka K, Cheng K. 2007. A temporal frequency-dependent functional architecture in human V1 revealed by high-resolution fMRI. *Nat. Neurosci.* 10(11):1404–6
Swisher JD, Gatenby JC, Gore JC, Wolfe BA, Moon C-H, et al. 2010. Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *J. Neurosci.* 30(1):325–30
Tanaka H, Ohzawa I. 2009. Surround suppression of V1 neurons mediates orientation-based representation of high-order visual features. *J. Neurophysiol.* 101(3):1444–62

Tong F, Pratte MS. 2012. Decoding patterns of human brain activity. *Annu. Rev. Psychol.* 63:483–509
Tootell R, Reppas J, Kwong K, Malach R, Born R, et al. 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* 15(4):3215–30
Ugurbil K. 2016. What is feasible with imaging human brain function and connectivity using functional magnetic resonance imaging. *Phil. Trans. R. Soc. B* 371(1705):20150361
van Bergen RS, Ma WJ, Pratte MS, Jehee JFM. 2015. Sensory uncertainty decoded from visual cortex predicts behavior. *Nat. Neurosci.* 18(12):1728–30
Vanni S, Henriksson L, Viikari M, James AC. 2006. Retinotopic distribution of chromatic responses in human primary visual cortex. *Eur. J. Neurosci.* 24(6):1821–31
Vintch B, Gardner JL. 2014. Cortical correlates of human motion perception biases. *J. Neurosci.* 34(7):2592–604
Vizioli L, Martino FD, Petro LS, Kersten D, Ugurbil K, et al. 2020. Multivoxel pattern of blood oxygen level dependent activity can be sensitive to stimulus specific fine scale responses. *Sci. Rep.* 10:7565
Walker GA, Ohzawa I, Freeman RD. 1999. Asymmetric suppression outside the classical receptive field of the visual cortex. *J. Neurosci.* 19(23):10536–53
Wandell BA, Winawer J. 2011. Imaging retinotopic maps in the human brain. *Vis. Res.* 51(7):718–37
Wang HX, Merriam EP, Freeman J, Heeger DJ. 2014. Motion direction biases and decoding in human visual cortex. *J. Neurosci.* 34(37):12601–15
Wardle SG, Ritchie JB, Seymour K, Carlson TA. 2017. Edge-related activity is not necessary to explain orientation decoding in human visual cortex. *J. Neurosci.* 37(5):1187–96
Warren SG, Yacoub E, Ghose GM. 2014. Featural and temporal attention selectively enhance task-appropriate representations in human primary visual cortex. *Nat. Commun.* 5:5643
Watkins DW, Berkley MA. 1974. The orientation selectivity of single neurons in cat striate cortex. *Exp. Brain Res.* 19(4):433–46
Watson AB, Ahumada AJ. 1985. Model of human visual-motion sensing. *J. Opt. Soc. Am.* 2(2):322–41
Weiner KS, Sayres R, Vinberg J, Grill-Spector K. 2010. fMRI-adaptation and category selectivity in human ventral temporal cortex: regional differences across time scales. *J. Neurophysiol.* 103(6):3349–65
Whitney D, Westwood DA, Goodale MA. 2003. The influence of visual motion on fast reaching movements to a stationary object. *Nature* 423(6942):869–73
Winawer J, Horiguchi H, Sayres RA, Amano K, Wandell BA. 2010. Mapping hV4 and ventral occipital cortex: the venous eclipse. *J. Vis.* 10(5):1
Winston JS, Henson RNA, Fine-Goulden MR, Dolan RJ. 2004. fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *J. Neurophysiol.* 92(3):1830–39
Yacoub E, Harel N, Ugurbil K. 2008. High-field fMRI unveils orientation columns in humans. *PNAS* 105(30):10607–12
Yacoub E, Shmuel A, Logothetis N, Ugurbil K. 2007. Robust detection of ocular dominance columns in humans using Hahn Spin Echo BOLD functional MRI at 7 Tesla. *NeuroImage* 37(4):1161–77
Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ. 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *PNAS* 111(23):8619–24
Yu Q, Shim WM. 2017. Occipital, parietal, and frontal cortices selectively maintain task-relevant features of multi-feature objects in visual working memory. *NeuroImage* 157:97–107
Yu Y, Huber L, Yang J, Jangraw DC, Handwerker DA, et al. 2019. Layer-specific activation of sensory input and predictive feedback in the human primary somatosensory cortex. *Sci. Adv.* 5(5):eaav9053