Main Manuscript for

NeuroModulation Modeling (NMM): Inferring the form of neuromodulation from fMRI tuning functions

Patrick Sadil*, David E. Huber[1], Rosemary A. Cowell[1]

University of Massachusetts, Amherst

[1]Authors DEH and RAC contributed equally to this work.

***Corresponding Author:** Patrick Sadil

Email: psadil@gmail.com

Competing Interest Statement: The authors have no relevant financial or non-financial interests to disclose.

Classification: Biological Sciences/Neuroscience; Social Sciences/Psychological and Cognitive Sciences

Keywords: fMRI; Encoding Model; Neuromodulation; Hierarchical Bayes

This PDF file includes:

Main Text
Figures 1 to 5

Abstract

Many cognitive neuroscience theories assume that changes in behavior arise from changes in the tuning properties of neurons (e.g., Dosher & Lu 1998, Ling, Liu, & Carrasco 2009). However, direct tests of these theories with electrophysiology are rarely feasible with humans. Non-invasive functional magnetic resonance imaging (fMRI) produces voxel tuning, but each voxel aggregates hundreds of thousands of neurons, and voxel tuning modulation is a complex mixture of the underlying neural responses. We developed a pair of statistical tools to address this problem, which we refer to as NeuroModulation Modeling (NMM). NMM advances fMRI analysis methods, inferring the response of neural subpopulations by leveraging modulations at the voxel-level to differentiate between different forms of neuromodulation. One tool uses hierarchical Bayesian modeling and model comparison while the other tool uses a non-parametric slope analysis. We tested the validity of NMM by applying it to fMRI data collected from participants viewing orientation stimuli at high- and low-contrast, which is known from electrophysiology to cause multiplicative scaling of neural tuning (e.g., Sclar & Freeman 1982). In seeming contradiction to ground truth, increasing contrast appeared to cause an additive shift in orientation tuning of voxel-level fMRI data. However, NMM indicated multiplicative gain rather than an additive shift, in line with single-cell electrophysiology. Beyond orientation, this approach could be applied to determine the form of neuromodulation in any fMRI experiment, provided that the experiment tests multiple points along a stimulus dimension to which neurons are tuned (e.g., direction of motion, isoluminant hue, pitch, etc.).

Significance Statement

The spatial resolution afforded by noninvasive neuroimaging in humans continues to improve, but the best available resolution is insufficient for testing theories in cognitive neuroscience; many theories are specified at the level of individual neurons, but magnetic resonance imaging aggregates over hundreds of thousands of neurons. With limited resolution, it is unclear how to test assumptions and predictions of these theories in humans. To bridge this gap, we developed a modeling framework that allows researchers to infer a key property of the neural code -- how stimulus features and cognitive states modulate neural tuning – given only noninvasive neuroimaging data. The framework is broadly applicable to constrain and test theories that link changes in behavior to changes in neural tuning.

Introduction

Neurons in mammalian sensory cortex often exhibit tuning functions, responding selectively to a narrow range of values of a stimulus feature such as line orientation. Moreover, experimental manipulations of perceptual or cognitive state can modulate these tuning functions [e.g., 1–3], providing crucial insights into the neural mechanisms of perception and cognition. For example, spatial attention achieves performance gains by multiplicative rescaling of orientation tuning functions in Macaque V4 [4], rather than through a bias shift (additive shift) or increased selectivity (sharpening). As another example, Figure 1A shows a neuron in mouse primary visual cortex, tuned to motion direction, that appears to undergo multiplicative rescaling as spatial frequency is manipulated (Figure 1A). However, single-cell electrophysiology is rarely feasible in humans, and non-invasive techniques such as functional magnetic resonance imaging (fMRI) reflect the activity of many cells: a patch of striate cortex the size of a typical fMRI voxel (2 mm^3) reflects the activity of approximately 300,000 – 500,000 neurons [5, 6]. Nonetheless,

fMRI studies reveal feature-selective tuning in voxels [7–10], which is assumed to arise from a non-uniform distribution of tuning preferences across the neurons contributing to a voxel.

Like neural tuning functions, voxel tuning functions are modulated by manipulations of perceptual state [10–14], and it is tempting to infer the form of *neural-level* modulation (hereafter, neuromodulation) directly from observed changes in voxel tuning. But, as many have acknowledged, directly inferring neural behavior from voxel behavior might be misleading [10, 15–17]. The relationship between voxel and neural tuning functions presents a many-to-one inverse problem: the same voxel tuning function can arise from many different combinations of the shape of the underlying neural tuning functions and the distribution of neurons across different preferred stimulus values (e.g., the number of neurons preferring vertical, horizontal or oblique orientations). Thus, fMRI tuning does not uniquely specify tuning at the neural level [17, c.f., 18, 16].

We developed two complementary techniques to address this inverse problem, drawing qualitative conclusions about neural-level response properties from the Blood Oxygenation-Level Dependent (BOLD) signal [19, 20]. The first uses Bayesian estimation and model comparison to estimate the form of neuromodulation underlying a change in BOLD across two experimental conditions. The second is a non-parametric check that complements the model comparison by visualizing the form of neuromodulation. Collectively, we term these approaches NeuroModulation Modeling (NMM). Rather than characterizing neural tuning quantitatively, NMM adjudicates between qualitatively different forms of neuromodulation, determining which provides the best account of the data. Like existing methods for uncovering tuning [21–23], NMM assumes that voxel tuning emerges from the combined activity of feature-tuned neurons (Figure 1B). With this assumption, we show how different kinds of neuromodulation (Figure 1C-D) can produce distinct patterns of BOLD data. In short, even though the voxel tuning function reflects the aggregated activity of hundreds of thousands of neurons, NMM leverages distinct effects of different forms of neuromodulation on changes to the BOLD data to identify the form of neuromodulation underlying those changes.

We tested NMM by applying it to a new dataset for which the "ground truth" neuromodulation was known from single-cell recordings: changes in visual contrast induce multiplicative scaling in orientation-tuned neurons [24–26]. Given BOLD data collected from the primary visual cortex of human participants viewing oriented gratings at high versus low contrast, the technique correctly implicated multiplicative scaling, despite working with only voxel-level data.

Linking Neural Tuning to Voxel Tuning

Similar to forward encoding models [21–23], NMM assumes that voxel tuning arises from a linear combination of underlying neural tuning functions (Figure 1B). Unlike previous approaches, which assume a discrete number of neural tuning functions, NMM includes an infinite number of neural tuning functions by specifying the contribution of these tuning function to a voxel's activity with a continuous distribution − the von Mises distribution (Figure 2). The von Mises distribution resembles a Gaussian distribution but is periodic[1], thereby accommodating the circular nature of orientations. For an orientation, r, the density function of the von Mises distribution is

[1] The von Mises density function has a period of 2π, and orientations have a period of π; a grating that is oriented to the upper vertical meridian is the same as a grating oriented to the lower vertical meridian. To account for this difference, orientations are doubled during model fitting.

$$f(r|\phi,\kappa_1) = \frac{\exp(\kappa_1 \cos(r-\phi))}{2\pi I_0(\kappa_1)} \quad (1)$$

where κ_1 is a concentration parameter, ϕ is the orientation of the peak value of the function, and $I_0(\cdot)$ is the modified Bessel function of the first kind, of order 0. The numerator of Equation 1 produces a periodic bell shape peaking at ϕ, with the concentration parameter determining its "sharpness", and the denominator keeping the integral of the function at one. Equation 1 describes the relative contribution to a voxel's activity of all neurons, and so we refer to it as a "weight distribution". With a flat weight distribution (low κ_1), the neurons within a voxel are in equal proportion, resulting in a complete absence of voxel tuning even if the contributing neurons are well-tuned. With a peaked weight distribution (high κ_1), the voxel samples mostly from neurons tuned to orientations around ϕ.

An equation similar to Equation 1 was used to specify neural tuning functions (Equation 2). Sharpness of tuning can be captured directly through the concentration parameter, but for the neural tuning function we included parameters to control the baseline activity and responsiveness. The parameter for baseline activity models the activity of a neuron in the absence of stimulation and is implemented as an additive offset, α. The parameter for responsiveness models how well a neuron responds to stimulation and is implemented as a gain parameter, γ. Combining baseline activity and responsiveness gives the neural tuning function, $NTF(\cdot)$, for a neuron expressed as a function of orientation relative to that neuron's preferred orientation (Equation 2).

$$NTF(r|\kappa_2,\gamma,\alpha) = \alpha + \gamma f(r|0,\kappa_2) \quad (2)$$

Within each voxel (but not across voxels), all neural tuning functions assume the same shape. The overall voxel tuning function is obtained by convolving Equations 1 and 2 (Figure 1B), which is well approximated by a function that has the same form as Equation 2 [27]. This results in Equation 3, the voxel tuning function, $VTF(\cdot)$, in a baseline condition (e.g., low-contrast stimuli).

$$VTF_{baseline}(r|\kappa_v,\gamma_v,\alpha_v,\phi_v) = \alpha_v + \gamma_v f(r|\phi_v,\kappa_v) \quad (3)$$

The subscript v indicates parameters that are estimated separately for each voxel. The concentration parameter, κ_v, is equal to neither κ_1 nor κ_2, but is a function of both [27]. This means that the model does not distinguish between voxels containing a concentrated distribution of orientation-tuned neurons each of which has diffuse tuning (high κ_1 and low κ_2), and voxels containing a diffuse distribution of orientation-tuned neurons each with concentrated tuning (low κ_1 and high κ_2). Specifying the concentrations of the weight distribution and neural tuning functions separately was not needed for determining which form of neuromodulation best explains the data, of the forms that we considered.

Models of Neuromodulation

Within the NMM framework, each form of neuromodulation is instantiated by incorporating additional parameters that modulate one or more parameters in the baseline voxel tuning function (Equation 3, Figure 1C-D). To test the framework's validity, we considered two forms of neuromodulation: multiplicative gain and additive shift. We included multiplicative

gain because it is known to be "ground truth", that is, changes in visual contrast induce multiplicative gain in single-neuron tuning functions [24–26]. To model multiplicative modulation, we included an extra, voxel-specific gain parameter, g_v (Equation 4a). In seeming contradiction to multiplicative modulation, an examination of raw voxel tuning functions suggests an additive shift with increases in visual contrast (Figure 3). Therefore, for an alternative model against which to compare multiplicative modulation, we selected additive modulation[2]. To model additive modulation, we included another, voxel-specific additive shift parameter, a_v (Equation 4b). In summary, Equation 3 captures the situation for low contrast in both models, whereas either Equation 4a or 4b captures the situation for high contrast.

$$Multiplicative: \quad VTF_{modulated}(r|\kappa_v, \gamma_v, \alpha_v, \phi_v, g_v) = a_v + g_v \gamma_v f(r|\phi_v, \kappa_v) \qquad (4a)$$
$$Additive: \quad VTF_{modulated}(r|\kappa_v, \gamma_v, \alpha_v, \phi_v, a_v) = a_v + \alpha_v + \gamma_v f(r|\phi_v, \kappa_v) \qquad (4b)$$

Equations 3, 4a, and 4b model voxel activity at each tested orientation, at each level of contrast. However, there is variability in voxel activation across runs, and this variability varies across voxels (some voxels are noisier than others, Figure S2). These sources of variability were included in the model, and all voxel-specific parameters were estimated hierarchically (SI Appendix, Figure S1).

Model Recovery

Given multiple forms of neuromodulation (e.g., multiplicative, additive), each instantiated with a different model (e.g., Equation 4a, 4b), NMM judges the relative plausibility of these forms with model comparison on empirical data. To assess the extent to which inherent properties of the models (e.g., their flexibility) allow reliable model comparison, we assessed model recovery via two kinds of simulation: data-informed and data-uninformed [28, 29].

Data-uninformed model recovery asks whether any model is overly flexible. First, synthetic data were generated with each model, using "weakly informative" priors for how real-world empirical data would be distributed (Figure S1). Next, the models were fit to these synthetic data, keeping in mind the "ground truth", data-generating model. A model would be considered overly flexible if it tended to fit the synthetic data best, even when the data were generated from a different model. If a model were overly flexible, an outcome in which the overly flexible model provides the best fit to the empirical data could not be taken at face value – this outcome may have arisen owing to the winner's excessive flexibility, rather than because it is more likely to be the correct model.

Data-informed model recovery repeats the recovery process using what has been learned about the models from the empirical data. Whereas data-uninformed model recovery assesses the fairness of model comparison across a wide range of datasets that are plausible given the prior distribution, data-informed recovery uses the posterior distribution. The two kinds of recovery simulation may differ if, for example, the empirical data suggest that the effect of contrast is small enough to be readily buried by noise.

---

[2] Other forms of neuromodulation could also be modeled. For example, a shift in orientation preference could be modeled by allowing ϕ_v to change across conditions, and a change in concentration could be modeled by allowing κ_v to change, both of which assume the weight distribution is unaffected by the contrast manipulation. However, both of these forms of neuromodulation were rejected for the current dataset because neither of them allows contrast to alter the average activity of a voxel across orientations – an effect that was clearly present in the empirical data (Figure 3).

Non-Parametric Check

In the Bayesian component of NMM, the hierarchical nature of the voxel tuning functions (Equations 3, 4a and 4b) provides a powerful way to combine information across voxels, allowing model comparison to automatically down-weight the noisier voxels and magnify the well-tuned voxels [30]. But the Bayesian model assumes that the weight distribution and neural tuning function follow specific functional forms, assumptions that may not hold in all datasets (e.g., with some voxel sizes, the weight distribution may be multimodal). Thus, to complement the parametric modeling, NMM includes a non-parametric method for visualizing the form of neuromodulation. To preview, regardless of the shape of the weight distribution, and regardless of the shape of the neural tuning function, if the neuromodulation is multiplicative, a within-voxel plot comparing high- versus low-contrast should have a slope greater than 1, whereas if the neuromodulation is additive, this plot should reveal a slope equal to 1.

The key difference between the models we compared is that the multiplicative model inherently allows the effect of contrast to vary by orientation whereas the additive model does not. Additive neuromodulation corresponds to an increase in neural activity at all orientations. Thus, regardless of the forms of the weight distribution and neural tuning function, additive neuromodulation causes the low-contrast tuning function to shift upwards uniformly across orientations (Figure 4, top left). Hence, a scatterplot of the voxel's response to high-contrast stimuli against its response to low-contrast stimuli has a slope of 1 (Figure 4, top right). In contrast, multiplicative neuromodulation corresponds to a greater increase in neural activity at the most preferred orientations (Figure 4, bottom left), and a scatterplot of a voxel's response to high versus low-contrast stimuli has a slope larger than 1 (Figure 4, bottom right). Therefore, the models can be differentiated by plotting high-contrast activity against low-contrast activity and calculating the slope of the best fitting line (Figure 4, right). A slope of one implies additive shift, but a slope greater than one implies multiplicative gain (see also Supplementary Information).

However, the application of this slope analysis to empirical data is not necessarily straightforward. First, the slope between high- and low-contrast cannot be determined using standard linear regression (i.e., estimating the line that minimizes the ordinary least squares error, the squared vertical distance between the line and the data). Standard linear regression estimates a relation between two sets of observations, such as the activity at low and high contrast. But when *both* of those observations are corrupted by noise, the estimated relationship is "diluted", biased to zero [31]. To address this dilution, the slope was estimated using *orthogonal* regression, by finding the line that minimizes the squared vertical and horizontal distances from the data [Supplementary Information, Methods, 31]. Second, the slope analysis cannot be done across voxels because voxels with a larger response in the low-contrast condition are likely to produce a larger contrast effect, and this tendency will bias an across-voxel analysis toward a slope greater than 1 (Figure 5A). To address this bias, slopes were determined separately for every voxel, followed by an analysis of the distribution of slopes. Third, even a within-voxel slope analysis may be problematic because most voxels are only weakly tuned, responding to all orientations almost equivalently. For voxels that fail to respond differently to different orientations, the data for the slope analysis will be an uncorrelated cloud of points, and in orthogonal regression the slope for a cloud of points can take on nearly any value. As seen in Figure 5B, when all voxels were included, the range of observed slope values was considerable, with the tails of the slope distribution extending from -100 to +300 (note the different scale of the x-axes in the left and right plots of Figure 5B). Given this variability, we first examined the

distribution of slopes for all voxels (Figure 5B, left), and also repeated this procedure using a subset of the voxels that were most sensitive to contrast (Figure 5B, right), voxels whose average activity (collapsed across orientations and runs) was most influenced by the contrast manipulation. That is, we analyzed either all voxels (Figure 5B, left), or only those voxels whose difference in average activity across levels of contrast exceeded the 90% quantile of differences (Figure 5B, right). The threshold was applied within participants, such that each participant contributed their 10% most active voxels.

Results

Behavior

Average accuracy for the spatial frequency change detection task was 72 % and 74 % for low and high-contrast gratings, respectively ($p = 0.13$). Eye-tracking was used to assess whether participants maintained adequate fixation in the scanner: over 90 % of all participants' fixations ended within 2° of the location of the run's average fixation (range: 93–99 %).

Model Comparison through Cross-Validation

Data-Uninformed Model Recovery. To evaluate the ability of the NMM Bayesian modeling procedure to recover both kinds of modulation, we conducted model recovery by simulating datasets from the prior distribution, fitting two models that allowed different kinds of modulation, and comparing the resulting fits using an approximation to cross-validation [29, SI Appendix, 32]. The models did not exhibit mimicry: model comparison picked the true data-generating model in every simulated dataset, providing strong evidence that the technique recovers the appropriate class of modulation when it is present in the data.

Application to real data. We next asked whether NMM recovers from the fMRI data the correct form of neuromodulation. Specifically, it is known from electrophysiology that the tuning functions of orientation-sensitive neurons undergo multiplicative neuromodulation with an increase in stimulus contrast [24–26]. Given this knowledge, model comparison on the collected BOLD data ought to favor the multiplicative model. Models were compared using the same approximation to cross-validation as for model recovery [32]. The multiplicative model provided the best fit to the data: a difference of 14.5 standard errors, in units of expected log pointwise-predictive density [32].

Data-Informed Model Recovery. Exploiting what was learned about the plausible parameter values by observing the data, we further probed the validity of model comparison through data-informed model recovery [29, 32, SI Appendix]. Synthetic datasets were generated from the posterior distributions of the multiplicative and additive models. Next, the multiplicative and additive models were refit to each of these simulated datasets. All datasets were fit best by the model whose posterior was used to generate the data.

Non-Parametric Check

The Bayesian analysis distinguishes additive versus multiplicative neuromodulation, but it requires assuming a specific neural tuning function and weight distribution. Even without those assumptions, these two forms of neuromodulation should be distinguishable based on what these forms predict for the slope of a plot of high-contrast activity against low-contrast activity (Figure 4, Introduction, Supplementary Information). The slope predicted by multiplicative neuromodulation is present in the empirical data. Figure 5B shows the distribution of within-voxel slopes, as estimated with orthogonal regression in both the high- and low-contrast

directions (Introduction, Methods). Considering either all voxels (Threshold 0) or only the most contrast-sensitive voxels (Threshold 0.9), the median slope is greater than one. This tendency for slopes to be larger than 1 was supported by an additional Bayesian hierarchical model of the slopes (Supplementary Information), and is predicted by multiplicative but not additive neuromodulation.

Discussion

Many theories of cognition predict specific kinds of neural tuning changes with different manipulations [e.g., 33, 34]. For instance, attention researchers do not only consider neural responses when people are in a state of high attention; this state is compared and contrasted to one of low attention, and theories of attention are informed by this comparison [e.g., 35, 36]. However, these predictions for neuromodulation – the manner in which neural tuning functions change – are rarely tested in humans because there are no non-invasive techniques for recording from individual neurons. Tuning functions, and modulation of tuning functions, can be measured non-invasively using fMRI, but the response of a voxel reflects hundreds of thousands of neurons. Thus, voxel tuning functions have been interpreted as population-level properties of stimulus representations in the brain, unreflective of the response properties of the underlying, individual neurons [10, 15–17].

Rather than attempting a precise measurement of neural-level tuning from fMRI data, we developed a set of procedures, NMM, that identifies the functional form of neuromodulation. NMM comprises two techniques: 1) hierarchical Bayesian modeling of BOLD data combined with model comparison; and 2) a qualitative, non-parametric check. The key principle underlying the NMM framework, used in both techniques, is a modeled link between the neural and voxel tuning functions (Figure 1B-D). This link justifies using the techniques to infer the form of neuromodulation given data from the voxel-level only. Even though the BOLD signal reflects the aggregated activity of many neurons, different forms of neuromodulation can produce identifiable signatures in voxels' BOLD activity (i.e., Equation 4a vs 4b, Figure 4), and so we can infer the form of neuromodulation by comparing models fit to the BOLD data. We tested, with model recovery, the reliability of those signatures, which revealed that model comparison would not conflate multiplicative gain with additive offset. Then, we applied NMM to a case where the neural-level "ground truth" modulation was known to be multiplicative gain from electrophysiology [24–26], and both techniques of the NMM framework correctly implicated multiplicative gain using only the BOLD signal.

The NMM framework is an important advance in fMRI analysis techniques because it identifies effects at the sub-voxel, neural level, in contrast to previous techniques that assume the voxel as the base unit. Long-established region-of-interest analysis methods average the BOLD signal across swaths of cortex to reach conclusions regarding the function of each region, and by doing so, they essentially assume that all voxels within the region behave in a similar manner [e.g., 37]. Decoding methods retain fine-grained spatial information by probing representations that may be distributed across voxels, representations that are not necessarily parceled into gross anatomical regions [38, 39]. Statistical parametric mapping improves upon region-level analyses by probing information contained in individual voxels and clusters of voxels [40, 41]. However, this progression of increasing spatial resolution nonetheless treats the voxel as the base unit of observation – a unit that is constrained by imaging technology rather than neurocognitive theories.

Several techniques have been developed to peer inside the voxel, but none of them statistically compares alternative theories of neuromodulation using a model that explicitly links neural tuning with voxel tuning. For example, biophysical models specify how the BOLD signal arises from the often nonlinear coupling between neural activity and vasculature [42, 43]. But, despite this biological realism, the neural activity has not yet been translated into neural tuning functions. In contrast to biophysical models, encoding models of fMRI data delve inside the voxel by assuming that voxel activity arises from the aggregated activity of a set of sub-voxel "channels" [7, 21–23], and, promisingly, enable researchers to use the BOLD signal to uncover features of neural tuning [18]. Encoding models can be applied to multiple conditions – for example through simulation [44], or through fitting to data and inverting [45] – thereby estimating how the channels are modulated by an experimental manipulation. Inverting an encoding model uses it as a decoder, estimating channel responses for other stimuli [21, 22, 46] or estimating how channels are modulated in other conditions [47]. But this use of inverted encoding models requires hard-wired (and arbitrary) assumptions about the state of the channel responses in the baseline condition. These assumptions mean the inferred channel modulation cannot be taken to reflect the modulations in underlying neural functions that would be measured with electrophysiology, as users of such models readily acknowledge [17, 48]. As a result, inverted encoding models have been used to pursue very different goals (i.e., understanding population-level representations) than the aim of the present study [17]. Beyond inverted encoding models, other researchers have compared the qualitative predictions of different forms of neuromodulation [44, 49, 50], analogous to the logic underlying NMM's non-parametric check. However, these techniques neither fit the alternative encoding models to empirical data, nor take into account model flexibility.

These previous attempts to peer inside the voxels are combined in NMM. Like biophysical models, NMM connects the latent neural activity to the observed BOLD, by assuming a linear coupling that is justified in our experimental design [51–55]. But unlike extant biophysical models, NMM uses an encoding model (Figure 1B-D, Equations 3, 4a, 4b) and characterizes modulations in tuning across different conditions. Unlike inverted encoding models, it avoids hard-wired assumptions about channels; it does so by attempting to identify the form of neuromodulation directly, bypassing any attempt to identify the exact shape of the underlying neural tuning functions. NMM leverages the qualitative differences between different forms of neuromodulation as a non-parametric check, but its main inference is based on a quantitative measure of the predictive abilities of each model, a measure that is sensitive to both the flexibility of each model and the signal to noise ratio of the data.

NMM is able to specify the form of neuromodulation because it not only acknowledges that different voxels have different stimulus preferences, but it goes further, by acknowledging and exploiting the fact that different voxels have different characteristic tuning functions. To understand this point, consider a voxel tuning function produced by a simple "averaging" across aligned voxels, which is an appropriate technique if all voxels have similar tuning functions (Figure 3). This "binning" alignment technique appears to reveal an additive shift with contrast, but this additive shift is caused by the inclusion of many voxels with poor orientation tuning, i.e., with an almost uniform distribution of neural tuning functions across preferred stimulus values. When the BOLD response is aggregated across a great many such "undiagnostic" voxels, the true form of neuromodulation is obscured (i.e., it appears additive rather than multiplicative). Happily, however, there is variability among voxels: not all voxels are poorly tuned. The two

techniques of NMM capitalize on this variation by detecting the well-tuned "diagnostic" voxels, and using them to identify the form of neuromodulation.

As is likely to be the case in any fMRI study of voxel tuning, the majority of the voxels in our study were poorly tuned (e.g., a large number of voxels near the origin of Figure 5A). Although we restricted the analyses to only those voxels whose population receptive fields overlapped with the stimulus (see Methods and Supplementary Information), the restriction was likely imperfect: some voxels were only weakly responsive to visual stimulation and, among those that were responsive, few were strongly affected by contrast. These voxel differences and other sources of voxel variation – including partial voluming [e.g., 56] , proximity of each voxel to blood vessels [39], accuracy of the assumed hemodynamic response [57], the responsiveness of neurons within each voxel, etc. – are present in any fMRI experiment. But in the presence of such variation, analyses that assume the same effect for all voxels may be misleading. In the current case, because most voxels exhibited nearly flat tuning, this altered the appearance of the average voxel tuning function (Figure 3), incorrectly suggesting that changing visual contrast causes an additive shift.

NMM guards against poor tuning by focusing on only the divergent predictions of the considered forms of neuromodulation. Because NMM explicitly addresses voxel variation, the predictions for the different forms of neuromodulation are assessed within each voxel (e.g., a voxel with poor tuning is predicted to produce what appears to be an additive shift even with multiplicative gain). All voxels help constrain the parameters of the Bayesian models. But voxels with flat tuning or a weak contrast effect are equally well accounted for by any form of neuromodulation, and so such voxels cannot adjudicate between forms. Only voxels with relatively sharp tuning and a sizable contrast effect can be uniquely matched to one form of neuromodulation. The hierarchical structure of the Bayesian model (and the Bayesian version of the non-parametric check, SI) allows the model comparison to automatically rely most heavily on the voxels that are most diagnostic for determining the form of neuromodulation (i.e., the ones with sharp tuning and a sizable contrast effect). Thus, rather than treating variation between voxels as a source of noise, NMM capitalizes on this variation to efficiently identify the form of neuromodulation.

Limitations and Extensions

The model comparison technique of the NMM framework approximates the distribution of neurons tuned to orientations within each voxel with a von Mises distribution, which is unimodal. This approximation provided a convenient formula for deriving the voxel tuning function, and the resulting function matched the data (e.g., Figure 3 looks unimodal). However, this assumption might not hold for other types of stimuli, other brain regions, or for other scan parameters. For example, presenting more orientations may reveal that some voxels have multimodal tuning functions. Incorporating more complex weight distributions, including multimodal functions, will require further research (e.g., it could be assumed that the weight distribution is a mixture of two Von Mises distributions with separate preferred orientations). Regardless, the non-parametric visualization check does not rely on this unimodal assumption, and it yielded the same result in our validation study.

Another limitation of NMM is that it can differentiate only between models that are formally part of the model comparison process. If the "true" form of neuromodulation is not included in the set of possibilities, the results might be misleading. Neuromodulation may be complex in some situations, involving a combination of factors [58–60]; for example, following extended adaptation, orientation tuning functions can both widen and shift their preferred

orientation [58]. In principle, the model comparison technique of NMM could handle these additional modulations by allowing more parameters to vary between each model, because both the width and preferred direction map on to individual parameters. But in applying NMM to explore neuromodulations beyond multiplicative gain and additive shift, further model recovery simulations would be required.

Like many fMRI analysis techniques, NMM makes several simplifying assumptions. For instance, it assumes that voxel tuning functions are a linear combination of the activity of the underlying neural tuning functions. Although this may be a reasonable assumption in some experiments [51–53], the relationship between neural firing rate and BOLD signal is not perfectly linear [61]. NMM also simplifies the situation by assuming that all neurons within each voxel have tuning functions that share the same shape (though it allows these functions to differ for different voxels). Considering the relatively wide tuning of inhibitory interneurons [62], this assumption is almost certainly incorrect. These simplifying assumptions are not strictly necessary, and future research could incorporate greater biological specificity. For example, multiple mathematical accounts have been proposed to explain the nonlinearities between the neural firing rate and the BOLD signal [61, 63–65], and these relationships could be built into NMM's equations. At least for this simple test case of visual contrast, even with these simplifying assumptions, NMM reached the correct conclusion.

Summary

Using a simple test case – modulation of orientation tuning by stimulus contrast – we validated the NMM method for deriving the form of neuromodulation from a BOLD dataset. Despite the simplicity of the test case, the method should be applicable to a broad range of domains and manipulations. The main requirement for using the technique is an experimental paradigm that: (1) produces measurable voxel tuning by testing the BOLD response at different levels of a single stimulus dimension (e.g., direction of motion, color, pitch), and (2) includes some cognitive manipulation of interest (e.g., with and without attention, before and after perceptual learning, complex versus simple stimuli) that modulates the voxel tuning from one condition to another. Now that NMM has been validated, it can be used to study modulations of neural tuning where electrophysiology has not yet provided an answer (e.g., studies of neuromodulation underlying tasks that cannot be taught to animals).

Materials and Methods

Participants

Seven participants (22 – 31 years old; 3 female, 2 did not report) completed three sessions for monetary compensation ($50 per 2-hour session). All participants had normal or corrected-to-normal vision. One additional participant completed a single session but exhibited substantial motion; their data were excluded. The procedure was approved by the University of Massachusetts Institutional Review Board.

Behavioral Stimulation and Recording

Stimuli were presented to participants with a gamma corrected 32" LCD monitor at 120 Hz refresh rate (Cambridge Research Systems). The experiment was designed using the Psychophysics toolbox [Version 3.0.14 66, 67] and custom MATLAB code (2018b, MathWorks). Behavioral responses were collected with a button box (Current Design). Eye-tracking data were recorded at a rate of 1000 Hz with the Eyelink 1000 Plus system on a long

range mount (SR Research), controlled using the Eyelink Toolbox extension to Psychtoolbox [68]. Due to a technical error, the button presses of one participant were not recorded.

Orientation scans. For the main experiment, participants completed 18 functional runs across three sessions. During each run, oriented grayscale gratings were presented twice at each of two levels of contrast (eight orientations at 50% or 100% Michelson contrast in all runs for six out of seven participants, eight orientations at 20% or 80% Michelson contrast in 12 runs for one participant, and seven orientations at 20% or 80% Michelson contrast in that participant's remaining 6 runs). Grating parameters replicated those of Rademaker et al. [69]. Gratings (spatial frequency of 2 cycles per degree) were masked with annuli (1.2° inner and 7° outer radii). The annulus edges were smoothed with an isotropic 2D Gaussian kernel (1° kernel, 0.5° standard deviation). Throughout each run, a magenta fixation dot was presented in the center of the screen (0.2°, RGB: 0.7843, 0, 0.8886).

In each trial, a counterphasing (5 Hz) grating was presented for five seconds. In the middle three seconds of each trial, the spatial frequency of the grating either increased or decreased (1 cycle per degree) for 200 ms. Participants were instructed to indicate via button press the direction of change as soon as they noticed it. Per run, gratings were presented at multiple orientations, twice at each combination of orientation and contrast. In most runs, there were eight orientations, but in one session of one participant (totaling six runs), only seven orientations were presented. A five-second fixation period preceded the first trial, and a fifteen-second fixation period succeeded the final trial. Each run lasted 490 seconds.

Population Receptive Field (pRF) Mapping Scans. We used pRF mapping to limit analyses to voxels that responded to the grating stimulus, but not to the edges of the stimulus nor to peripheral regions outside the stimulus. pRF mapping scans followed the protocol of Benson et al. [70]. Briefly, natural images [from 71] were overlaid on pink noise and viewed through a series of apertures. Within one run per session, the apertures enabled view of either a moving bar or rotating wedges and rings that expanded and contracted. In bar runs, a bar traversed the central region in cycles. During each cycle, the bar was visible for 28 seconds, followed by a 4-second blank period. The bar moved in one of eight directions (east, north, west, south, northeast, northwest, southwest, or southeast, in that order). A 16-second blank period preceded the first cycle, a 12-second blank period followed the fourth cycle, and there was a 16-second blank period at the end of all cycles (300 seconds in total).

In the second pRF scan, the apertures were either wedges that rotated clockwise or counterclockwise, or they were rings that expanded or contracted. These runs started with a 16-second blank period, followed by two, 32-second cycles of a counterclockwise rotating wedge, two 28-seconds of expanding rings (each followed by a 4-second blank period), two 32-second clockwise wedge rotations, and two 28-second cycles of contracting rings (followed by 4 and 26 seconds of blank, respectively). The total run time was 300 seconds.

Throughout the pRF scans, the color of a central fixation dot (0.2°) changed between black, white, and red. Participants were instructed to monitor the color of the fixation dot and press a button when the dot turned red. To help participants maintain fixation, a circular fixation grid was presented throughout.

fMRI Data Acquisition

MRI data were collected on a 3 T Siemens Skyra scanner with a 64-channel head coil. In each of the three sessions we collected field-mapping scans, functional scans, and a T1-weighted anatomical scan (MPRAGE, FOV 256 × 256, 1 mm isotropic, TE 2.13 ms, Flip Angle 9°). The anatomical scan was used to align field-mapping and functional images parallel to the calcarine

sulcus. Gradient recall echo scans estimated the magnetic field. The pRF and primary functional data were collected with the same scan parameters (TR 1000 ms, TE 31 ms, flip angle 64°, FOV 94 × 94, 2.2 mm isotropic, 2.2 mm slice gap, Multiband Acceleration Factor 4). To aid alignment of functional and anatomical images, single-band reference images were collected before each functional run for all but three participants (TR 8000 ms, TE 65.4 ms, flip angle 90°, FOV 94 × 94, 2.2 mm isotropic, 2.2 mm slice gap).

Preprocessing of images was performed with *fMRIPrep* 1.4.0 [72, 73], which relies on *Nipype* 1.2.0 [74, 75] and *Nilearn* 0.5.2 [RRID:SCR_001362, 76], see SI Appendix.

Population Receptive Field Mapping

To mitigate the effects of stimulus vignetting – whereby orientation information differs between the edge and central portions of a grating [77, 78, cf 79] – we restricted the analyses to only voxels whose population receptive fields (pRFs) did not overlap with the stimulus edges. We estimated the pRFs of each voxel with standard methods. First, the preprocessed functional data for the pRF scans were converted into percent mean signal change within a run (across voxels). The compressive spatial summation model was fit to each voxel using analyzePRF HCP7TRET [64, osf.io/bw9ec/ 70]. Following Benson et al. (2018), the compressive exponent of this model was set to 0.05. The resulting pRF parameters were combined with an anatomical prior for a Bayesian estimation of the parameters [80 neuropythy 0.94]. Only the parameters of voxels for which the pRF explained more than 10 % of the variance of the run were used as empirical parameters for Bayesian estimation; the remaining voxels' posterior pRF parameters were determined entirely by the prior.

The resulting pRF parameters determined whether a voxel would be retained for analyses. The pRF resembles an isotropic, bivariate Gaussian. The three pRF sessions were analyzed separately, resulting in three sets of three pRF parameters per voxel. Within a set of parameters, two indicate the center location of the pRF, and the third determines its size – the standard deviation of the Gaussian. A voxel was retained only if a circle centered on its pRF with radius equal to two standard deviations was entirely contained by the grating stimulus in all three sessions.

Estimated Voxel Tuning Function

The voxel tuning function (Figure 3) was estimated according to a "binning" method [e.g., 10]. First, each voxel's activity was z-scored, separately for each voxel. Then, one run was held out and the voxel's preferred orientation was calculated with the remaining runs. The preferred orientation was estimated as the orientation that produced the largest activity, averaged across runs and levels of contrast. Next, the activity in the held-out run was plotted against stimulus orientation translated into a function of distance from the voxel's preferred orientation. This procedure was repeated once for each available run, holding out a different run on each repetition. Averages and repeated-measures confidence intervals, across voxels and ignoring participants, were then calculated [81], treating contrast and distance from preferred orientation as within-voxel factors.

In Figure 3, one session of one participant was excluded; the figure shows only sessions in which the participant was shown eight orientations.

Neural Tuning Function Model

Each of the voxel-specific parameters (controlling the tuning function, modulations to those tuning functions, and variability in the output of the tuning function) were estimated with a hierarchical Bayesian model. All models were estimated with R (R Core Team) and its interface

(RStan, 2.18.2) to the Stan language [82, 83]. The Stan language provides a robust variation of Hamiltonian Monte Carlo [84, 85], which efficiently and accurately approximates Bayesian posterior distributions [86]. Two diagnostics were used to assess the accuracy of the estimation. First, we calculated the split-$\hat{R}$ for each parameter in the model [87]. This value is analogous to an F-score in an analysis of variance, and values close to 1 imply that the different Markov chains are sampling from the same distribution. The split-$\hat{R}$ was below 1.1 for all parameters [30]. Second, chains were monitored for divergences, whose presence would indicate that the samples, even if they are from a common distribution, likely misrepresent the true posterior distribution [86]. There were no divergences in the fits of the multiplicative and additive models to real data.

In all fits, four chains were initialized with random values for the parameters (RStan defaults), and the sampling algorithm was given 1000 draws to adapt. After adaptation, each chain was used to draw 2000 samples from the posterior distribution. All model comparison was done with Pareto-smooth importance sampling, leave-one-out cross-validation plus [32].

Model Recovery

For data-uninformed model recovery, 100 datasets were simulated from the priors of both models. The datasets contained one participant contributing 100 voxels measured in 18 runs at 8 levels of orientation. Both the additive and multiplicative models were fit to each simulated dataset, resulting in 200 simulated datasets and 400 model fits.

Data informed model recovery was implemented analogously to the data-uninformed model recovery. One hundred datasets were generated from the posterior of the multiplicative and additive models. The datasets contained six participants, each contributing 200 voxels measured in 18 runs at 8 levels of orientation. The multiplicative and additive models were fit to these simulated datasets and model comparison was conducted as above.

Orthogonal Regression Estimate of Slope

A derivation of the slope for orthogonal regression can be found in Casella and Berger, chapter 12. Here, we report the result. Let x_i be the activity of a voxel at low contrast on trial i, out of n trials; let y_i be the activity of a voxel at high contrast on an analogous trial (i.e., the beta weight for the same orientation in the same run); let $\bar{x}$ be the voxel's average activity at low-contrast; and let $\bar{y}$ be the voxel's average activity at high-contrast. The following are the sums of squares and the cross-product.

$$S_{xx} = \sum_{i}^{n} (x_i - \bar{x})^2$$

$$S_{yy} = \sum_{i}^{n} (y_i - \bar{y})^2$$

$$S_{xy} = \sum_{i}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

The slope, g, that solves the total least squares problem is given by

$$g = \frac{-(S_{xx} - S_{yy}) + \sqrt{(S_{xx} - S_{yy})^2 + 4S_{xy}^2}}{2S_{xy}} \quad 5$$

As stated in the main text, a separate slope was estimated for each voxel. In one session, one participant was shown seven rather than eight orientations. When calculating the slopes for this participant, the activity from those sessions were excluded.

Acknowledgements

The work was supported by NIH award 1RF1MH114277-01 to R.A.C. and D.E.H.

References

1. Hsiao SS, Lane J, Fitzgerald P (2002) Representation of orientation in the somatosensory system. Behav Brain Res 135:93–103. https://doi.org/10.1016/S0166-4328(02)00160-2

2. Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. J Physiol 148:574–91. https://doi.org/10.1113/jphysiol.1959.sp006308

3. Yang T, Maunsell JHR (2004) The effect of perceptual learning on neuronal responses in monkey visual area V4. J Neurosci 24:1617–1626. https://doi.org/10.1523/JNEUROSCI.4442-03.2004

4. McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. J Neurosci 19:431–41

5. Goense JBM, Logothetis NK (2008) Neurophysiology of the BOLD fMRI signal in awake monkeys. Curr Biol 18:631–640. https://doi.org/10.1016/j.cub.2008.03.054

6. Leuba G, Garey LJ (1989) Comparison of neuronal and glial numerical density in primary and secondary visual cortex of man. Exp Brain Res 77:. https://doi.org/10.1007/BF00250564

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

8. Engel SA, Glover GH, Wandell BA (1997) Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb Cortex 7:181–192. https://doi.org/10.1093/cercor/7.2.181

9. Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679–685. https://doi.org/10.1038/nn1444

10. Serences JT, Saproo S, Scolari M, et al (2009) Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. NeuroImage 44:223–231

11. Saproo S, Serences JT (2010) Spatial attention improves the quality of population codes in human visual cortex. J Neurophysiol 104:885–895. https://doi.org/10.1152/jn.00369.2010

12. Scolari M, Serences JT (2010) Basing perceptual decisions on the most informative sensory neurons. J Neurophysiol 104:2266–2273. https://doi.org/10.1152/jn.00273.2010

13. Serences JT, Saproo S (2010) Population response profiles in early visual cortex are biased in favor of more valuable stimuli. J Neurophysiol 104:76–87. https://doi.org/10.1152/jn.01090.2009

14. Zhang J, Meeson A, Welchman AE, Kourtzi Z (2010) Learning alters the tuning of functional magnetic resonance imaging patterns for visual forms. J Neurosci 30:14127–14133. https://doi.org/10.1523/JNEUROSCI.2204-10.2010

15. Çukur T, Nishimoto S, Huth AG, Gallant JL (2013) Attention during natural vision warps semantic representation across the human brain. Nat Neurosci 16:763–70. https://doi.org/10.1038/nn.3381

16. Nevado A, Young MP, Panzeri S (2004) Functional imaging and neural information coding. NeuroImage 21:1083–1095. https://doi.org/10.1016/j.neuroimage.2003.10.043

17. Sprague TC, Adam KCS, Foster JJ, et al (2018) Inverted Encoding Models Assay Population-Level Stimulus Representations, Not Single-Unit Neural Tuning. Eneuro 5:ENEURO.0098-18.2018. https://doi.org/10.1523/ENEURO.0098-18.2018

18. Keliris GA, Li Q, Papanikolaou A, et al (2019) Estimating average single-neuron visual receptive field sizes by fMRI. Proc Natl Acad Sci 116:201809612. https://doi.org/10.1073/pnas.1809612116

19. Ogawa S, Lee TM, Kay AR, Tank DW (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci 87:9868–9872. https://doi.org/10.1073/pnas.87.24.9868

20. Ogawa S, Tank DW, Menon R, et al (1992) Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci 89:5951–5955. https://doi.org/10.1073/pnas.89.13.5951

21. Brouwer GJ, Heeger DJ (2009) Decoding and Reconstructing Color from Responses in Human Visual Cortex. J Neurosci 29:13992–14003

22. Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352–355. https://doi.org/10.1038/nature06713

23. Mitchell T, Shinkareva S, Carlson A (2008) Predicting human brain activity associated with the meanings of nouns. Science 320:1191–1195. https://doi.org/10.1126/science.1152876

24. Alitto HJ, Usrey WM (2004) Influence of Contrast on Orientation and Temporal Frequency Tuning in Ferret Primary Visual Cortex. J Neurophysiol 91:2797–2808. https://doi.org/10.1152/jn.00943.2003

25. Sclar G, Freeman RD (1982) Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. Exp Brain Res 46:457–461

26. Skottun BC, Bradley A, Sclar G, et al (1987) The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. J Neurophysiol 57:773–786

27. Mardia KV, Jupp PE (1999) Directional statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA

28. Jang Y, Wallsten TS, Huber DE (2012) A stochastic detection and retrieval model for the study of metacognition. Psychol Rev 119:186–200. https://doi.org/10.1037/a0025960

29. Wagenmakers E-J, Ratcliff R, Gomez P, Iverson GJ (2004) Assessing model mimicry using the parametric bootstrap. J Math Psychol 48:28–50. https://doi.org/10.1016/j.jmp.2003.11.004

30. Gelman A, Carlin J, Stern H, et al (2013) Bayesian Data Analysis. Chapman and Hall/CRC

31. Casella G, Berger RL (2001) Statistical inference. Duxbury Press

32. Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput 27:1413–1432. https://doi.org/10.1007/s11222-016-9696-4

33. Dosher BA, Lu Z-L (1998) Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. Proc Natl Acad Sci 95:13988–13993. https://doi.org/10.1073/pnas.95.23.13988

34. Ling S, Liu T, Carrasco M (2009) How spatial and feature-based attention affect the gain and tuning of population responses. Vision Res 49:1194–1204. https://doi.org/10.1016/j.visres.2008.05.025

35. Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cognit Psychol 12:97–136. https://doi.org/10.1016/0010-0285(80)90005-5

36. Schneider W, Shiffrin RM (1977) Controlled and automatic human information processing: I. Detection, search, and attention. Psychol Rev 84:1–66. https://doi.org/10.1037/0033-295X.84.1.1

37. Tootell R, Reppas J, Kwong K, et al (1995) Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J Neurosci 15:3215–3230. https://doi.org/10.1523/JNEUROSCI.15-04-03215.1995

38. Haxby JV, Gobbini MI, Furey ML, et al (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430. https://doi.org/10.1126/science.1063736

39. Kriegeskorte N, Cusack R, Bandettini P (2010) How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatiotemporal filter? NeuroImage 49:1965–1976. https://doi.org/10.1016/j.neuroimage.2009.09.059

40. Friston KJ, Jezzard P, Turner R (1994) Analysis of functional MRI time-series. Hum Brain Mapp 1:153–171. https://doi.org/10.1002/hbm.460010207

41. Friston KJ (1995) Statistical parametric mapping: ontology and current issues. J Cereb Blood Flow Metab 15:361–370

42. Buxton RB, Wong EC, Frank LR (1998) Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. Magn Reson Med 39:855–864. https://doi.org/10.1002/mrm.1910390602

43. Stephan KE, Harrison LM, Penny WD, Friston KJ (2004) Biophysical models of fMRI responses. Curr Opin Neurobiol 14:629–635. https://doi.org/10.1016/j.conb.2004.08.006

44. Alink A, Abdulrahman H, Henson RN (2018) Forward models demonstrate that repetition suppression is best modelled by local neural scaling. Nat Commun 9:. https://doi.org/10.1038/s41467-018-05957-0

45. Brouwer GJ, Heeger DJ (2011) Cross-orientation suppression in human visual cortex. J Neurophysiol 106:2108–2119

46. Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. NeuroImage 56:400–410. https://doi.org/10.1016/j.neuroimage.2010.07.073

47. Sprague TC, Saproo S, Serences JT (2015) Visual attention mitigates information loss in small- and large-scale neural codes. Trends Cogn Sci 19:215–226. https://doi.org/10.1016/j.tics.2015.02.005

48. Gardner JL, Liu T (2019) Inverted encoding models reconstruct an arbitrary model response, not the stimulus. eNeuro 6:ENEURO.0363-18.2019. https://doi.org/10.1523/ENEURO.0363-18.2019

49. Ramirez FM, Cichy RM, Allefeld C, Haynes J-D (2014) The Neural Code for Face Orientation in the Human Fusiform Face Area. J Neurosci 34:12155–12167. https://doi.org/10.1523/JNEUROSCI.3156-13.2014

50. 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 103:3349–3365. https://doi.org/10.1152/jn.01108.2009

51. Logothetis NK, Wandell BA (2004) Interpreting the BOLD signal. Annu Rev Physiol 66:735–769. https://doi.org/10.1146/annurev.physiol.66.082602.092845

52. Heeger DJ, Huk AC, Geisler WS, Albrecht DG (2000) Spikes versus BOLD: what does neuroimaging tell us about neuronal activity? Nat Neurosci 3:631–633. https://doi.org/10.1038/76572

53. Rees G, Friston K, Koch C (2000) A direct quantitative relationship between the functional properties of human and macaque v5. Nat Neurosci 3:716–723. https://doi.org/10.1038/76673

54. Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human v1. J Neurosci 16:4207–21

55. Logothetis NK (2003) The underpinnings of the BOLD functional magnetic resonance imaging signal. J Neurosci 23:3963–3971. https://doi.org/10.1523/JNEUROSCI.23-10-03963.2003

56. González Ballester MÁ, Zisserman A, Brady M (2000) Segmentation and measurement of brain structures in MRI including confidence bounds. Med Image Anal 4:189–200. https://doi.org/10.1016/S1361-8415(00)00013-X

57. Handwerker DA, Ollinger JM, D'Esposito M (2004) Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. NeuroImage 21:1639–1651. https://doi.org/10.1016/j.neuroimage.2003.11.029

58. Dragoi V, Sharma J, Sur M (2000) Adaptation-induced plasticity of orientation tuning in adult visual cortex. Neuron 28:287–298. https://doi.org/10.1016/S0896-6273(00)00103-3

59. Patterson CA, Wissig SC, Kohn A (2013) Distinct effects of brief and prolonged adaptation on orientation tuning in primary visual cortex. J Neurosci 33:532–543. https://doi.org/10.1523/JNEUROSCI.3345-12.2013

60. Wissig SC, Kohn A (2012) The influence of surround suppression on adaptation effects in primary visual cortex. J Neurophysiol 107:3370–3384. https://doi.org/10.1152/jn.00739.2011

61. Buxton RB, Frank LR (1997) A Model for the Coupling between Cerebral Blood Flow and Oxygen Metabolism during Neural Stimulation. J Cereb Blood Flow Metab 17:64–72. https://doi.org/10.1097/00004647-199701000-00009

62. Webb BS (2005) Early and late mechanisms of surround suppression in striate cortex of macaque. J Neurosci 25:11666–11675. https://doi.org/10.1523/JNEUROSCI.3414-05.2005

63. Bao P, Purington CJ, Tjan BS (2015) Using an achiasmic human visual system to quantify the relationship between the fMRI BOLD signal and neural response. eLife 4:1–21. https://doi.org/10.7554/elife.09600

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

65. Friston KJ, Mechelli A, Turner R, Price CJ (2000) Nonlinear Responses in fMRI: The Balloon Model, Volterra Kernels, and Other Hemodynamics. NeuroImage 12:466–477. https://doi.org/10.1006/nimg.2000.0630

66. Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436. https://doi.org/10.1163/156856897X00357

67. Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442. https://doi.org/10.1163/156856897X00366

68. Cornelissen FW, Peters EM, Palmer J (2002) The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox. Behav Res Methods Instrum Comput 34:613–617. https://doi.org/10.3758/BF03195489

69. Rademaker RL, Chunharas C, Serences JT (2019) Coexisting representations of sensory and mnemonic information in human visual cortex. Nat Neurosci 22:1336–1344. https://doi.org/10.1038/s41593-019-0428-x

70. Benson NC, Jamison KW, Arcaro MJ, et al (2018) The Human Connectome Project 7 Tesla retinotopy dataset: Description and population receptive field analysis. J Vis 18:23. https://doi.org/10.1167/18.13.23

71. Kriegeskorte N, Mur M, Ruff DA, et al (2008) Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron 60:1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

72. Esteban O, Markiewicz C, Blair RW, et al (2018) {fMRIPrep}: a robust preprocessing pipeline for functional {MRI}. Nat Methods. https://doi.org/10.1038/s41592-018-0235-4

73. Esteban O, Blair R, Markiewicz CJ, et al (2018) fMRIPrep. Software. https://doi.org/10.5281/zenodo.852659

74. Gorgolewski K, Burns CD, Madison C, et al (2011) Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. Front Neuroinformatics 5:13. https://doi.org/10.3389/fninf.2011.00013

75. Gorgolewski KJ, Esteban O, Markiewicz CJ, et al (2018) Nipype. Software

76. Abraham A, Pedregosa F, Eickenberg M, et al (2014) Machine learning for neuroimaging with scikit-learn. Front Neuroinformatics 8:. https://doi.org/10.3389/fninf.2014.00014

77. Roth ZN, Heeger DJ, Merriam EP (2018) Stimulus vignetting and orientation selectivity in human visual cortex. eLife 7:1–25. https://doi.org/10.7554/eLife.37241

78. Carlson TA (2014) Orientation Decoding in Human Visual Cortex: New Insights from an Unbiased Perspective. J Neurosci 34:8373–8383. https://doi.org/10.1523/JNEUROSCI.0548-14.2014

79. 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 37:1187–1196. https://doi.org/10.1523/JNEUROSCI.2690-16.2016

80. Benson NC, Winawer J (2018) Bayesian analysis of retinotopic maps. eLife 7:0–45. https://doi.org/10.7554/eLife.40224

81. Morey RD (2008) Confidence intervals from normalized data: a correction to cousineau (2005). Tutor Quant Methods Psychol 4:61–64. https://doi.org/10.3758/s13414-012-0291-2

82. Carpenter B, Gelman A, Hoffman MD, et al (2016) Stan : A Probabilistic Programming Language. J Stat Softw VV:1–43. https://doi.org/10.18637/jss.v076.i01

83. Stan Development Team (2019) RStan: the R interface to Stan

84. Duane S, Kennedy AD, Pendleton BJ, Roweth D (1987) Hybrid Monte Carlo. Phys Lett B 195:216–222. https://doi.org/10.1016/0370-2693(87)91197-X

85. Hoffman M, Gelman A (2014) The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res 15:1593–1623

86. Betancourt M (2018) A conceptual introduction to Hamiltonian Monte Carlo. arXiv. https://doi.org/arXiv:1701.02434v2

87. Gelman A, Rubin DB (1992) Inference from Iterative Simulation Using Multiple Sequences. Stat Sci 7:457–472. https://doi.org/10.1214/ss/1177011136

88. Allen Institute for Brain Science (2016) Allen Brain Observatory. observatory.brain-map.org/visualcoding/

89. Allen Institute for Brain Science (2015) Allen SDK

90. Kohonen T (2000) Self organizing maps, Third. Springer, Berlin-Verlag

91. Boynton GM (2005) Imaging orientation selectivity: decoding conscious perception in v1. Nat Neurosci 8:541–542. https://doi.org/10.1038/nn0505-541

Figures

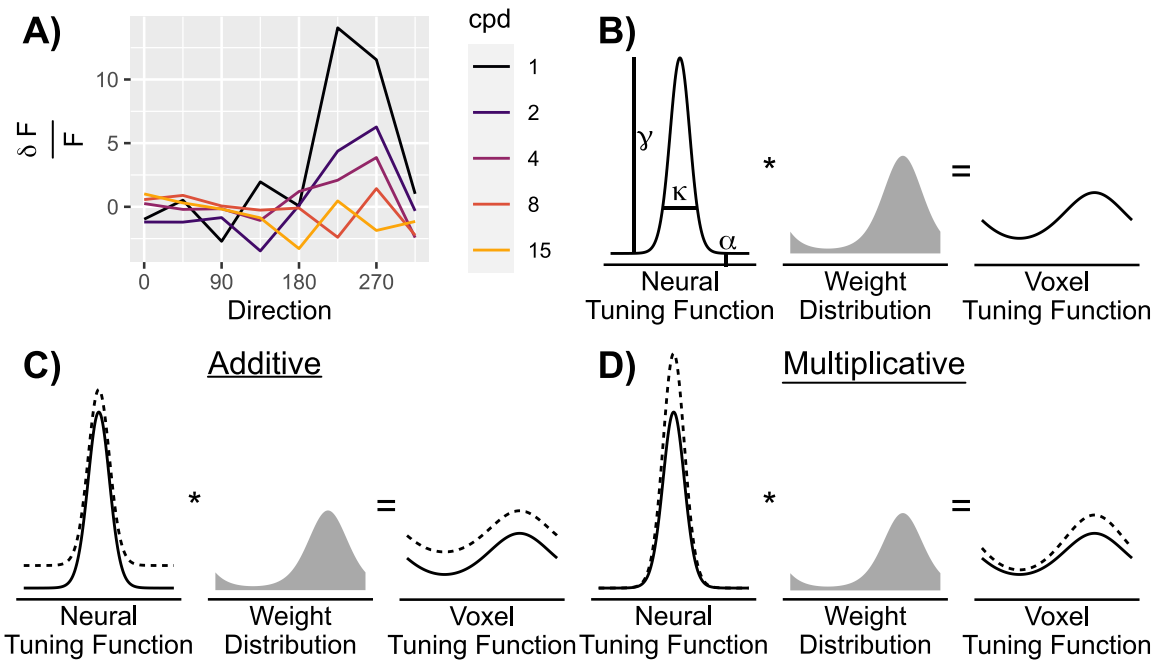

Figure 1 **A)** Modulation of neural tuning. In this neuron (mouse primary visual cortex), direction tuning interacts multiplicatively with spatial frequency (cpd: cycles per degree; $\delta F/_F$: average change in fluorescence) [88, 89]. **B)** Convolution ($*$) of neural tuning functions with a weight distribution leads to voxel tuning functions. The modeling procedure of NMM uses the observed voxel tuning functions to estimate parameters describing the shape of the neural tuning functions and the distribution of neurons across preferred orientations (the weight distribution) within each voxel. Compare with Equation 3. **C)** An additive shift in the neural tuning function causes an additive shift in the voxel tuning function. **D)** Multiplicative gain in the neural tuning function causes multiplicative gain in the voxel tuning function. Compare C and D with Equation 4.

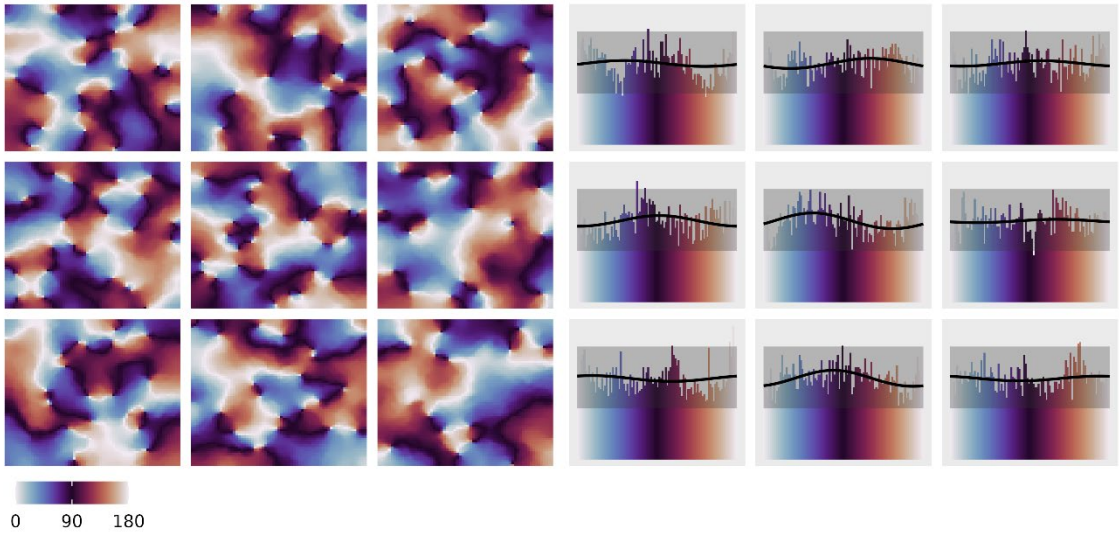

Figure 2 Simulation of continuous orientation preferences. **Left:** A Kohonen network was used to simulate orientation preference maps [90, SI]. Each panel represents a separate voxel. **Right:** Weight distribution for each voxel. The histograms represent the tallies of each preferred orientation in the left panels. The black line is the best fitting (maximum likelihood estimation) von Mises density function (Equation 1). The horizontal gray (shaded) ribbons give the 99% confidence interval for a flat distribution (i.e., the range of heights that we would expect when a simulated voxel contains an equal number of neurons tuned to all orientations). Given the amount of cortex sampled by each stimulated voxel, most weight distributions tend to be flat, but the fitted von Mises density function can capture individual modes [see also 91].

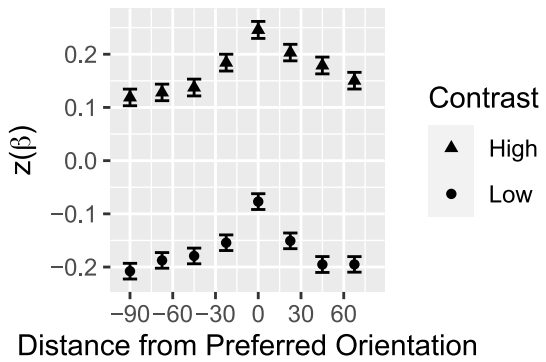

Figure 3 Observed Voxel Tuning Function at Low and High Contrast. Error bars give 95% confidence interval around the mean.

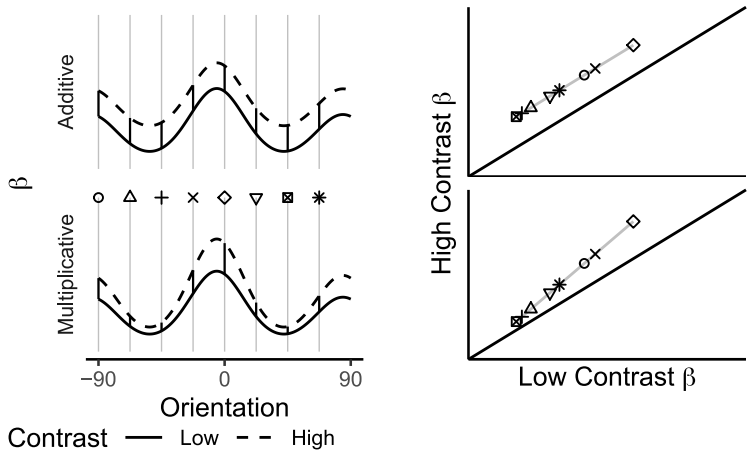

Figure 4 Plotting a voxel's response to high versus low contrast orientation uncovers the form of neuromodulation. **Left:** Simulated voxel tuning functions in which higher levels of contrast induce either an additive (top) or multiplicative (bottom) neuromodulation. The eight vertical lines are eight hypothetical orientations at which these voxel tuning functions might be probed, which would produce eight responses per level of contrast, High vs. Low. Note that, whereas the Bayesian voxel tuning function model assumes that the weight distribution and neural tuning function follow specific shapes (e.g., unimodal distributions), the slope test does not (e.g., the voxel tuning function can be bimodal, as displayed). **Right:** The two kinds of neuromodulation reveal different signatures when the responses to high contrast stimuli are plotted against the responses to low-contrast stimuli. The diagonal line corresponds to no effect of contrast. A line drawn through the points produced by the additive model necessarily has a slope equal to 1 (top); under this neuromodulation. the effect of contrast does not depend on the orientation. A line drawn through the points produced by the multiplicative model necessarily has a slope greater than 1 (bottom); under this neuromodulation, the effect of contrast is largest at those orientations that are closest to the voxel's preferred orientation.

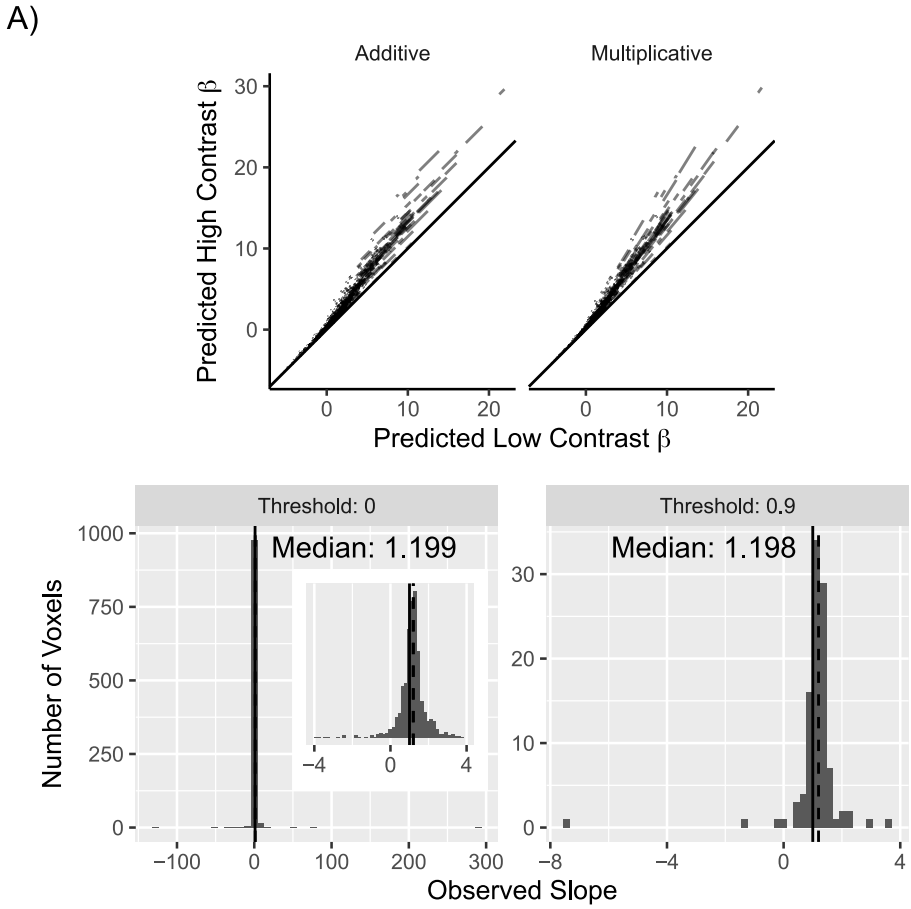

Figure 5 Multiplicative model wins because it alone can capture the interaction between orientation and contrast within voxels, evident as slopes greater than 1 in a plot of low- against high-contrast activity. Compare to Figure 4. **(A)** Posterior distribution of the predicted responses from the two models, plotted as a separate line for each voxel that connects the average predicted responses at each of the tested stimulus orientations. Most voxels are weakly responsive at each level of contrast and are only weakly influenced by the experimental manipulation (i.e., lines cluster in the lower left corner, close to the diagonal). Across voxels (i.e., comparing separate lines), both models can capture how contrast has a larger effect on the most responsive voxels: predictions lying further rightward along the x-axis are located further away from the diagonal. However, only the multiplicative model can capture the interaction between orientation and contrast within voxels. That is, some voxels (individual lines) have slope larger than one in the multiplicative model, but never in the additive model. **(B)** Distribution of Within-Voxel Slopes. The distributions include either all voxels (left) or only the 10% most responsive voxels (right). Thresholding voxels (according to their difference in average activity across levels of contrast, see main text) reduces the variability of the distribution (compare the x-axis ranges in the left and right plots), but in both cases the median slope is above 1. In the left plot, the inset shows a close-up of the distribution around the median, demonstrating the similarity of the median to that shown in the right plot.