


Model mimicry limits conclusions about neural tuning and can mistakenly imply unlikely priors

Received: 31 January 2024

Michael J. Wolff  & Rosanne L. Rademaker 

Accepted: 5 June 2025

Published online: 01 July 2025

 Check for updatesARISING FROM W. J. Harrison et al. *Nature Communications* <https://doi.org/10.1038/s41467-023-41027-w> (2023)

In a recent issue of *Nature Communications*, Harrison, Bays, and Rideaux use electroencephalography (EEG) to infer population tuning properties from human visual cortex, and deliver a major update to existing knowledge about the most elemental building block of visual perception – orientation tuning. Using EEG together with simulations in an approach they refer to as generative forward modeling, the authors adjudicate between two competing population tuning schemes for orientation tuning in visual cortex. They claim that a redistribution of orientation tuning curves can explain their observed pattern of EEG results, and that this tuning scheme embeds a prior of natural image statistics that exhibits a previously undiscovered anisotropy between vertical and horizontal orientations. If correct, this approach could become widely used to find unique neural coding solutions to population response data (e.g., from EEG) and to yield a true population tuning scheme deemed generalizable to other instances. However, here we identify major flaws that invalidate the promise of this approach, which we argue should not be used at all.

First, we will examine the premise of Harrison and colleagues¹, to subsequently explain why generative forward modeling cannot circumvent model mimicry pitfalls and can deliver many possible solutions of unknowable correctness. Finally, we show a tentative alternative explanation for the data.

Invasive neural recording techniques are the gold standard and only direct measurement tool for quantifying neural orientation tuning properties in visual cortex. Harrison and colleagues¹ point to previous research as precedence for the overrepresentation of horizontal compared to vertical orientations, citing work showing higher contrast energy for horizontal compared to vertical orientations in natural images^{2,3}, and inversely related differences in behavioral sensitivity when using images with broadband orientation contents^{2,4}. These theoretical findings could imply a cardinal asymmetry at the neural level. Indeed, Harrison and colleagues¹ also prominently feature an apparent overrepresentation of horizontal compared to vertical selective neurons in visual areas of mouse⁵, cat⁶, and macaque⁷ in Fig. 1B of their paper. How these data were derived and plotted is not

described, and statistical tests of vertical-horizontal anisotropies are not reported. However, Harrison and colleagues seem to misinterpret data from these physiology studies. While all three studies provide evidence for an overrepresentation of neurons tuned to cardinal (horizontal and vertical) compared to oblique (diagonal) orientations (the well-known oblique effect⁸), they do not, in fact, set out to test for or convincingly show anisotropies between vertical and horizontal orientations. In mouse, Roth and colleagues⁵ show a weak trend of more horizontal versus vertical selectivity in mouse V1, but an *opposite* trend in a later visual area (Posteromedial area). In cat, Wang and colleagues⁶ show a similar weak trend favoring horizontal selectivity in cat visual cortex, but other cat studies that show no or opposite trends can be easily found (e.g.,^{9,10}). In macaque, Fang and colleagues⁷ analyzed data from a total of 48 V1 hemispheres and 38 V4 hemispheres (in over 34 animals), showing no trends favoring horizontal orientations in V4, and an opposite trend in V1. (The first data figure in Fang and colleagues⁷ shows a single example V4 hemisphere where a trend for more horizontal-preferring neurons can be observed, similar to Fig. 1b in Harrison and colleagues. This trend is absent in the full V4 data.) This higher selectivity for vertical was statistically significant, but likely due to concurrent radial biases. That is, neurons fire preferentially for orientations aligned with the radial angle between their receptive field location and fixation (radial bias), and Fang and colleagues⁷ sampled V1 neurons at radial angles closer to the vertical meridian more densely than those closer to the horizontal meridian. Thus, even this small selection of animal physiology research from a field that spans decades (starting with Hubel & Wiesel in 1962¹¹) shows no consistent evidence for or against differences between vertical and horizontal neural selectivity, and it is unclear to us how Harrison and colleagues can imply otherwise. If systematic neural anisotropies between vertical and horizontal selectivity exist, we are not aware of research that has systematically evaluated the existing literature or actually applied quantitative tests, but this would certainly be worth looking into. Furthermore, while some orientation selectivity is present innately¹² the implied evolutionary justification for an anisotropy favoring

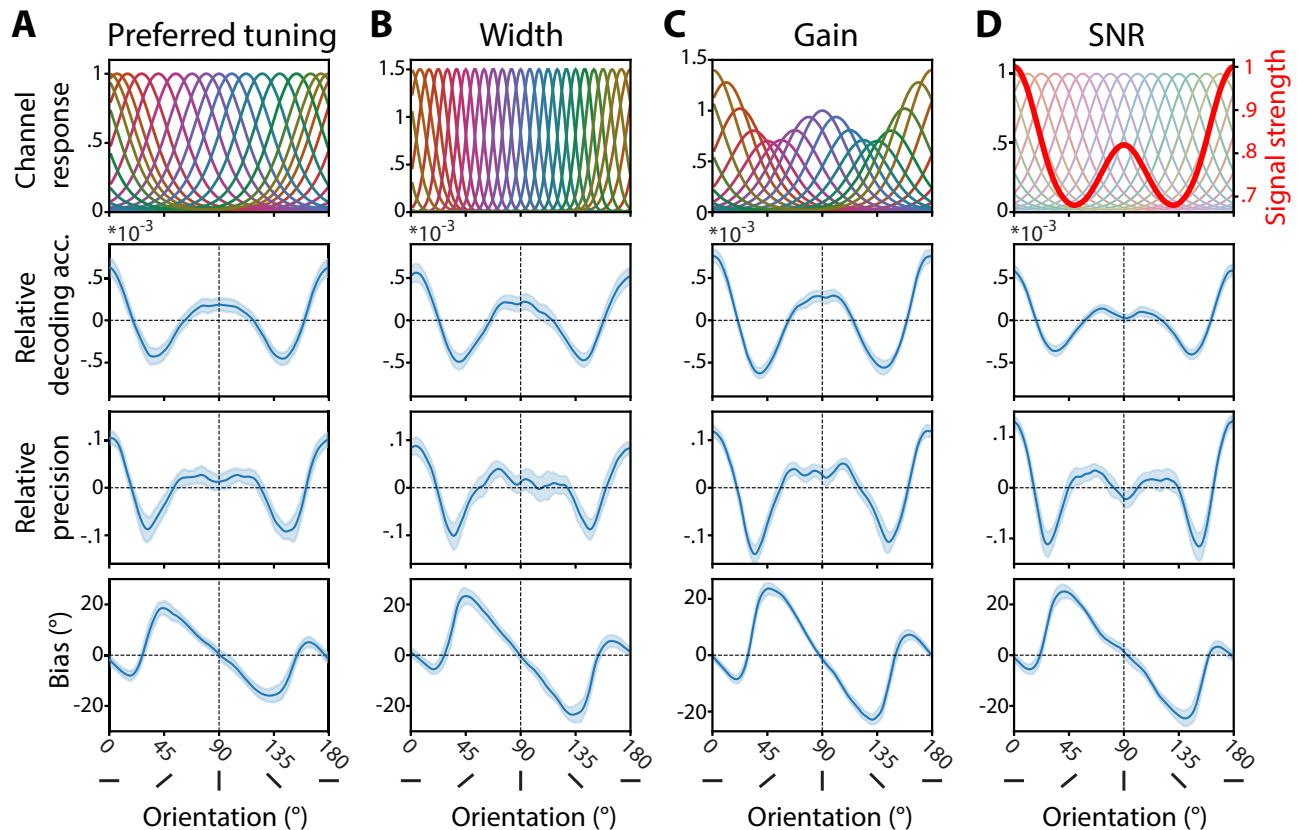


Fig. 1 | Model mimicry: many models produce the same pattern of results. In generative forward modeling, EEG data are simulated from models that use different sets of orientation tuning functions (top row). Decoding results (mean-centered decoding accuracy, mean-centered precision, and bias) as a function of orientation are shown (3 bottom rows) for simulations using different underlying example models. Blue error areas are 95% confidence intervals of the mean of the simulated instances ($n = 36$) of each model for each decoding metric. **A** Preferred tuning model: Tuning functions are unevenly spaced along the orientation space,

with more clustering at vertical, and even more at horizontal orientations. This is the best fitting model from Harrison and colleagues¹. **B** Width model: Tuning curve widths are uneven, with narrowest tuning for obliques, wider tuning for vertical and widest tuning for horizontal. **C** Gain model: Uneven tuning curve gain across orientations space, with more gain at cardinals that is highest for horizontal orientations. **D** Signal-to-noise (SNR) model: Tuning curves are uniform, but signal strength is orientation specific. Source data are provided as a Source Data file.

horizontal over vertical orientations that is seemingly mirrored in the statistics of natural but not man-made scenes, does not take into account physiological research emphasizing the extent to which sensory input during development can shape orientation tuning^{12–14}. (The justification for an embedded prior based on natural scene statistics (i.e., the green line in Harrison and colleagues Figure 7B) comes from measurements by Girshick and colleagues¹⁵ over 6 levels of image resolution. It is unclear which resolution the green line is based on or why).

Aside from its premise, the central flaw in Harrison and colleagues¹ lies with the fact that EEG decoding results cannot inform about the underlying neural or population tuning, due to an inherent inverse problem and model mimicry. The inverse problem is where an underlying cause cannot be inferred from a (measurable) effect, such as the inability to estimate neural causes from non-invasive imaging results. The inability to model single neurons and other limitations of population response models are discussed at length in refs. 16–19. Relatedly, model mimicry (or model degeneracy) refers to cases where many possible models can generate the same, or very similar, outcomes and model fits. To test what population tuning properties explain their pattern of EEG results, Harrison and colleagues¹ claim they can use generative forward modeling (see also²⁰) to differentiate between two possible population tuning schemes: differences in tuning widths and differences in tuning preferences. We do not believe this claim is justified. Using the same simulation approach as Harrison and colleagues (but a slightly different decoder, see “Methods” and

Supplementary Fig. 1), we show several examples of population tuning schemes that all yield the same pattern of results at the macro-level (Fig. 1). Importantly, this includes the population tuning scheme with different tuning widths that Harrison and colleagues argued could not fit their data (Fig. 1B). This is because they did not consider the entire parameter space for this model, omitting schemes with wider tuning at cardinals compared to obliques. In addition to models with differences in tuning width or preference, a physiologically sensible model with differences in gain modulation²¹ could also explain the data (Fig. 1C), but was not considered by Harrison and colleagues¹. In fact, even a uniform set of tuning curves can approximate the data, as long as the signal-to-noise ratio (SNR) is modulated across orientation space (Fig. 1D). Indeed, we argue that it boils down to orientation-specific SNR differences that need to be simulated to approximate the EEG results. This can be done either by explicitly changing the signal strengths (Fig. 1D) or by modulating the tuning curves in some way (preferred tuning, width, and gain as in Fig. 1A–C, or any combination thereof). To complicate matters further, when adjudicating between models, or even when deciding if a single model approximates the data, there are many possibly sensible, but fairly arbitrary parameters to pick from (e.g., the number of tuning functions, the range of tuning function widths, etc.) that can all generate outcomes that mimic each other or impact SNR. Furthermore, model specification is not limited to sensible choices only—the tuning functions used by Harrison and colleagues¹ may be well-motivated models¹⁶, but a set of arbitrarily shaped functions could also be used to simulate data and/or recover

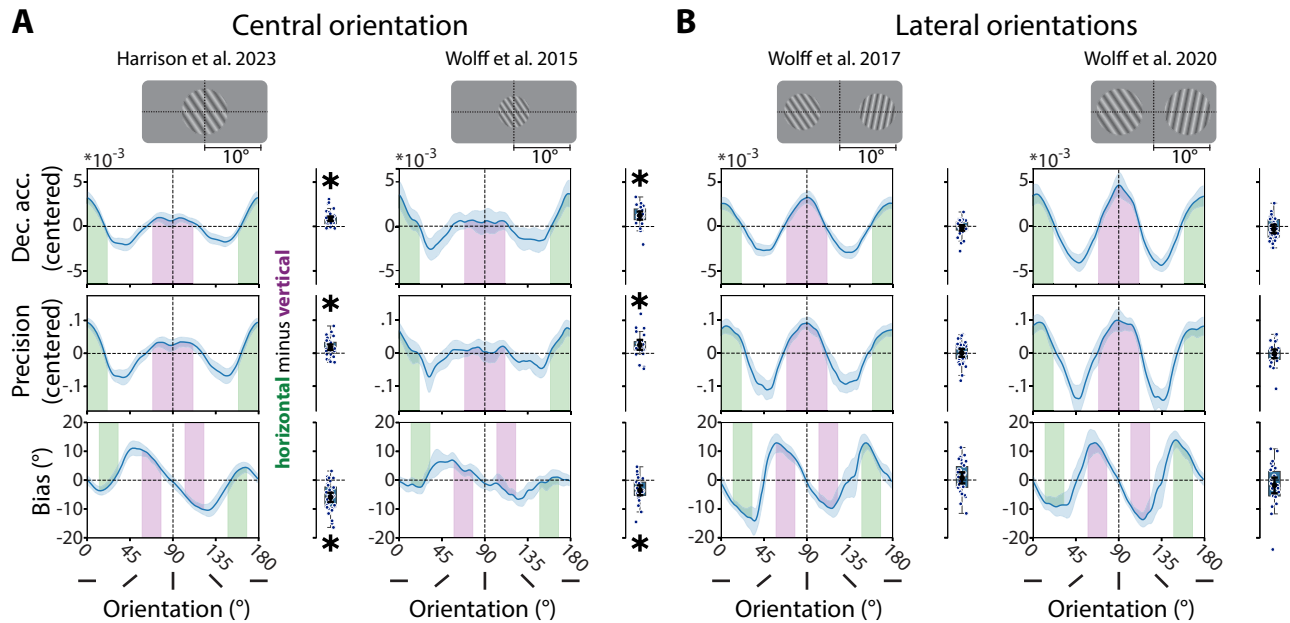


Fig. 2 | Cardinal anisotropies for orientation decoding are specific to central stimuli in EEG. Stimulus positions and sizes (top) are shown to scale. Dashed lines depict the vertical and horizontal meridians. Black horizontal line illustrates 10° visual angle from fixation. **A** Re-analyses of experiments reported by Harrison and colleagues¹ (left: $n = 36$) and Wolff and colleagues²³ (right: $n = 24$) where central orientations were shown to participants. Line plots show mean-centered Mahalanobis distance-based decoding metrics as a function of orientation, with shaded areas indicating the cardinal orientation bins used to compute differences between horizontal (green) and vertical (purple) orientations. Blue error areas are 95% confidence intervals (C.I.). The box plots show decoding metric differences for horizontal minus vertical orientations, with box limits indicating the upper and lower quartiles of the data, whiskers indicating 1.5 times the inter-quartile range, and blue dots representing individual subjects. The superimposed black circle and error-bars indicate the mean and 95% C.I. *Top*: Mean-centered accuracy (mean-centered cosine vector mean of pattern similarity curve), *Middle*: Mean-centered

precision (1 minus the circular standard deviation of decoded orientation across trials), *Bottom*: Bias of pattern similarity curves, in degrees. Both data-sets show statistically significant differences between horizontal and vertical orientations, with higher decoding for horizontal orientations (both $p < 0.001$), higher precision for horizontal orientations (left: $p < 0.001$, right: $p = 0.003$), and a stronger attraction toward vertical orientations (both: $p < 0.001$). Tests were two-sided permutation t-tests (10,000 permutations). No adjustments for multiple comparisons were made. **B** Re-analyses of experiments with orientations presented laterally^{24,25} (left: $n = 30$; right: $n = 26$) to the left and right of fixation, at an eccentricity of 6.69° or 6.08° (for data from²⁴ and²⁵, respectively). Same conventions as in A. No consistent differences between horizontal and vertical orientations. (Decoding accuracy difference, left: $p = 0.315$, right: $p = 0.232$; precision difference, left: $p = 0.837$, right: $p = 0.895$; attraction difference, left: $p = 0.43$, right: $p = 0.236$; two-sided, not corrected for multiple comparisons). Source data are provided as a Source Data file.

decoding metrics¹⁷. Thus, even when a model clearly fails to approximate the data, like the width model with sharpening around cardinal orientations from Harrison and colleagues¹, one should be careful to falsify such a model outright. We aren't entirely confident that even this sharpening-at-cardinal model might not still fit when combined with a different set of (sensible or arbitrary) parameter choices that compensate for the model's low SNR at cardinals. Importantly, our model mimicry argument should not be taken to discount all modeling of (neuroimaging) data, as modeling can serve many useful purposes. Models often are the best description of neuroscientific theory that we have available and can guide experiments.

Across two EEG data sets from previously published manuscripts^{20,22}, Harrison and colleagues¹ show that horizontal orientations result in notably better and more consistent decoding than vertical orientations. (Note that generative forward modeling was only applied to the data from Rideaux et al.²⁰, and not to the data from King & Wyart²². The latter shows a diverging pattern of relative precision across orientation space, with peaks at cardinals and obliques – a pattern that the model does not reproduce.) Given that generative forward modeling cannot be used to infer orientation tuning anisotropies in human visual cortex as a plausible explanation, what other factors might be driving these EEG results? Using the same decoding method as for the simulations (Fig. 1), we replicate the pattern of results both for the dataset²⁰ used by Harrison and colleagues¹, as well as for another openly available EEG dataset where orientated grating stimuli were presented centrally²³ (Fig. 2A). However, this effect is not replicated for EEG data sets^{24,25} where orientation gratings were

presented laterally, just to the left and right of fixation along the horizontal meridian (Fig. 2B), where no decoding differences between vertical and horizontal orientations are evident (see Supplementary Fig. 2 for overlaid grating sizes and positions to scale of all experiments).

Thus, simply presenting gratings slightly off-center leads to markedly different results that do not align with Harrison and colleagues' central premise of more neurons tuned to horizontal compared to vertical⁷. Note that it has been found that at the single neuron level, eccentricity does not seem to have an impact on orientation selectivity in primate V1, where selectivity does not differ between cells with receptive fields closer ($<5.2^\circ$) or further ($>5.2^\circ$) away from fixation⁷. So why do we see this striking difference between centrally and laterally presented stimuli in the EEG data? We hypothesize that the cardinal anisotropy seen only for centrally presented gratings could be driven by V1 surface area anisotropies – i.e., anisotropies that affect processing of visual field location instead of orientation. Human V1 has about double the cortical surface area dedicated to the horizontal compared to the vertical meridian, and human visual performance is higher for stimuli presented along the horizontal compared to the vertical meridian, especially in the periphery^{26,27}. A central stimulus drives responses all around fixation, including both the horizontal and vertical meridians. This means central stimuli are susceptible to V1 surface area anisotropies, with not all parts of the stimulus processed equally. Conversely, laterally presented stimuli fall along only a single meridian, meaning no surface area anisotropies. In EEG, further anisotropies may arise due to the organization of the visual field map in cortex, which

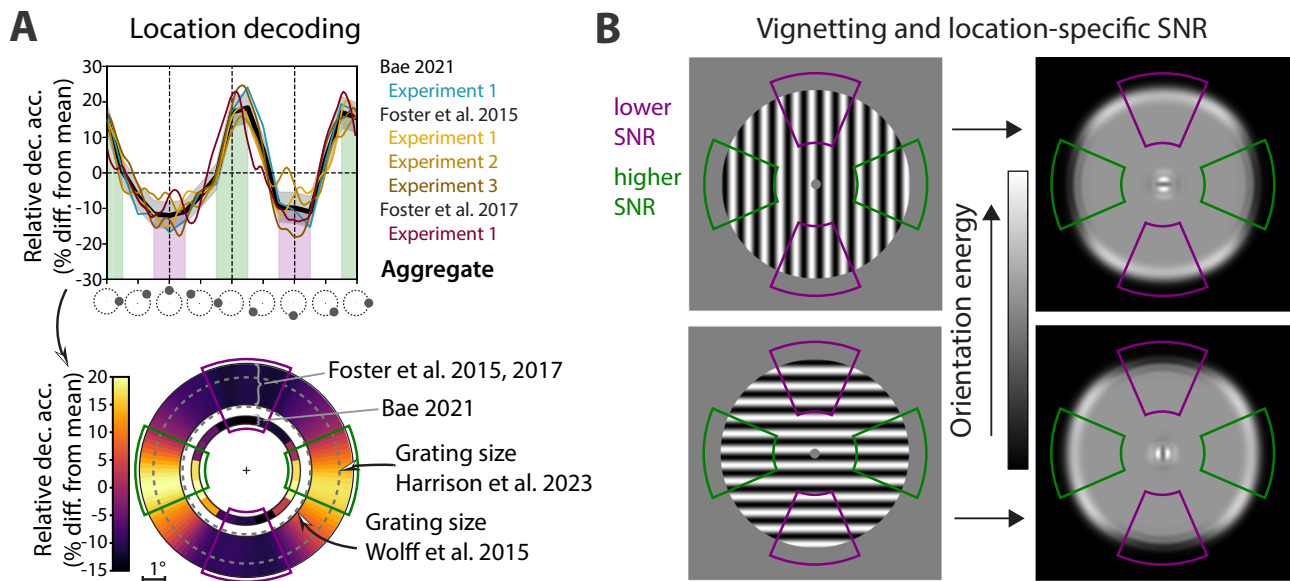


Fig. 3 | Differences in measurement sensitivity across the visual field that interact with stimulus vignetting can explain decoding differences between vertical and horizontal orientations that are presented centrally. **A Top:** Relative location decoding accuracy (percentage difference from the mean) as a function of presented stimulus location (re-analyses of^{31–33}, $n = 77$ over all experiments). Green and purple shadings highlight stimuli presented on the horizontal and vertical meridians, respectively. Black error shading of the aggregate is the 95% C.I. of the mean of all participants. **Bottom:** Relative location decoding across the visual field (percentage difference from the mean), replotted to scale for the various experiments: Outer, thicker ring represents possible locations of dot stimuli used in

Foster and colleagues^{32,33}, presented at 3.8°–4° eccentricity and 1.6° in diameter. Inner ring represents possible locations of the dot stimuli used in Bae³¹, presented at 2.3° eccentricity with 0.35° diameter. Dashed gray circles represent stimulus sizes of the central orientations used in Wolff and colleagues²³ and Harrison colleagues¹ (radii of 2.88° and 4.2°, respectively). **B** Vignetting³⁴ for gratings presented at the center of the screen: Orientation energy is highest on the stimulus edges aligned with the orientation. This means relatively higher orientation energy along the vertical meridian for vertical orientations (*top*) where SNR is low, and along the horizontal meridian for horizontal orientations (*bottom*) where SNR is high. Source data are provided as a Source Data file.

determines how well activity from different portions of cortex are captured by EEG scalp electrodes. For example, locations along the vertical meridian are processed closer to, and inside of, the longitudinal fissure²⁸, which is more difficult to measure with scalp electrodes. This could mean that EEG is particularly ill-suited for measuring orientation sensitivity, also because different orientations are predominantly processed in different regions of cortex^{29,30}, which may be unevenly sampled by scalp electrodes.

That differences between the vertical and horizontal meridians of the visual field play a role in EEG measurements, becomes evident when looking at location decoding from EEG signals. We re-analyzed multiple openly available EEG datasets where participants were presented with a single dot at one of many possible locations around fixation^{31–33}. We see clear and systematic differences in location decoding accuracies, with highest relative decoding for locations close to the horizontal meridian, and lowest for locations close to the vertical meridian (Fig. 3A). The eccentricity at which dot stimuli were presented in four of these datasets (outer ring in Fig. 3A, *bottom*) overlaps with the edge of the grating stimulus used by Harrison and colleagues (overlaid gray dotted line in Fig. 3A, *bottom*). The same is true for the eccentricity of dot stimuli in the dataset from Bae³¹, shown as the inner ring in Fig. 3A (*bottom*), which is close to edge of the full-field gratings used in Wolff and colleagues²³ where cardinal anisotropies are also observed (Fig. 2A). Concretely, for centrally presented stimuli, this difference in sensitivity across the visual field means lower SNR at the upper and lower stimulus edges than at the left and right stimulus edges (Fig. 3B, *left*).

Importantly, these location-specific SNR differences can interact with second order stimulus properties (stimulus edge effects or vignetting) that have been argued to at least in part be related to the decoded signal obtained from non-invasive neuroimaging^{34,35} (but see also refs. 29,36). Vignetting refers to the interaction between stimulus orientation and stimulus aperture, such that for circular gratings the

orientation energy is strongest on the edges of the grating aligned with the orientation (Fig. 3B, *right*). A vertically orientated grating presented centrally will therefore evoke more activity in the periphery of the vertical meridian, a visual field location where sensitivity is lower, leading to relatively lower decoding. A centrally presented horizontal grating will evoke more activity in the periphery of the horizontal meridian, where sensitivity is higher, leading to relatively higher decoding (Fig. 2A). This may not be true for laterally presented stimuli, where the decoded orientation energy falls into a part of the visual field where measurement sensitivity is more evenly distributed (i.e., along a single meridian). Unaffected by large SNR differences between the horizontal and vertical meridians, decoding results for lateral stimuli are similar for horizontal and vertical orientations (Fig. 2B).

We do not claim that this explanation is definitive or exhaustive. Rather, we want to highlight the importance of considering stimulus and measurement biases that can interact with orientation decoding. For example, spatial attention to the endpoints of orientated gratings³⁷ could also interact with visual field anisotropies in a manner similar to vignetting effects. Other factors may interact with measurement of orientation selectivity as well, such as stimulus contrast³⁸ or radial bias^{7,39}. However, radial bias should impact central and lateral stimuli similarly: Central stimuli are preferentially processed along the horizontal meridian, given the location-specific SNR differences described above. Lateral stimuli to the left and right of fixation are entirely processed along the horizontal meridian. Thus, both types of stimuli should have higher decoding for horizontal (radial from fixation) compared to vertical (tangential from fixation) orientations if radial bias had a measurable impact on orientation decoding with EEG. Yet, our analyses of lateral orientations do not show any decoding difference between horizontal and vertical orientations at all (Fig. 2B). Finally, the oblique effect which describes better perceptual performance for cardinal over oblique orientations⁸ and is mirrored in the overrepresentation of orientation-tuned neurons that prefer cardinal

over obliques^{5–7}, aligns with the EEG results for both centrally and laterally presented gratings (see Fig. 2). This implies that at least some form of well-established orientation anisotropy may be genuinely measurable with EEG. That said, the observed attenuation of decoding metrics for vertical compared to horizontal orientations, specific for centrally presented gratings, is likely driven by location-specific measurement differences that could interact with second-order stimulus properties (such as vignetting) or other factors.

In conclusion: Despite decades of research, invasive neural recordings in animals have not found the anisotropies between vertical and horizontal orientations seen in the EEG data reported by Harrison and colleagues¹. This pattern of results cannot be explained on the basis of the underlying orientation tuning, because generative forward modeling^{1,20} suffers from an inherent inverse problem, where many possible population tunings (including many physiologically plausible ones) can approximate the patterns of reported EEG data equally well. Given that the pattern of EEG results does not replicate for laterally presented stimuli, cardinal anisotropies are likely driven by other factors, such as differences in visual field sensitivity between the vertical and horizontal meridian and their interaction with second-order stimulus effects.

Methods

Simulations (generative forward modeling)

We simulated different population tuning models using largely the same approach as in Harrison and colleagues¹ by adapting their published Matlab scripts. Briefly, data for 36 subjects, from 32 EEG channels and 6480 trials per subject, was simulated for each model (see below). Each model had a given number of tuning functions, or model channels. The modeled channel responses to each orientation (1° to 180°, in steps of 1°) shown to a given model were transformed to EEG sensor space via matrix multiplication between the orientation-specific model response of each trial, and a random weights matrix (number of channel functions by number of EEG channels, sampled from a uniform distribution over 0 to 1). Trial-specific noise was added to each simulated EEG channel sampled from a normal distribution (s.d. = 6), which also ensures differences in simulated responses to trials on which identical orientations were shown.

For all models we will only mention any deviations from Harrison and colleagues¹ Preferred-tuning-model. The purpose of our simulations was to demonstrate that there is no unique model that best describes the data, even when only considering models that could be argued to be plausible. Our models (Fig. 1) are by no means the best fitting models, as searching for best fitting solutions in this very large parameter space would be computationally intractable. Indeed, we derived at our models through mere trial and error and stopped once we obtained decoding results that resembled Harrison and colleagues¹ Preferred-tuning-model. Thus, the models and their parameters described below should not be considered definitive; they are snapshots out of many more possibilities.

Preferred-tuning-model: For this we used the exact script published by Harrison and colleagues¹, which generates the preferred tuning model. This model consisted of 16 tuning functions with constant widths ($\kappa = 2$). Preference was modulated by shifting the tuning functions based on the sum of two von Mises derivative functions ($\kappa = 0.5$) centered on 0° (amplitude = 14) and on 90° (amplitude = 8), which has the effect that there are relatively more tuning functions around horizontal (0°) compared to vertical (90°), and fewest tuning functions around obliques (45° and 135°). Note that these values in the scripts uploaded by the original authors at the time of writing differ slightly from the values described in the manuscript (which states the amplitudes were 15 and 10). The resulting difference between these two parameter settings is marginal however, and we decided to stick to those parameters in the script as uploaded by the authors, without changing anything.

Width-model: Instead of changing the tuning preferences across the orientation space, tuning functions were evenly spaced, but their widths were modulated. This modulation was derived from the *inverse* of the sum of two von Mises functions ($\kappa = 0.5$), one centered on 0° (amplitude = 15) and one centered on 90° (amplitude = 4). Given the inversion, tuning widths were wider for cardinals than for obliques, with horizontal widths being wider than vertical widths. The possible tuning widths were rescaled such that they ranged from $\kappa = 7$ (the widest) to $\kappa = 19$ (the narrowest). The number of tuning functions increased to 24 (from the original 16) and every tuning function was scaled to range from 0 to 1.5.

Gain-model: The gain-model comprises 16 evenly spaced tuning functions with constant widths ($\kappa = 2$), but differences in scaled amplitude (i.e., gain). Gains were modulated from the sum of two von Mises functions ($\kappa = 0.5$), one centered on 0° (amplitude = 15) and the other on 90° (amplitude = 8). The range of gains were scaled from 0.7 (at the obliques) to 1.4 (at 0 degrees, which is horizontal).

SNR-model: Here we used a uniform distribution of 16 identical tuning functions, all with the same width ($\kappa = 2$) and all with the same gain (amplitude of 1). Unlike the models above, here we do not manipulate the underlying tuning response functions but instead modify the signal strengths of the simulated activity patterns across the EEG channels. Specifically, the signal strength was modulated for simulated response patterns generated from each of the 180 orientations using the sum of two von Mises functions ($\kappa = 0.5$), centered on 0 degrees (amplitude = 15) and on 90 (amplitude = 8). Signal strength modulation ranged from 0.68 (68% signal strength) to 1 (100% of signal strength). The signal strengths of the orientation-specific patterns were modulated *after* transforming activations from the tuning response functions to every possible orientation (1°–180°) into sensor space (as described in Harrison and colleagues¹ and above), meaning that the orientation-specific patterns of the simulated EEG sensors were multiplied by the corresponding signal strengths (0.68 to 1), before adding the same amount of Gaussian noise to each (s.d. = 6).

EEG data

We reanalyzed openly available EEG datasets of 9 experiments across 7 publications^{1,23–25,31–33}, where human participants viewed either circular orientation gratings, or locations. For the present manuscript, the stimulus sizes and locations that participants viewed while EEG was recorded are of particular interest, and are described in more detail below. Other details are available in the methods sections of the original publications.

Harrison et al.¹: Participants ($N = 36$, 23 female; age (years): $M = 23.8$, $SD = 4.6$) viewed serially presented, randomly orientated circular gratings (4.2° radius) centered around fixation. Each grating was presented for 50 ms, with an ISI of 150 ms between consecutive gratings. The task was to detect gratings with a lower spatial frequency.

Wolff et al.²³: Participants ($N = 24$, 12 female; age (years): $M = 22.2$, range 18–38) performed a visual working memory task, where the orientation of a grating had to be memorized for up to 2.6 seconds. Each circular grating was centrally presented (2.88° radius) for 200 ms, followed by a blank delay of at least 1.17 seconds.

Wolff et al.²⁴: Only experiment 1 was reanalyzed. Here, participants ($N = 30$, 13 female; age (years): $M = 24.9$, range 18 to 38) performed a retro-cue visual working memory task. Two randomly orientated circular gratings (radius of 3.345° each) were simultaneously presented on the horizontal meridian at 6.69° eccentricity. The presentation time was 250 ms, followed by a blank delay of 800 ms. The orientations of both gratings were behaviorally relevant during encoding.

Wolff et al.²⁵: Participants ($N = 26$, 17 female; age (years): $M = 25.8$, range 20 to 42) also performed a retro-cue visual working memory task with laterally presented, randomly orientated circular gratings. The gratings (radius of 4.255° each) were presented at 6.08° eccentricity

for 200 ms followed by a blank delay of 400 ms. The orientations of both gratings were behaviorally relevant during encoding.

Foster et al.³²: Participants performed a spatial working memory task in all three experiments. The visual stimulus on each trial in all three experiments was a dark gray circle (0.8° radius) presented on a random location of an invisible circle at 4° eccentricity. The participants' task was to memorize the location for a delay of at least 1 s (variable across experiments). In experiment 1 and 3, the circle was presented for 250 ms and in experiment 2 for 1 s. Sample size was $N = 15$ in all experiments (age range 18 to 35 years). Information about the sex of participants is not provided in the original manuscript.

Foster et al.³³: We reanalyzed experiment 1 ($N = 10$, 7 female; age range 18 to 35 years). Here, in each trial a single a randomly colored circle (0.8° radius) was presented on a random location of an invisible circle at 3.8° eccentricity. Stimulus duration was 100 ms, followed by a 1.2 s blank delay. The participants' task was to memorize and report the color of the colored circle in each trial.

Bae³¹: We reanalyzed experiment 1, where participants ($N = 22$, 16 female; mean age range 18 to 30), performed a spatial working memory task. The visual stimulus was a small circle (0.175° radius) presented for 200 ms on one of 16 discrete locations on an invisible circle at 2.3° eccentricity. A blank delay (1.3 s) followed after the offset of the circle. The task was to memorize and report the location of the circle on each trial.

Preprocessing

For all experiments, we used the voltage data the way it was published and preprocessed by the original authors.

For the subsequent decoding analyses, we used the voltage traces from 50 to 450 ms relative to stimulus onsets from the posterior electrodes, in line with Harrison and colleagues¹. Data from refs. 23–25 and the reanalyzed experiment 1 from ref. 31 all used the same electrode coverage, and the same 17 posterior channels were included in the corresponding analyses (P7, P5, P3, P1, Pz, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz and O2). The same electrodes were included for the data of ref. 1 in addition to the electrodes Iz, P9, and P10. The electrode coverage was lower for the reanalyzed experiments in refs. 32,33 and the included posterior electrodes for these data-sets were PO3, PO4, P3, P4, O1, O2, POz, and Pz.

Instead of decoding at each time-point separately within the time-window of interest and then averaging (as in Harrison and colleagues¹), we first reformatted the data in a manner similar to previous work²⁵ before feeding it to the decoder: To take advantage of the fact that stimulus-specific information is not only present in the activity patterns across electrodes, but also in the *temporal* pattern of the evoked voltage changes, we combined the channel and temporal dimensions to improve the sensitivity of the decoder. To do so, we first down-sampled the signal from the time-window of interest (50 ms to 450 ms, relative to stimulus onset) to 50 (51.2 Hz for Harrison and colleagues¹, due to the original sampling rate of 1024 Hz), and removed the mean activity level within each trial and electrode. The resulting, mean-centered 20 voltage values of each channel in each trial were then combined with the channel dimension. The number of dimensions for the decoder increased therefore 20-fold (number of down-sampled time-points by number of channels).

Stimulus decoding

None of the various decoding metrics (accuracy, precision, bias) in Harrison and colleagues¹ are specific to the commonly used inverted encoding model (IEM) in their paper. We used a Mahalanobis distance decoder²⁵ that yields qualitatively similar results as the IEM decoder (Supplementary Fig. 1). We made other minor analysis changes to improve consistency and robustness, such as using wider orientation bins, using repeated stratified random folds to split both real and simulated data, etc (see below).

We used the same approach to decode orientations from the simulated data and orientations/locations from the spatiotemporal signal from the EEG data sets. Location decoding was the same as orientation decoding apart from taking into account that orientations are in 180° space, while locations are in 360° space. We used an 8-fold cross-validation approach. First, trials were assigned to the closest of 16 evenly spaced orientations/locations (variable, see below). The trials were then randomly split into 8 folds using stratified sampling. The trials of 1 fold were held out for testing and the trials of the remaining 7 folds were part of the training data. The covariance of the train trials was estimated using a shrinkage estimator⁴⁰, before the number of trials in each orientation/location bin of the train data was equalized through random subsampling. The subsampled trials within each bin of the training set were then averaged. And the averaged bins were then convolved with a half cosine basis set raised to the 15th power⁴¹ to pool information across similar orientations/locations. The Mahalanobis distances between the left-out test trials and the averaged train bins were then computed. This procedure was repeated for all train/test fold combinations. The experiment of one dataset³¹ used exactly 16 evenly spaced locations. Here the original location labels were used, rendering the aforementioned binning unnecessary. All remaining datasets used random orientations/locations, for which the above procedure was run separately for 8 possible ways of binning the orientations (with bins centered at 0° to 168.75°, at 1.40625° to 170.1563°, at 2.8125° to 171.5625°, at 4.2188° to 172.9688°, at 5.625° to 174.375°, at 7.0313° to 175.7813°, at 8.4375° to 177.1875°, or at 9.8438° to 178.5938°, each in 16 steps of 11.25) or location spaces (same as for orientation, but converted to 360° space by multiplying all values by two). This means that for each trial, we obtained 16 times 8 = 128 Mahalanobis distance bins (with the exception of the dataset with only 16 discrete orientations, which resulted in exactly 16 distances per trial). Given the randomness of the initial folds and the subsampling within folds, the above procedure was repeated 20 times to obtain more robust results. Once all distances were obtained and averaged over repetitions, distances for each trial were mean centered by subtracting the average distance across all Mahalanobis bins from each. The distances were then ordered as a function of angular difference between test and train bin, obtaining pattern-similarity-curves for each trial. For experiments with two simultaneously presented orientation gratings, one on each side^{24,25}, each orientation was decoded separately.

Centered decoding accuracy, centered precision, and bias

Decoding accuracy was obtained for each trial by computing the cosine vector mean of the pattern-similarity-curve²⁴. Decoding accuracy was then averaged as a function of orientation/location using a sliding window (width = 11.25° for orientations, width = 25° for locations) that moved over angular space in steps of 1.40625°/2.8125° for orientations/locations. Mean-centered decoding accuracy was obtained by mean-centering the resulting orientation/location-specific decoding accuracy curve. Precision was obtained by taking the circular means of the trial-wise pattern-similarity-curves and calculating the inverse circular standard deviation over these means. Mean-centered precision was obtained by mean-centering the precision curve (same as for relative decoding accuracy). Bias was obtained by computing the circular mean of the averaged pattern-similarity-curves of all trials within each angular window (same as above).

For location decoding we obtained relative decoding accuracy (% difference from mean decoding) by subtracting the average decoding accuracy (of all locations) from each location bin and then dividing the mean-centered decoding accuracies by the average to obtain proportion-difference. This was multiplied by 100 to obtain %-difference.

For visualization, the angular relative decoding, relative precision and bias curves were smoothed across orientations/locations with a Gaussian smoothing kernel (s.d. = 2°/4° for orientations/locations).

To explicitly test differences in decoding accuracy and precision between vertical and horizontal orientations (as shown in Fig. 2), the respective decoding metrics were averaged from -22.5° to $+22.5^\circ$ relative to 0° and 90° degrees. For the bias we assumed that, given equal attraction towards each cardinal, the effect should be maximal for orientations 22.5° away from the cardinals, i.e., halfway the distance to the obliques, where the influence of each cardinal should be canceled out. We thus averaged the bias values from -12.5° to 12.5° relative to 22.5° and 157.5° for horizontal orientations, and relative to 67.5° and 112.5° for vertical orientations, after sign reversing bias values such that positive values always correspond to attraction to the nearest cardinal.

Edge effects (vignetting)

We used the perfect-cube-model³⁴ to illustrate a possible relationship between location-specific SNR differences, and orientation energy, strongest at the edges for circular gratings. We used the exact stimulus and model parameters as described in ref. 34. Briefly, two sine-wave gratings (one vertical, the other horizontal) were convolved with eight distinctly oriented 2D Gabor filters (0° to 157.5° , in steps of 22.5°), which all had the same spatial frequency as the sine-wave gratings. The output of each filter was normalized before taking the sum over all eight, resulting in the 2D orientation energy plot in Fig. 3B.

Statistical significance testing

The reported differences between horizontal and vertical orientations (Fig. 2) were tested for significance using a permutation t-test with 10,000 permutations as implemented by the python toolbox MNE. All tests were two-sided and the statistical significance threshold was $p < 0.05$.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data used in this study is openly available^{23–25,31–33}. For convenience, data from^{23–25} was reduced in size by only including electrodes and time-points of interest, and is available at <https://osf.io/bdf74/>. Source data are provided with this paper.

Code availability

The code used to generate the figures and results reported in this manuscript are available at <https://github.com/mijowolff/model-mimicry-and-unlikely-priors>⁴².

References

- Harrison, W. J., Bays, P. M. & Rideaux, R. Neural tuning instantiates prior expectations in the human visual system. *Nat. Commun.* **14**, 5320 (2023).
- Hansen, B. C. & Essock, E. A. A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes. *J. Vis.* **4**, 5 (2004).
- Harrison, W. J. Luminance and contrast of images in the THINGS database. *Perception* **51**, 244–262 (2022).
- Hansen, B. C., Essock, E. A., Yufeng, Z. & J Kevin, D. Perceptual anisotropies in visual processing and their relation to natural image statistics. *Netw. Comput. Neural Syst.* **14**, 501 (2003).
- Roth, M. M., Helmchen, F. & Kampa, B. M. Distinct functional properties of primary and posteromedial visual area of mouse neocortex. *J. Neurosci.* **32**, 9716–9726 (2012).
- Wang, G., Ding, S. & Yunokuchi, K. Difference in the representation of cardinal and oblique contours in cat visual cortex. *Neurosci. Lett.* **338**, 77–81 (2003).
- Fang, C., Cai, X. & Lu, H. D. Orientation anisotropies in macaque visual areas. *Proc. Natl. Acad. Sci.* **119**, e2113407119 (2022).
- Lennie, P. Distortions of Perceived Orientation. *Nat. N. Biol.* **233**, 155–156 (1971).
- Berman, N. E., Wilkes, M. E. & Payne, B. R. Organization of orientation and direction selectivity in areas 17 and 18 of cat cerebral cortex. *J. Neurophysiol.* **58**, 676–699 (1987).
- Dragoi, V., Turcu, C. M. & Sur, M. Stability of cortical responses and the statistics of natural scenes. *Neuron* **32**, 1181–1192 (2001).
- Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154 (1962).
- Blakemore, C. & Cooper, G. F. Development of the brain depends on the visual environment. *Nature* **228**, 477–478 (1970).
- Hirsch, H. V. B. & Spinelli, D. N. Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. *Science* **168**, 869–871 (1970).
- Sengpiel, F., Stawinski, P. & Bonhoeffer, T. Influence of experience on orientation maps in cat visual cortex. *Nat. Neurosci.* **2**, 727–732 (1999).
- Girshick, A. R., Landy, M. S. & Simoncelli, E. P. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* **14**, 926–932 (2011).
- Sprague, T. C., Boynton, G. M. & Serences, J. T. The importance of considering model choices when interpreting results in computational neuroimaging. *eNeuro* **6**, 0196–19 (2019).
- Gardner, J. L. & Liu, T. Inverted encoding models reconstruct an arbitrary model response, not the stimulus. *eNeuro* **6**, 0363–18 (2019).
- Liu, T., Cable, D. & Gardner, J. L. Inverted encoding models of human population response conflate noise and neural tuning width. *J. Neurosci.* <https://doi.org/10.1523/JNEUROSCI.2453-17.2017> (2017).
- Sprague, T. C. et al. Inverted encoding models assay population-level stimulus representations, not single-unit neural tuning. *eNeuro* <https://doi.org/10.1523/ENEURO.0098-18.2018> (2018).
- Rideaux, R., West, R. K., Rangelov, D. & Mattingley, J. B. Distinct early and late neural mechanisms regulate feature-specific sensory adaptation in the human visual system. *Proc. Natl. Acad. Sci.* **120**, e2216192120 (2023).
- Wei, X.-X. & Stocker, A. A. A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nat. Neurosci.* **18**, 1509–1517 (2015).
- King, J.-R. & Wyart, V. The human brain encodes a chronicle of visual events at each instant of time through the multiplexing of traveling waves. *J. Neurosci.* **41**, 7224–7233 (2021).
- Wolff, M. J., Ding, J., Myers, N. E. & Stokes, M. G. Revealing hidden states in visual working memory using electroencephalography. *Front. Syst. Neurosci.* **9**, 123 (2015).
- Wolff, M. J., Jochim, J., Akyürek, E. G. & Stokes, M. G. Dynamic hidden states underlying working-memory-guided behavior. *Nat. Neurosci.* **20**, 864–871 (2017).
- Wolff, M. J., Jochim, J., Akyürek, E. G., Buschman, T. J. & Stokes, M. G. Drifting codes within a stable coding scheme for working memory. *PLOS Biol.* **18**, e3000625 (2020).
- Himmelberg, M. M., Winawer, J. & Carrasco, M. Polar angle asymmetries in visual perception and neural architecture. *Trends Neurosci.* **46**, 445–458 (2023).
- Himmelberg, M. M., Winawer, J. & Carrasco, M. Linking individual differences in human primary visual cortex to contrast sensitivity around the visual field. *Nat. Commun.* **13**, 3309 (2022).
- Wandell, B. A., Dumoulin, S. O. & Brewer, A. A. Visual field maps in human cortex. *Neuron* **56**, 366–383 (2007).
- Roth, Z. N., Kay, K. & Merriam, E. P. Natural scene sampling reveals reliable coarse-scale orientation tuning in human V1. *Nat. Commun.* **13**, 6469 (2022).

30. Sasaki, Y. et al. The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron* **51**, 661–670 (2006).
31. Bae, G.-Y. Neural evidence for categorical biases in location and orientation representations in a working memory task. *NeuroImage* **240**, 118366 (2021).
32. Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K. & Awh, E. The topography of alpha-band activity tracks the content of spatial working memory. *J. Neurophysiol.* **115**, 168–177 (2015).
33. Foster, J. J., Bsaies, E. M., Jaffe, R. J. & Awh, E. Alpha-band activity reveals spontaneous representations of spatial position in visual working memory. *Curr. Biol.* **27**, 3216–3223 (2017).
34. Carlson, T. A. Orientation decoding in human visual cortex: new insights from an unbiased perspective. *J. Neurosci.* **34**, 8373–8383 (2014).
35. Roth, Z. N., Heeger, D. J. & Merriam, E. P. Stimulus vignetting and orientation selectivity in human visual cortex. *eLife* **7**, e37241 (2018).
36. Pratte, M. S., Sy, J. L., Swisher, J. D. & Tong, F. Radial bias is not necessary for orientation decoding. *NeuroImage* **127**, 23–33 (2016).
37. Bae, G.-Y. & Luck, S. J. Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials. *J. Neurosci.* **38**, 409–422 (2018).
38. Maloney, R. T. & Clifford, C. W. G. Orientation anisotropies in human primary visual cortex depend on contrast. *NeuroImage* **119**, 129–145 (2015).
39. Freeman, J., Heeger, D. J. & Merriam, E. P. Coarse-scale biases for spirals and orientation in human visual cortex. *J. Neurosci.* **33**, 19695–19703 (2013).
40. Ledoit, O. & Wolf, M. Honey, I shrunk the sample covariance matrix. *J. Portf. Manag.* **30**, 110–119 (2004).
41. Myers, N. E. et al. Testing sensory evidence against mnemonic templates. *eLife* **4**, e09000 (2015).
42. Wolff, M. J. et al. Code for ‘Model mimicry limits conclusions about neural tuning and can mistakenly imply unlikely priors. Zenodo <https://doi.org/10.5281/zenodo.15480845> (2025).

Acknowledgements

We want to thank the multiple groups of scientists whose work we re-analyzed here. We were able to test our alternative hypotheses thanks to their readily available and well-documented data. We’d also like to thank Tommy Sprague and John Serences for reading an early version of this manuscript, and for their valuable feedback. This study was supported by funding from the Max Planck Society awarded to R.L.R.

Author contributions

M.J.W. and R.L.R. conceived the study and wrote the text. M.J.W. analyzed and modeled the data.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-60859-2>.

Correspondence and requests for materials should be addressed to Michael J. Wolff or Rosanne L. Rademaker.

Peer review information *Nature Communications* thanks Marc Himelberg and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025