



Published in final edited form as:

Nat Neurosci. 2020 January ; 23(1): 113–121. doi:10.1038/s41593-019-0544-7.

Binocular viewing geometry shapes the neural representation of the dynamic three-dimensional environment

Kathryn Bonnen^{1,2,3}, Thaddeus Czuba^{1,2}, Jake A Whritner², Adam Kohn⁴, Alexander C Huk^{1,2}, Lawrence K Cormack^{1,2}

¹Department of Neuroscience, University of Texas, Austin, TX, USA

²Department of Psychology, University of Texas, Austin, TX, USA

³Center for Neural Science, New York University, New York, NY, USA

⁴Dominick Purpura, Department of Neuroscience, Department of Ophthalmology and Visual Sciences, Department of Systems and Computational Biology Albert Einstein College of Medicine, The Bronx, NY, USA

Abstract

Sensory signals give rise to patterns of neural activity which the brain uses to infer properties of the environment. For the visual system, considerable work has focused on the representation of frontoparallel stimulus features and binocular disparities. But inferring the properties of the physical environment from retinal stimulation is a distinct and more challenging computational problem – this is what the brain must actually accomplish to support perception and action. Here we develop a computational model that incorporates projective geometry, mapping the three-dimensional (3D) environment onto the two retinae. We demonstrate that this mapping fundamentally shapes the tuning of cortical neurons and corresponding aspects of perception. For 3D motion, the model explains strikingly non-canonical tuning present in existing electrophysiological data and distinctive patterns of perceptual errors evident in human behavior. Decoding the world from cortical activity is strongly affected by the geometry that links the environment to the sensory epithelium.

For an animal to behave effectively in its environment, its nervous system must encode information well enough to support interactions with the dynamic world in real time. In the mammalian visual system, it is clear that early levels of cortical motion processing take as primitives the dynamic patterns of stimulation that fall upon the left and right retinae. Then subsequent decoding processes must allow the animal to interact with the dynamic, three-dimensional (3D) world. It is thus not the retinal motion that is ultimately important, but

Contributions

T.C. and A.K. collected the electrophysiology data. K.B., L.K.C., and A.C.H. built the computational model. K.B., T.C., A.K., A.C.H., and L.K.C. interpreted the results from electrophysiology and modeling. K.B., A.C.H. and L.K.C. designed the human psychophysical experiments. K.B. and J.A.W. collected the human psychophysical data. K.B., J.A.W., A.C.H., and L.K.C. performed the analysis and interpreted the results. K.B., L.K.C., and A.C.H. wrote the paper. K.B., T.C., J.A.W., A.K., A.C.H and L.K.C. edited the paper.

Competing Interests

The authors declare no competing interests.

rather inferring the 3D environmental motion that gave rise to the retinal stimulation and subsequent cortical activity.

In some cases, the stimulation upon the sensory epithelium is a fairly direct proxy for the stimulus in the environment. Some tactile perception works this way (i.e., if you feel a poke on your forearm, then something is poking your forearm). So too with a stimulus moving on a computer monitor: the mapping from monitor position to retinal position is straightforward. However, for most vision, there is a many-to-one mapping of 3D world positions (and velocities) to retinal positions (and velocities). Thus, for the visual system to work outside of the context of a laboratory's frontoparallel computer screen, decoding of the 3D environment requires additional computation to infer the properties of the world that gave rise to the stimulation upon the two retinae [1, 2].

In this work, we show how projective geometry, which maps the 3D environment to 2D images on each retina, results in strikingly discontinuous tuning functions for 3D motion in the Middle Temporal area (MT) of primate visual cortex. This encoding is starkly different in form from tuning functions observed for the reduced case of frontoparallel motion. Furthermore, predictions for the perception and estimation of 3D direction which result from these tuning curves show a distinctive dependence of error on 3D direction and systematic misperceptions of depth – patterns we then observe in human perceptual behavior. Theoretical analysis reveals that a key feature of the encoding-decoding computations for recovering 3D direction from the slightly different patterns of retinal stimulation are the small but ubiquitous differences in monocular sensitivities observed in cortical neurons (the simplest being ocular dominance) which have thus far been a well-established phenomenon lacking any clear functionality. Together, this framework shows that even visual perception, long taken as a model system for its apparently simple stages of image formation and transduction, involves idiosyncratic encoding which is shaped by geometric projection at the earliest stages of stimulation upon the sensory epithelium. It therefore requires corresponding non-canonical decoding mechanisms downstream, in the service of reconstructing the 3D environment well enough to inform perception and guide action.

Results

We developed a computational model of MT responses to motion which incorporates the geometric relationship between the world and the two retinae, acknowledging the fact that retinal stimulation is the result of light bouncing off objects and surfaces in the 3D environment and being projected through the pupil onto the back of the eyes. This is distinct from earlier work which often assumed that visual patterns presented on flat screens in front of a subject were a sufficiently complete proxy for the dynamic patterns of stimulation that fall on the retinae. Our model starts with environmental representations of object motion, works through the projective geometry upon both of the retinae, and then takes known responses to monocular velocities and binocular combination into account. This predicts non-canonical tuning forms for single neuron encoding of 3D direction and correspondingly non-homogeneous estimation performance when decoding from populations with these tuning functions.

Highly atypical tuning structure for 3D environmental velocities in macaque MT

Recent work across electrophysiology and fMRI has implicated MT in the processing of motion off the frontoparallel plane [3, 4, 5]. Here we perform a closer examination of neural recordings in macaque MT in order to characterize the functional form of tuning for 3D motion direction (specifically xz -motion directions, see Figure 1a). The black points in Figure 1b show the measured tuning curves (i.e., the average neural response) to the presentation of different 3D motion directions (i.e., motion on the xz -plane) for six example neurons. The firing rate of these neurons is modulated by changes in 3D motion direction. Notice that the tuning curves are characterized by steep transitions in four locations on the motion direction axis (roughly between each pair of adjacent cardinal directions; right, away, left, toward) with relatively little change in firing everywhere else. Given that the vast majority of tuning to simple sensory features takes on a Gaussian form [6, 7, 8], including MT responses to frontoparallel directions of motion, this at first glance appears to be a bizarrely “terraced” tuning structure. However, we can explain this tuning structure by considering the relationship between 3D environmental velocities and the resulting velocities which fall on the retina.

Atypical tuning structure for 3D environmental velocities is predicted by a model which incorporates environment-to-retina geometry

To understand the non-canonical 3D tuning curves, we developed a model for encoding 3D motion which incorporates the projective geometry from the environment onto the two retinæ. It then applies the canonical log-Gaussian tuning of MT neurons to the pair of retinal velocities that correspond to a particular 3D direction [9, 10], and takes the linear combination of those monocular responses. When a particular 3D motion direction is presented on the xz -plane at a given viewing distance, the geometric projection onto the retinæ results in separate left and right eye retinal velocities (Figure 2a–b). The direction and speed of the retinal velocities are dependent on the 3D motion’s environmental velocity as well as its distance to the eyes. Correspondingly, any egocentric representation of 3D motion direction along the xz -plane must consider the locations of the eyes, as well as the viewing distance.

Here, we use a coordinate system that is egocentric, in which the frontoparallel rightward-leftward and 3D towards-away motion axes of the xz -plane are always anchored to the cardinal axes (0 & 180 and 270 & 90, respectively). In the first part of the paper (including Figures 1–3), we employ a scaling that makes the ocular axes of the left and right eye (i.e., motions directly toward/away from either eye) orthogonal to one another; placing each midway between the 3D and frontoparallel axes. In effect, this also divides the space equally into regions with the same- or oppositely-signed motion in the two eyes. We begin with this representation because it matches prior work [11, 12, 4] and because the layout makes it very easy to examine the relationship between the motion in the environment and the motion that falls on the retinæ. This coordinate system can be interpreted in environmental terms as having an implausibly short effective viewing distance: half the average inter-pupillary

distance (~3.25cm in humans and ~1.63cm in macaques; see Figure 2a–b). We emphasize that these conventional axes are not based on environmental interpretations, but on the uniform sampling of monocular velocity ratios across the two eyes (as presented in Figure 2 of [12]). In subsequent sections, we consider more realistic viewing distances in the model and human behavior.

In the model, the projected retinal velocity in each eye produces a neural response derived directly from the monocular tuning curve (Figure 2 rows c–e, left panels); both of these monocular-velocity responses can then be replotted as functions of 3D motion direction (Figure 2 rows c–e, middle panels). The predicted binocular response is a linear combination of the corresponding monocular responses (Figure 2c–e, right panels; Online Methods equation 1). Combining binocular projective geometry and canonical tuning for retinal stimulation within a simple linear model results in tuning curves with abrupt discontinuities, characterized by multiple plateaus separated by steep cliffs (e.g., Figure 2 rows c–e, right panels). This shape deviates drastically from the classical smooth unimodal (i.e., bell-shaped) tuning observed across virtually all sensory features and systems [6, 7, 8]. Though atypical in appearance, the model tuning curves bear a striking resemblance to MT responses to binocular 3D motion stimulation, capturing the qualitative deviations from bell-shaped tuning curves ([4]; e.g., Figure 1, black curves).

Given the qualitative success of this model, we further quantified its ability to describe a full electrophysiological data set ($n=236$ neurons; 4500 responses collected per neuron) collected in macaque MT ([4]; see also Online Methods). We predicted the binocular response to 3D motion direction by summing the average monocular responses to the corresponding retinal velocities (see Online Methods equation 1, with $c_L, c_R = 1$). Note that this is a *parameter-free* prediction of the neural response to 3D motion direction. Relying solely on the geometric transformations from environment to retinae (and the assumption of inter-ocular additivity), the model accounts for 76% of the variance in the data (187-of-236 units with 50% of the variance explained; median root mean-squared error 7.0 spikes/s). Fitting the binocular combination coefficients as free parameters (c_L, c_R in Online Methods equation 1 using least-squares and Monte Carlo cross-validation) results in a modest improvement to account for 82% of the variance in the data (190-of-236 units with 50% of the variance explained; median root mean-squared error 4.6 spikes/s).

Consider the sample tuning curves shown in Figure 1, now noting that the purple curves depict the predictions of our model using the fitted combination coefficients (purple dots/line). Supplementary figure 1 shows additional example fits for neurons which are well fit and poorly fit by this model. The von Mises tuning curve (i.e., circular normal) is the canonical tuning curve used to describe 2D direction tuning curves in MT. For comparison to our principled 3D model, we also fit von Mises tuning curves to the MT data, despite the fact that they lack the plateaus and cliffs evident in many of the neural tuning curves. The von Mises model explained 80% of the variance in the data (190-of-236 units with 50% of the variance explained; median root mean-squared error 4.9 spikes/s). A direct statistical model comparison using AIC and BIC analyses (see Online Methods) further supports the conclusion that the 3D encoding model performs better than the canonical von Mises (3D model: $AIC = 274$, 95% CI [197, 348], $BIC = 173$, 95% CI [173,320]; von Mises model:

$AIC = 380$, 95% CI [302, 482], $BIC = 427$, 95% CI [345, 530]; noting that, e.g. a difference in $BIC > 10$ corresponds to “very strong” evidence in favor of one model over another). It is also interesting to note that the performance of the two models is not uniform over the different motion directions presented. This is related to the von Mises model’s failure to capture the qualitative shape (cliffs/plateaus) of many of the neurons we observed. In particular, Supplementary figure 2 shows that the 3D model is better at capturing the responses to toward and away motions.

Although the 3D model was quantitatively superior to the descriptive bell-shaped fits from a conventional (von Mises) function, the most important differences between these two models are largely qualitative in nature. The 3D encoding model can capture the qualitative shapes of many of the neurons we observed, including the abrupt cliffs and the long plateaus in the tuning curves. Furthermore, the 3D model directly implements binocular combination and suggests explicit mechanisms for the construction 3D motion direction tuning. In contrast, the von Mises tuning model is purely descriptive. Thus the 3D encoding model is a better model quantitatively, qualitatively, and mechanistically. In the upcoming sections, we also show that the 3D encoding model makes predictions that are consistent with behavior, but which are not explained by a von Mises tuning model.

Estimating 3D motion direction from this atypical tuning structure reveals a sufficient but idiosyncratic encoding

Our model’s success describing MT raises the question of whether a population with such idiosyncratic tuning curves could be used to estimate 3D direction (Figure 3a). To investigate this, we built a population based on model fits to the neurons recorded in [4], assuming Poisson output noise (Online Methods equations 1–3). We simulated population responses to motion (5cm/s) around the xz-plane of 3D directions sampled at 1 degree intervals, at a viewing distance of $\frac{\text{interpupillary distance}}{2}$, and used a standard maximum log-likelihood decoder to estimate the 3D velocity (direction, speed) from the resulting population response (e.g., [13]; Online Methods equations 11–12). Despite the unconventional encoding of 3D directions, this decoder successfully recovered 3D motion direction (Figure 3b; estimates near the unity line).

The unusual structure of the 3D direction encoding has important ramifications because the underlying tuning curves do not represent changes in 3D direction with equal fidelity. This is evident in how decoding performance varies as a function of the true motion direction (see Figure 3c), with regions of higher precision near the steepest portions of the tuning curves. These regions correspond to what we deem the “ocular axes”, which are the directions for which the retinal velocities flip sign when the 3D direction changes. When an object’s 3D direction is very close to moving directly toward one of the eyes, small changes in 3D direction can correspond to categorical (direction) changes upon that retina. Because MT neurons respond more strongly to one direction than another, these direction changes in one eye give rise to the steep transitions present in the binocular tuning curves. The resulting heterogeneous pattern of precision is notably distinct from decoding based on canonical

tuning (Figure 3d–f; Supplementary figure 3), which predicts consistent estimation error across all values of a stimulus feature.

3D direction tuning depends on environmental position

The central contribution of our model to the existing understanding of motion processing is the incorporation of the environment-to-retina projection geometry, which results in tuning that is expressed with respect to the environment. An important consequence of this is that the tuning structures are location dependent, a factor which has also been ignored in standard “retinocentric” models of MT and direction selectivity. Because our model includes viewing distance as a parameter, we can simulate and decode at multiple, more realistic viewing distances using the same model population, Poisson output noise, and maximum log-likelihood decoder. Figure 4 shows the dramatic effect of viewing distance. Viewing distance changes the retinal projections (see figure 4), which affects the shape of individual tuning curves (4c compared to 3a), and drastically changes model decoding performance (see figure 4d compared to 3b). At a further (and more perceptually realistic) viewing distance, the systematic biases and errors of the model estimation results reveal two notable features: (1) a coarse-scale ‘X’ pattern indicating errors that are orthogonal to the line of unity (which delineates perfectly accurate estimation), and (2) square structures in the clouds of points which reflect finer-scale deviations from unity (see also, [14, 15]). These patterns can be thought of as depth-sign errors and a bounded bias away from frontoparallel motion, respectively (see figure 4d). Next, we describe how both of these initially perplexing patterns of errors are understandable consequences of environmentally-referenced decoding that are already evident in the behavior of our simple model.

The reason for the rather striking depth-sign error is related to the geometric consequences of viewing distance. As viewing distance increases, retinal velocities decrease, and the angle between the visual axes of the two eyes decreases (i.e., there is a reduced phase shift in the environment-to-retinal velocity mappings between the two eyes; compare Figure 4a at a 67cm viewing distance to Figure 2b, left panel at $\frac{1}{2}$ **interpupillary distance* or a 3.25cm viewing distance for a human). For a fixed environmental velocity, any single tuning curve is dependent on the resulting retinal velocities, and thus on viewing distance. At larger viewing distances, the steep transitions of binocular tuning curves shift closer to the toward and away environmental motion directions (Figure 4b–c); resulting in a more symmetrical tuning curve (see symmetry line Figure 4b). The depth-sign errors are due to this increasing symmetry across the neural representation in the presence of noise. Note the approximate mirror symmetry in Figure 4c, with lines of symmetry at *left* (←) and *right* (→).

In addition to the depth-sign errors, a subtler but equally telling idiosyncrasy is present in the form of systematic bias of estimates away from purely frontoparallel. This is easiest to see in the roughly square cloud of points in the center of Figure 4d: when a leftward direction was presented (middle of the x-axis), estimates (y-axis) are repulsed from frontoparallel, but cannot be mistaken for motion containing a rightward component (i.e. the decoder does not make x-axis sign flip errors.) Thus, the estimates are bounded at the toward and away

directions (evident in the horizontal bands at the top and bottom edges of that central square). Analogous patterns for rightward motion are present in the corners.

This “frontoparallel repulsion” is also explainable by our model, and is a distinct consequence of the same underlying dependence on retinal velocities in the encoding scheme. Figure 5a–d shows model estimates at different viewing distances (3.25cm, 20cm, 31cm, and 67cm), color coded by their corresponding environmental speed estimate. The systematic bias for toward/away motion at the farthest viewing distance is lawfully related to a systematic overestimation of environmental motion speed: for a perfectly-frontoparallel estimate to be generated, the monocular velocities would have to match exactly. But given noisy monocular estimates, the resulting 3D direction estimates will be repulsed from frontoparallel, either towards or away, depending on which monocular channel’s noise yielded a larger/smaller response. More detailed examination of the corresponding monocular velocities reveals that variability of the monocular velocity estimates roughly follows Weber’s law, regardless of the viewing distance (see Figure 5e–h). However, different viewing distances result in a different mapping between retinal velocities and environmental motion (see Figure 5i–l, and equations 4–5 which are dependent on viewing distance, z). Thus, at far viewing distances the same variability plays out as a larger systematic bias for the model: estimates of motion that are too fast and too close to the toward and away directions.

Although this explication of the model builds intuition for these errors in the decoder’s performance, it may seem unreasonable to predict that human observers would exhibit these patterns of performance and particularly that they would make the same depth-sign and frontoparallel-repulsion errors as this model decoder. However, existing psychophysical results have established that humans do make depth-sign errors [16], and in the next section we not only confirm the existence of both depth-sign and frontoparallel-repulsion errors, but show that these perceptual distortions emerge and obey the quantitative functional dependence on position in the environment (i.e., viewing distance) implied by our model.

Human performance on a 3D motion direction estimation task exhibits the signatures of the proposed environment-to-retina model of 3D motion tuning

We tested whether human perception exhibits signatures of the environment-to-retina encoding-decoding model: performance in 3D direction estimation should be a function of both motion direction and location/viewing distance. We designed a perceptual experiment to examine human 3D motion direction estimation at several viewing distances. Observers estimated the 3D motion direction of random dots within a 3D spherical volume (5 degrees in frontoparallel diameter; at 5% contrast; rendered with looming and expansion cues; motion direction at $0^\circ, 5^\circ, \dots$ or 355° on the xz -plane; with a motion speed of 5cm/s) at three different viewing distances (20cm, 31cm, or 67cm). These 3D motion volumes are analogous to the classical 2D motion apertures found in classic studies of 2D motion. Motion was presented for 1 second and observers reported their estimate of the dots’ 3D motion direction using a knob to adjust the angle of a stereoscopically-rendered indicator on

the screen. Supplementary Video 1 and Supplementary Video 2 provide high contrast examples of the motion stimuli.

Figure 6 shows estimation performance at three different viewing distances collapsed across three observers (top row) and model performance at the same three viewing distances for comparison (bottom row). Human observers did exhibit depth-sign errors and biases for toward/away motion that fully emerge as a function of viewing distance. (Performance of individual subjects is shown in Supplementary figure 4.) Panel g of Figure 6 illustrates the increase in depth-sign errors with increased viewing distance and compares performance to the the predictions of the 3D model and the von Mises model. While there are almost no depth-sign errors predicted by the von Mises model, the 3D model predictions increase in step with the psychophysical results.

Subtle tuning differences across the two eyes enable the toward-vs-away aspect of decoding for 3D direction

By separately manipulating parameters of the simulated population (see Online Methods eq. 1 – 3), we were able to examine which aspects of neural tuning in MT neurons affect estimation of 3D motion direction. For example, a population with identical monocular tuning parameters (i.e., the same speed preference, tuning bandwidth, response amplitude, and baseline firing rate in the two eyes), correctly identifies the x-component – the frontoparallel component – of 3D motion. But, in the absence of any implicit eye of origin signatures playing out in such parameters, this “equal-monocular” encoding cannot recover the direction for the depth component above chance levels because there is no differentiating information for toward versus away motion components (see Figure 7b).

However, merely incorporating subtly differential monocular tuning (at the levels measured in [4]) reveals that small, seemingly trivial differences in response amplitude, tuning bandwidth, or speed preference between the two eyes are each in principle sufficient for representing 3D motion direction (see Figure 7c–e respectively and Figure 7a for comparison). Differences in untuned components–such as the baseline firing rate from the two monocular response components– do not provide differential toward/away information (see Figure 7f). Thus, small and seemingly innocuous mismatches between left and right eye tuning may play a key role in encoding the 3D environment. In particular, we note that small differences in response amplitude are more commonly called “ocular dominance”, a phenomenon that has been well-documented in visual cortex [17, 18], but has rarely been posited as a scheme for carrying information [19]. This theoretical finding points to a potentially important role for these subtle ocular imbalances in visual processing.

Discussion

We have introduced a framework for making inferences about environmental properties given knowledge of the neural sensitivity to features of retinal stimulation. Specifically, we examined tuning for 3D motion in primate MT, and observed atypical tuning structures for 3D motion. We found that an encoding model that combines the relationship between the environment and the retina with the known retinal encoding of 2D motion explains this

strikingly atypical tuning structure. This encoding model was then shown to be sufficient for estimating 3D motion direction. Furthermore, a decoding analysis predicted 3D motion direction estimation performance that varies as a function of motion direction and location/distance, which we showed is consistent with human perceptual judgements and is in stark contrast with default (Gaussian/von Mises) tuning models that have homogeneous sensitivity across all 3D directions. Thus, the predictions made by extending sensory encoding and decoding to incorporate the geometry of the spatiotemporal environment naturally account for what are, at first glance, rather odd aspects of both neural tuning curves and human perception.

Previous work in the perceptual literature has reported the frontoparallel bias and depth-sign errors that we observe to be prevalent at longer viewing distances [20, 16]. Bayesian observer models that rely on slow speed priors have provided plausible explanations for the set of biases and errors observed in human perceptual experiments [21, 14, 15]. The use of binocular velocities for 3D motion direction discrimination/estimation was also proposed by Beverley and Regan [22], with supporting psychophysical experiments which tested 3D motion direction discrimination and demonstrated increased direction sensitivity in line with the location of the two eyes. Our model provides a more complete explanation in three important ways: the model is built upon the tuning structure of a known neural population, the model does not need to invoke a prior, and the model makes explicit the location-dependent nature of primate 3D motion direction estimation (i.e. how performance changes with viewing distance).

Previous work in the electrophysiological literature established that MT neurons with some '3D tuning' (as defined by a preferred direction calculated using a vector average of responses) were more likely to exhibit nonlinear binocular summation [4]. They concluded that these nonlinearities were likely critical for 3D motion sensitivity. Despite the fact that the binocular combination included in our model is purely linear and does not take into account these nonlinearities, our model accounts for over half the variance in most neurons. We do find that there are some neurons are not well-fit by our model. This is at least partially due to nonlinearities in binocular combination, which likely sharpen 3D motion sensitivity. However, the theoretical exercise described in this manuscript reveals fundamental contributions of binocular projection geometry and ocular imbalance that give rise to the non-canonical tuning structures observed in MT.

The model of 3D motion tuning proposed here examines how 3D motion information can be read out from the different retinal velocities which fall on the two eyes. The field has named this binocular information about 3D motion: interocular velocity differences (IOVDs) [23, 24]. Given that the model presented here relies on binocular summation and ocular tuning imbalances across the two eyes, the term interocular velocity differences is a bit of a misnomer and potentially confusing (which is why we have avoided mentioning it). The mechanism representing this type of information does not engage in any differencing per se, though it does rely on the fact that the velocities are different.

The work presented here provides a phenomenologically compelling model of the representation of 3D direction, supported by both electrophysiological and psychophysical

evidence. However, future work will need to examine more directly the relationship between physiology and perception in awake behaving primates using tools such as micro-stimulation (e.g. [25]). Such experiments will also provide an important opportunity to further characterize monocular and binocular tuning characteristics of neurons, as well as potential dependencies on viewing distance, in order to test and refine the model proposed here.

In conclusion, our findings emphasize the importance of recognizing the nervous system's ultimate need to infer the properties of the environment to guide behavior. Such inference is based on sensory information that is fundamentally constrained by the geometric relationship between the environment and the sensory organ. We considered the case of 3D motion direction as an example, demonstrating that a geometrically-constrained encoding model for 3D motion direction is consistent with electrophysiological recordings of neurons in MT and human performance on direction estimation tasks. Furthermore, we found evidence that small differences in tuning across the two eyes can support 3D motion direction estimation. The geometric framework presented here can be applied to other visual features. For example, slanted and tilted patterns project differential patterns of orientation upon the two retina, which shape the environmental meaning of canonical orientation tuning functions. Thus, a large number of important cortical encoding modules may not be implemented by banks of units with bell-shaped tuning when the decoding of environmental properties (rather than retinal image properties) is required, as is the case for visually-guided behaviors in the natural world.

Data/Code Availability

The modeling code/simulations and the human psychophysical data/analysis is available here: <https://github.com/kbonnen/BinocularViewing3dMotion>.

Online Methods

Electrophysiological Data

Several analyses performed in this paper rely on an electrophysiological data set (n=236) collected in the middle temporal area of 2 adult male macaque (*macaca fascicularis*, age 3 and 4 years) under anesthesia by [4]. These recordings include the neural responses to 3D motion in 28 directions on the xz-plane (with varying environmental speeds; fully crossed manipulation of retinal velocities in the two eyes: $-10^\circ/s$, $-2^\circ/s$, $-1^\circ/s$, $1^\circ/s$, $2^\circ/s$, $10^\circ/s$), as well as the responses to the corresponding monocular velocities. The stimulus was constructed using drifting gratings at 6 different orientations (0° , 30° , 60° , 90° , 120° , 150°), all drifting orthogonal to grating orientation. Each stimulus was repeated 25 times. For the purposes of our analyses we included all data except those collected using the horizontally-oriented grating, which doesn't have a proper binocular velocity signal. Additional details about these experiments can be found in the original paper

Computational Model

Encoding.

Here we describe the single neuron encoding model used to generate the model predictions for responses to 3D motion direction (e.g., Figures 3–7). 3D motion refers to motion on the xz -plane (see Figure 1a). A velocity on this plane is specified by (θ, m) where θ is the xz -direction (deg.) and m is the magnitude of the motion (cm/s). The binocular response function for 3D motion, $f_B(\theta, m)$ (e.g., Figure 2c–e, right panel), can be written as a weighted combination of the monocular responses due to the retinal velocities that fall onto each of the eyes:

$$f_B(\theta, m) = c_L * f_L(\theta, m) + c_R * f_R(\theta, m) \quad (1)$$

where $f_L(\theta, m)$ and $f_R(\theta, m)$ are the monocular responses (spike rate; e.g., Figure 2c–e, i–ii) to the corresponding left and right eye retinal velocities (see Figure 2b); c_L and c_R are the coefficients for linear combination. These combination coefficients allow for suppression or amplification of one or both eyes during the binocular response.

Monocular velocity tuning curves in MT are well-fit by log-Gaussian functions [10] and thus we parameterize the monocular response functions ($f_L(\theta, m)$, $f_R(\theta, m)$) using log-Gaussian curves (e.g., Figure 2c–e, i). The motion confined to the xz -plane gives rise to monocular velocities to the right or left at different speeds. Because MT neurons exhibit diversity in their direction selectivity, the log-Gaussian function must be simultaneously fit to both directions with coefficients to modulate the relative amplitude of the neural response:

$$f_L(\theta, m) = \begin{cases} \frac{a_{L+}}{d\theta_L(\theta, m)\sigma_l} e^{-\frac{(\log d\theta_L(\theta, m) - \mu_l)^2}{2\sigma_l^2}} + b_L & d\theta_L(\theta, m) \geq 0 \\ \frac{a_{L-}}{|d\theta_L(\theta, m)|\sigma_l} e^{-\frac{(\log |d\theta_L(\theta, m)| - \mu_l)^2}{2\sigma_l^2}} + b_L & d\theta_L(\theta, m) < 0 \end{cases} \quad (2)$$

$$f_R(\theta, m) = \begin{cases} \frac{a_{R+}}{d\theta_R(\theta, m)\sigma_r} e^{-\frac{(\log d\theta_R(\theta, m) - \mu_r)^2}{2\sigma_r^2}} + b_R & d\theta_R(\theta, m) \geq 0 \\ \frac{a_{R-}}{|d\theta_R(\theta, m)|\sigma_r} e^{-\frac{(\log |d\theta_R(\theta, m)| - \mu_r)^2}{2\sigma_r^2}} + b_R & d\theta_R(\theta, m) < 0 \end{cases} \quad (3)$$

where μ_L , σ_L , μ_R , and σ_R are the parameters of the log-Gaussian function; a_{L+} , a_{L-} , a_{R+} , and a_{R-} are the coefficients modulating the relative amplitude of the neural response; b_L and b_R are the baseline firing rates; $d\theta_L(\theta, m)$ and $d\theta_R(\theta, m)$ are functions that give the retinal velocities for the left and right eyes respectively (see below, see also Figure 2a), given the xz velocity (θ, m) .

$$d\theta_L(\theta, m) = \frac{\cos(\theta) * m * z - \sin(\theta) * m * (x + \frac{ipd}{2})}{(x + \frac{ipd}{2})^2 + z^2} \quad (4)$$

$$d\theta_R(\theta, m) = \frac{\cos(\theta) * m * z - \sin(\theta) * m * (x - \frac{ipd}{2})}{(x - \frac{ipd}{2})^2 + z^2} \quad (5)$$

where (x, z) is the motion location (cm) and ipd is the inter-pupillary distance (6.5cm in humans, 3.25 in macaque).

The equation for the retinal velocities ($d\theta_L, d\theta_R$) given an environmental velocity (θ, m) comes from taking the derivative on the angular relationship between the eye in question and the motion location (for schematic, see Supplementary figure 5):

$$\tan(\theta_r) = -\frac{z}{x - \frac{ipd}{2}} \quad (6)$$

$$\theta_r = \tan^{-1}\left(-\frac{z}{x - \frac{ipd}{2}}\right) \quad (7)$$

To find the velocity for the right eye, take the derivative (i.e. $d\theta_r$; note that

$$\frac{d}{dx}\tan^{-1}(f(x)) = \frac{1}{1 + f(x)^2} * f'(x):$$

$$d\theta_r = \frac{1}{1 + \left(\frac{z}{x - \frac{ipd}{2}}\right)^2} * \left(z * dx * \left(x - \frac{ipd}{2}\right)^{-2} - dz * \left(x - \frac{ipd}{2}\right)^{-1}\right) \quad (8)$$

$$d\theta_r = \frac{\left(x - \frac{ipd}{2}\right)^2}{\left(x - \frac{ipd}{2}\right)^2 + z^2} * \left(\frac{z * dx}{\left(x - \frac{ipd}{2}\right)^2} - \frac{dz}{\left(x - \frac{ipd}{2}\right)}\right) \quad (9)$$

$$d\theta_r = \frac{z * dx - dz * \left(x - \frac{ipd}{2}\right)}{\left(x - \frac{ipd}{2}\right)^2 + z^2} \quad (10)$$

Substituting dx and dz for $\cos(\theta) * m$ and $\sin(\theta) * m$ respectively gives equation 5 above. The derivation for $d\theta_L$ follows the same logic except the location of the eye has changed (i.e. $\left[x - \frac{ipd}{2}\right] \rightarrow \left[x + \frac{ipd}{2}\right]$).

Decoding.

3D motion direction estimation was performed by finding the xz -velocity (θ, m) associated with the maximum log-likelihood value, given the assumption of independent Poisson noise on the 3D binocular tuning curve:

$$\log L(\theta, m) = \log \left(\prod_{i=1}^N p(\mathbf{r}_i | \theta, m) \right) = \sum_{i=1}^N \log \left(\frac{f_{B_i}(\theta, m)^{\mathbf{r}_i} e^{-f_{B_i}(\theta, m)}}{\mathbf{r}_i!} \right) \quad (11)$$

$$= \sum_{i=1}^N \log (f_{B_i}(\theta, m)) r_i - \sum_{i=1}^N f_{B_i}(\theta, m) - \sum_{i=1}^N \log(r_i!) \quad (12)$$

where \mathbf{r} is the population response, a vector composed of the spike count for N neurons; and f_B are the binocular tuning curves for 3D motion (see [13]). Motion direction and magnitude were jointly estimated by maximizing the log-likelihood: $\operatorname{argmax}_{\theta, m} \log L(\theta, m)$.

von Mises model.

Here we describe a double von Mises encoding model for single neurons. This model produces classical bell-shaped tuning for 3D direction and allows for two peaks of different amplitudes separated by 180° . It is the model used in the comparison in Figure 3, Supplementary figures 1, 2, and 3. The response function (f_{von}) is given by the following equation:

$$f_{von}(\theta) = a_1 \frac{e^{K * \cos(\theta - \mu)}}{2\pi * I_0(K)} + a_2 \frac{e^{K * \cos(\theta - \mu - \pi)}}{2\pi * I_0(K)} + b \quad (13)$$

where μ is the preferred direction of the neuron, K is a measure of concentration (analogous to $\frac{1}{\sigma^2}$), b is the baseline firing rate and a_1, a_2 control the relative amplitudes of the preferred and anti-preferred directions (allowing for the type of mixed direction selectivity typically reported in MT tuning for 2D – xy – motion direction).

Model Comparison

The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were calculated for fits of both the von Mises and 3D encoding model to the binocular response data from [4]. AIC and BIC are designed for comparisons of models with differing numbers of parameters. The 3D model is a 2 parameter model given by equation 1 where c_L, c_R are the parameters and $f_L(\theta, m), f_R(\theta, m)$ are given by the monocular data. The von Mises model is the 5 parameter model as shown in equation 13, but is actually a 25 parameter model because a different 5 parameters must be learned for each of the five grating orientations used in the analysis. We performed Monte Carlo cross-validation ($n=50$) to estimate AIC and BIC for each neuron and then took the mean across all neurons to calculate the population AIC and BIC for both models.

Psychophysical Methods

Observers.

Data were collected from three psychophysical observers (including two of the authors and one naive subject; ages 20–28 yrs; 1 female and two male). Each of the observers had good stereopsis and normal or corrected-to-normal vision. All observers participated with written informed consent and were treated according to the principles set forth in the Declaration of Helsinki of the World Medical Association. All procedures were approved by the University of Texas at Austin Institutional Review Board.

Apparatus.

Stimuli were presented stereoscopically using a ProPixx 3D projector (120Hz per eye, 74.5cm x 132.5cm; St. Bruno, Canada) and a Screen Tech ST-PRO-DCF black acrylic glass screen (Hamburg, Germany). We designed a rail system for mounting both the screen and projector that can be easily adjusted to viewing distances from 20cm to 120cm without moving the subject. Supplementary figure 6 shows a schematic of this system.

Stimulus.

The stimuli were fixed spherical dot motion volumes analogous to the dot motion apertures in the classical (frontoparallel) motion literature. Supplementary Videos 1–2 show high contrast examples of this stimulus. Supplementary Video 1 can be free-fused and is a high-contrast version of the stimulus shown to subjects during our experiments. Supplementary Video 2 is a 2D video rendered with shading on the dots to give a stronger sense of the depth percept. Both videos show 8 motion epochs (0° , 45° , 90° , 135° , 180° , 225° , 270° , 315°). Spherical motion volumes were 5° (frontoparallel; 1.78cm, 2.75cm, and 5.95cm at the 3 viewing distances) in diameter at 5° eccentricity left or right from fixation. In order to avoid performing experiments in a stereomotion scotoma [26], stereomotion tests were performed at both locations and the stimulus was placed in the location with highest performance. Dots within the spherical volume were at 5% contrast (half with luminance above the background luminance and half with luminance below), moving at one of three speeds (5cm/s 7.75cm/s or 16.75cm/s), in one of 72 directions (0° , 5° , 10° , ... 350° , 355°), at one of three viewing distances (20cm, 31cm or 67cm), rendered with looming and expansion cues.

Procedure.

Each trial consisted of a motion epoch lasting one second. Subjects reported the motion direction of the dots using a knob to adjust the angle of an indicator on the screen. The indicator was rendered stereoscopically and consisted of a vector arrow that could be oriented radially around a ring on the xz-plane. This indicator was presented slightly below the location of fixation from the motion epochs. An example of the motion indicator used during the experiment can be seen in Supplementary Video 2. It is the figure below the motion cloud which is indicating the direction of motion of the dot cloud. The experiment was completed in blocks. Each block consisted of 72 trials at single viewing distance, with pseudo-randomly interleaved trials of different speeds and directions (the fastest speed, 16.75cm/s was presented only at the farthest viewing distance). Each condition (direction/

speed/viewing distance combination) was repeated 5 times. The experiment was conducted in 35 blocks for a total of 2520 trials.

Statistics.

No statistical methods were used to predetermine sample sizes but our sample similar to those reported in previous publications [20, 16]. The psychophysical experiment was completed in blocks. Within blocks, trials with different speeds and directions were interleaved. Blocks were performed in a random order. Between blocks the screen was set at the appropriate distance for the upcoming trials. Because participants were aware that the screen was at different viewing distances, data collection and analysis was not performed blind to the conditions of the experiments. In this paper we present the data from the 5cm/s trials, since that speed was presented at all 3 viewing distances. Otherwise, no data was excluded. We did not explicitly test for normality.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded by the National Eye Institute at the National Institutes of Health (EY020592, LKC, ACH, AK), the National Science Foundation (DGE-1110007, KB), the Harrington Fellowship program (KB), and the National Institutes of Health (T32 EY21462-6, TBC, JAW). Special thanks to Jacqueline Fulvio and Bas Rokers for many insightful conversations and for sharing their psychophysical data early on in this project.

References

- [1]. von Helmholtz Hermann. *Treatise on Physiological Optics*. 1867.
- [2]. Gibson James Jerome. *The Perception of the Visual World*, 1950.
- [3]. Rokers Bas, Cormack Lawrence K, and Huk Alexander C. Disparity- and velocity-based signals for three-dimensional motion perception in human MT+. *Nature Publishing Group*, 12 (8):1050–1055, 8 2009.
- [4]. Czuba Thaddeus B, Huk Alexander C, Cormack Lawrence K, and Kohn Adam. Area MT encodes three-dimensional motion. *The Journal of Neuroscience*, 34(47):15522–15533, 11 2014. [PubMed: 25411482]
- [5]. Sanada Takahisa M and DeAngelis Gregory C. Neural representation of motion-in-depth in area MT. *The Journal of Neuroscience*, 34(47):15508–15521, 11 2014. [PubMed: 25411481]
- [6]. Hubel DH and Wiesel TN. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148:574–591, 10 1959. [PubMed: 14403679]
- [7]. Bacon JP and Murphey RK. Receptive fields of cricket giant interneurons are related to their dendritic structure. *The Journal of physiology*, 352:601–623, 7 1984. [PubMed: 6747901]
- [8]. Jacobs GA and Theunissen FE. Functional organization of a neural map in the cricket cercal sensory system. *Journal of Neuroscience*, 16(2):769–784, 1 1996. [PubMed: 8551359]
- [9]. Maunsell JH and Van Essen DC. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, 49(5):1127–1147, 5 1983. [PubMed: 6864242]
- [10]. Nover Harris, Anderson Charles H, and DeAngelis Gregory C. A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *The Journal of Neuroscience*, 25(43):10049–10060, 10 2005. [PubMed: 16251454]

- [11]. Beverley KI and Regan D. Evidence for the existence of neural mechanisms selectively sensitive to the direction of movement in space. *The Journal of physiology*, 235(1):17–29, 11 1973. [PubMed: 4778134]
- [12]. Cynader M and Regan D. Neurons in cat parastriate cortex sensitive to the direction of motion in three-dimensional space. *The Journal of physiology*, 274(1):549–569, 1 1978. [PubMed: 625008]
- [13]. Graf Arnulf B A, Kohn Adam, Jazayeri Mehrdad, and Movshon J Anthony. Decoding the activity of neuronal populations in macaque primary visual cortex. In *Nature Neuroscience*, pages 239–247. New York University, New York, United States, 2 2011.
- [14]. Cooper Emily A, van Ginkel Marcus, and Rokers Bas. Sensitivity and bias in the discrimination of two-dimensional and three-dimensional motion direction. *Journal of Vision*, 16(10):5–11, 8 2016.
- [15]. Rokers Bas, Fulvio Jacqueline M, Pillow Jonathan W, and Cooper Emily A. Systematic misperceptions of 3-D motion explained by Bayesian inference. *Journal of Vision*, 18(3): 23–23, 3 2018.
- [16]. Fulvio Jacqueline M, Rosen Monica L, and Rokers Bas. Sensory uncertainty leads to systematic misperception of the direction of motion in depth. *Attention Perception & Psychophysics*, 77 (5):1685–1696, 7 2015.
- [17]. Hubel DH and Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–154, 1 1962. [PubMed: 14449617]
- [18]. Katz Lawrence C and Crowley Justin C. Development of cortical circuits: lessons from ocular dominance columns. *Nature Reviews Neuroscience*, 3(1):34–42, 1 2002. [PubMed: 11823803]
- [19]. Lehky Sidney R. Unmixing binocular signals. *Frontiers in human neuroscience*, 5:78, 2011. [PubMed: 21886618]
- [20]. Welchman Andrew E, Val L Tuck, and Harris Julie M. Human observers are biased in judging the angular approach of a projectile. *Vision research*, 44(17):2027–2042, 8 2004. [PubMed: 15149836]
- [21]. Welchman Andrew E, Lam Judith M, and Bulthoff Heinrich H. Bayesian motion estimation accounts for a surprising bias in 3D vision. *Proceedings of the National Academy of Sciences of the United States of America*, 105(33):12087–12092, August 2008. [PubMed: 18697948]
- [22]. Beverley KI and Regan D. The relation between discrimination and sensitivity in the perception of motion in depth. *The Journal of physiology*, 249(2):387–398, 7 1975. [PubMed: 1177097]
- [23]. Cumming BG and Parker AJ. Binocular mechanisms for detecting motion-in-depth. *Vision research*, 34(4):483–495, 2 1994. [PubMed: 8303832]
- [24]. Harris Julie M, Nefs Harold T, and Grafton Catherine E. Binocular vision and motion-in-depth. *Spatial vision*, 21(6):531–547, 2008. [PubMed: 19017481]
- [25]. Salzman CD, Britten KH, and Newsome WT. Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 1990.

Methods References

- [26]. Barendregt Martijn, Dumoulin Serge O, and Rokers Bas. Stereomotion scotomas occur after binocular combination. *Vision research*, 105:92–99, 12 2014. [PubMed: 25286416]

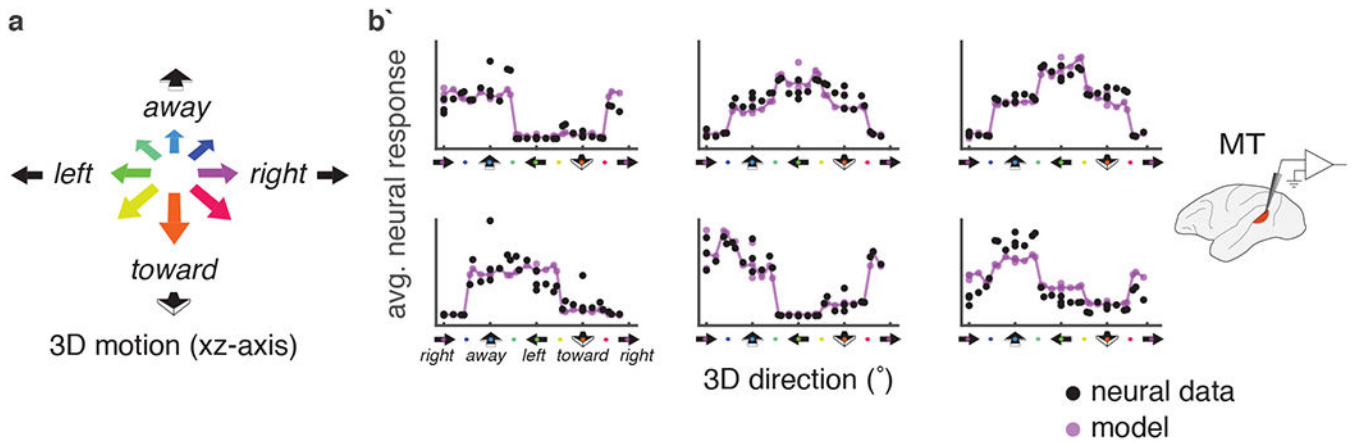


Figure 1: MT neurons exhibit atypical “terraced” tuning structure for environmental velocities in 3D.

a. For the purposes of this study, 3D motion refers to velocities which fall on the xz plane. This allows us to unwrap the motion direction onto a linear axis (as is typically done with frontoparallel motion): right, away, left, toward, right. **b.** Average neural response to 3D (xz) motion direction for 6 example neurons in macaque MT [4]. Each panel depicts the average response of a single example neuron to the presentation of different 3D motion directions (black dots). Predictions of the model proposed here are plotted for comparison (purple). Stimuli consisted of binocular presentations of motions consistent with a wide array of directions in the x-z axes (fully crossed manipulation of retinal velocities in the two eyes: $-10^\circ/\text{s}$, $-2^\circ/\text{s}$, $-1^\circ/\text{s}$, $1^\circ/\text{s}$, $2^\circ/\text{s}$, $10^\circ/\text{s}$). This results in motions presented in 28 unique directions (of varying environmental speeds), with each of the three cardinal directions (right, away, left, toward) repeated at 3 different speeds. These motion stimuli were presented at 6 different grating orientations (0° , 30° , 60° , 90° , 120° , 150°), all drifting orthogonal to grating orientation. Each stimulus was repeated 25 times. In the examples here, we have plotted the data from the vertically-oriented grating orientation. For the purposes of our analyses we included all data except those collected using the horizontally-oriented grating, which doesn't have a proper binocular velocity signal. Additional details about these experiments can be found in the original paper. [4].

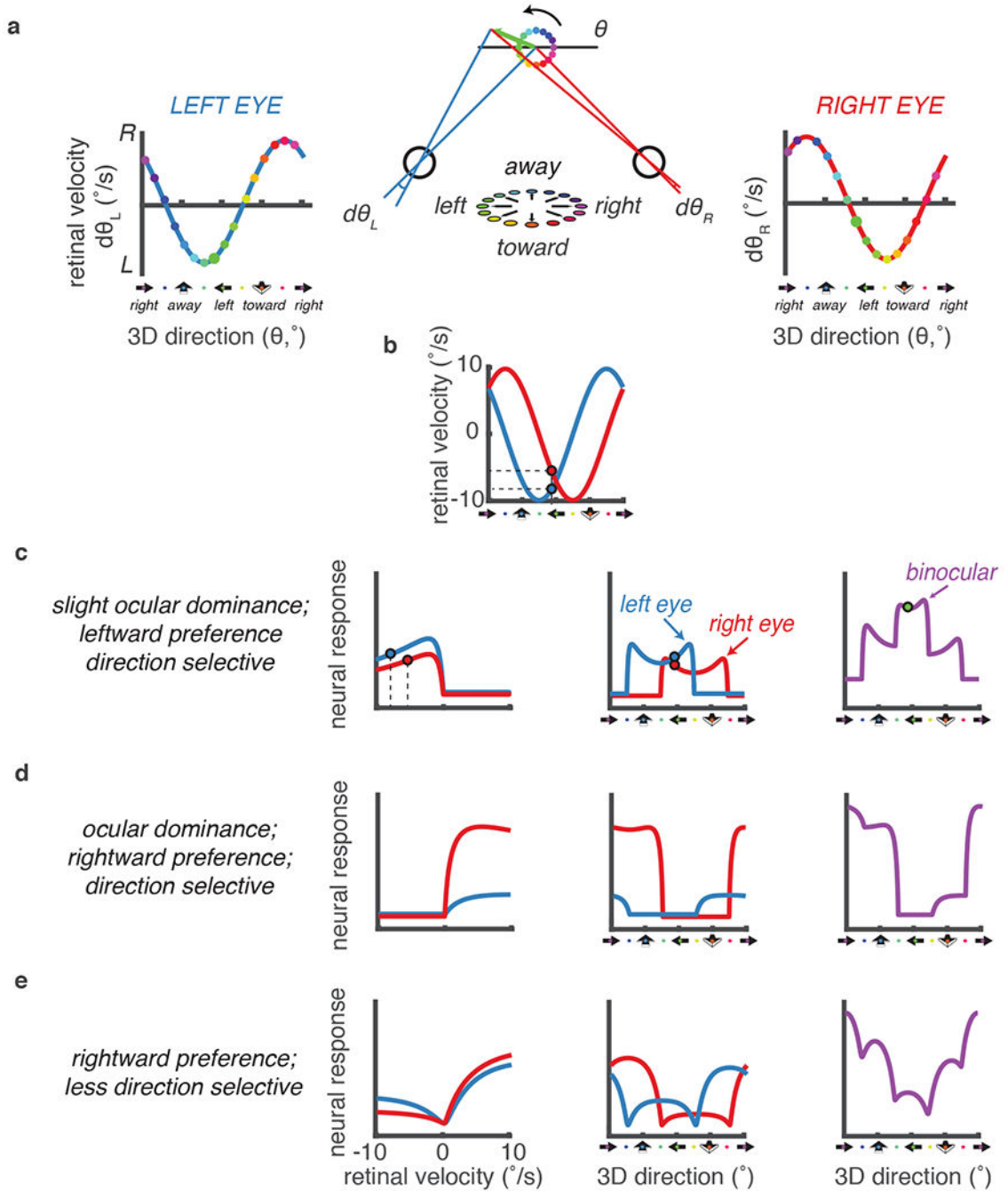


Figure 2: An encoding model that incorporates the environment-to-retina geometry of 3D motion predicts atypical structures for binocular 3D motion tuning curves.

a. Diagram of the projection of 3D motion (confined to the xz -plane; middle panel) onto the left eye (blue; left panel) and the right eye (red; right panel). The color wheels in the middle panel identify 16 xz -directions and those directions are also marked on the retinal velocity panels for the left and right eye. For simplicity, velocities are plotted in a world-motion reference frame, i.e., leftward motion in the world is also plotted as ‘leftward’ in retinal velocity panels. The assumption that the ocular axes are 90° apart results in an effective

viewing distance of $\frac{1}{2} * \textit{interpupillary distance}$. **b.** Left and right eye retinal velocities as a function of 3D motion direction. These are replotted from the left and right eye panels in **a.** **c-e.** Each row represents an example model neuron generated from fits to 3 neurons found in [4]. **c.** A 3D model neuron that exhibits slight ocular dominance, leftward preference, and is direction selective. (Left panel) Monocular retinal velocity tuning curves for the left and right eye. (Middle panel) Monocular neural responses as a function of 3D motion direction, built from the composition of the functions depicted in **b** and panel **i.** (Right panel) Binocular 3D motion direction tuning curve computed from a weighted linear combination of monocular responses in the middle panel. Data points (circles) trace the transformation of a single 3D direction from **b** through all three panels in **c.** **d.** A 3D model neuron that exhibits strong ocular dominance, rightward preference, and is direction selective. **e.** A 3D model neuron that exhibits rightward preference, and is less direction selective.

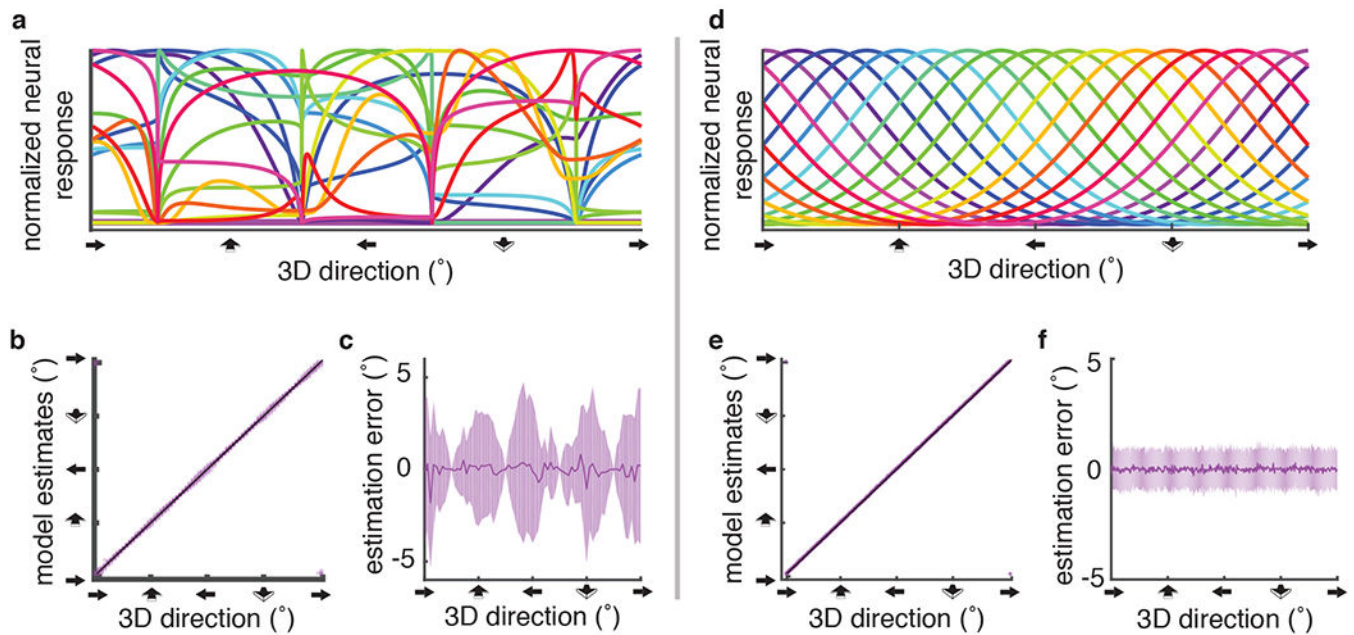


Figure 3: A 3D model decoder successfully estimates 3D motion direction.

However, the resulting pattern of estimates is distinct from an idealized Gaussian (von Mises) model. **a.** Binocular tuning curves from the computational model for decoding 3D motion direction, assuming a viewing distance of $\frac{1}{2}$ * interpupillary distance. These 16 example 3D direction tuning curves were chosen because their preferred direction (as calculated by the vector average) were closest to tiling 3D direction with 16 evenly spaced values in the xz-plane ($0^\circ, 22.5^\circ, 45^\circ, \dots, 337.5^\circ$). **b.** The decoder successfully estimates 3D motion direction; estimates (dots) fall on the unity line (dashed white line). **c.** The mean estimation error (purple line) and standard deviation (purple cloud) are plotted as a function of 3D direction ($n=36000$; 100 independent estimates per 360 directions tested). The standard deviation of the estimates (purple cloud) varies cyclically as a function of the motion direction presented. This is a consequence of the binocular projective geometry. **d.** For comparison to a-c: Idealized population of neurons with Gaussian tuning for 3D motion direction. Here we show 16 evenly spaced Gaussian tuning curves (with preferred directions: $0^\circ, 22.5^\circ, 45^\circ, \dots, 337.5^\circ$); 236 evenly spaced neurons were used in the simulated population. This matches the number of neurons in the recorded population, and simulated in the computational model. **e.** Gaussian decoder successfully estimates 3D motion direction; estimates (purple dots) fall on the unity line (dashed white line). **f.** The mean estimation error (purple line) and standard deviation of estimates (purple cloud) are plotted as function of 3D direction ($n=36000$; 100 independent estimates per 360 directions tested). Note that the standard deviation of the estimation error does not vary as a function of the motion direction presented (compare to 3c.)

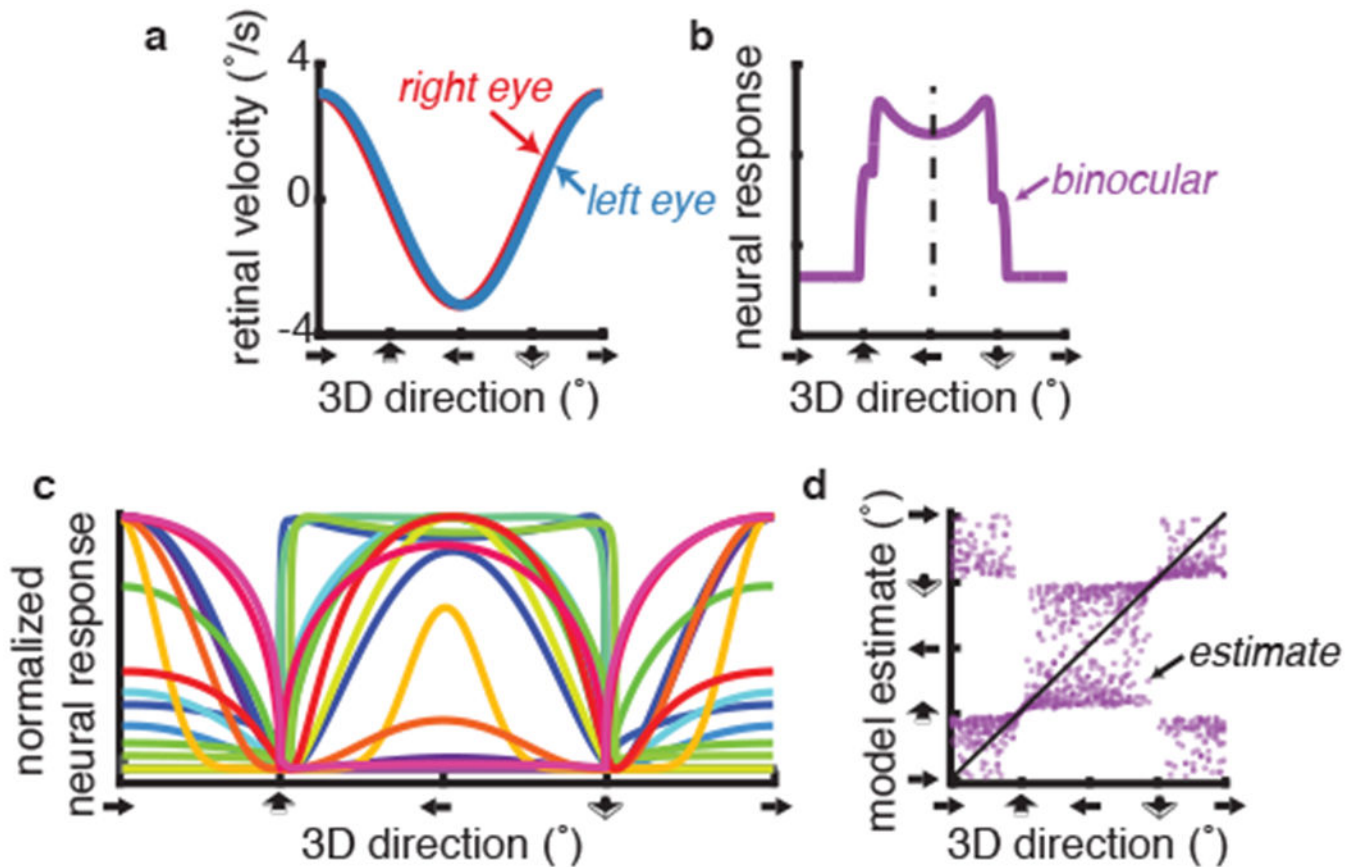


Figure 4: Model estimates of 3D motion direction change with viewing distance, resulting in surprising model errors at far viewing distances.

a. At a larger (67cm) viewing distance, the retinal velocities are smaller in magnitude and the difference between the left and right eye retinal velocities is drastically reduced. **b.** The effect of increased viewing distance on individual tuning curves is a convergence of steep transitions on the toward/away motion directions. This results in a relatively symmetrical function except close to the toward and away directions. This symmetry is present across the whole population (because it is a lawful consequence of binocular projective geometry; e.g., **c**) and it leads to the unusual model errors evident in **d**. **c.** Binocular tuning curves for 3D motion direction at a viewing distance of 67cm. These 16 3D direction tuning curves are the same example units as those shown in figure 3a. **d.** Model estimates of 3D motion direction for a viewing distance of 67cm ($n=15$ per 72 directions tested). A pattern of biases and depth-sign errors emerges, forming a 'X' pattern of results.

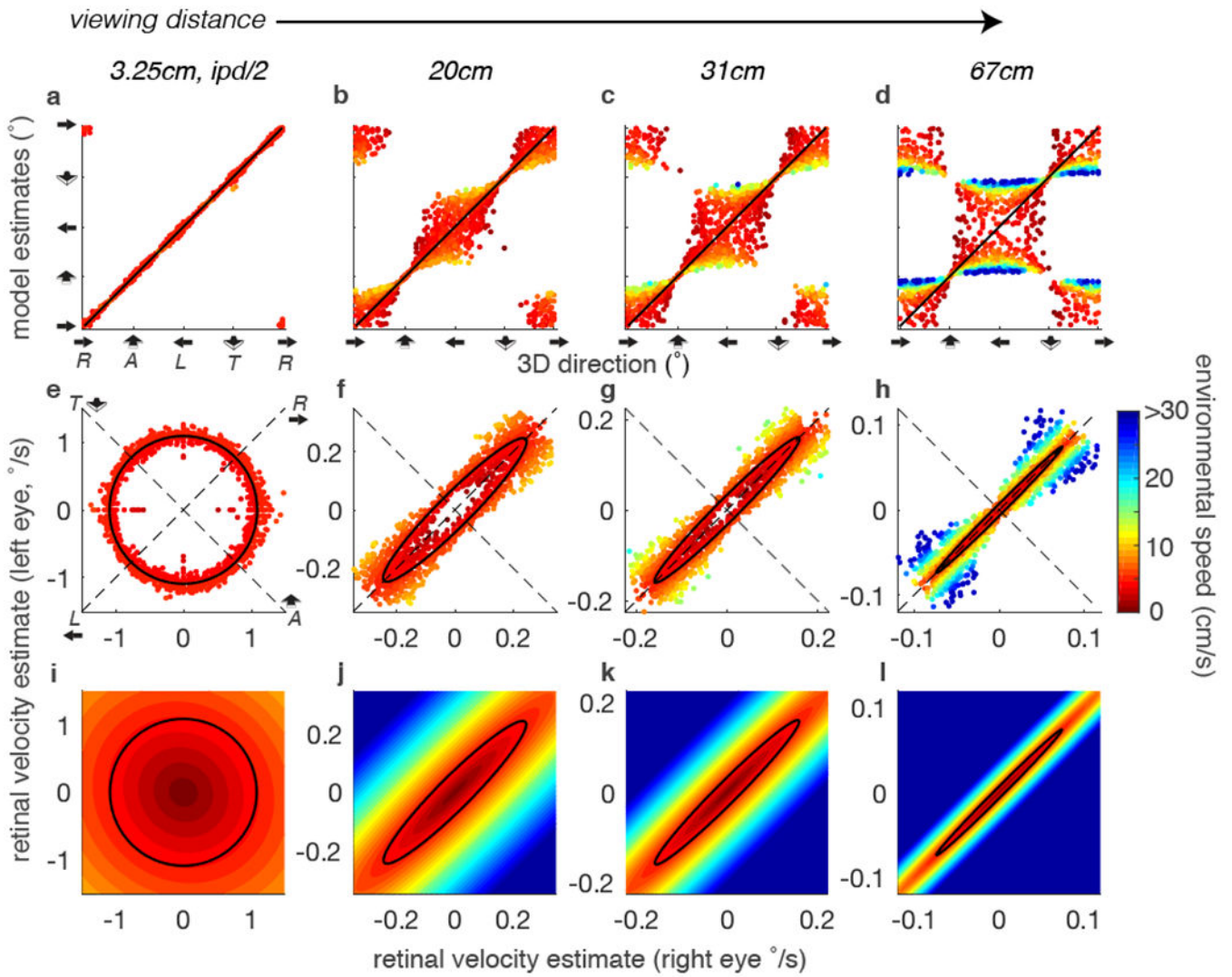


Figure 5: Systematic biases for toward/away motion emerges with increased viewing distances. **a-d.** Model performance for motion direction estimation for a single environmental speed (5cm/s) at four different viewing distances (3.25cm, 20cm, 31cm, 67cm). Colors indicate model estimates of environmental speed. The unity line (black) marks the presented motion directions. **e-h.** The same model and estimates as a-h, but plotted as a function of the corresponding left and right eye retinal velocities. Again the thick black line represents the presented motion. The dashed lines indicated the axes of toward/away motion and left/right motion. From this representation, it is evident that the variability around the retinal velocities is similarly shaped across viewing distances but that the transformation to the environmental velocity results in systematic differences in model estimation performance for environmental velocities at different viewing distances. **i-l.** The mapping from retinal velocities to environmental velocities at different viewing distances. Again the thick black line represents the presented motion.

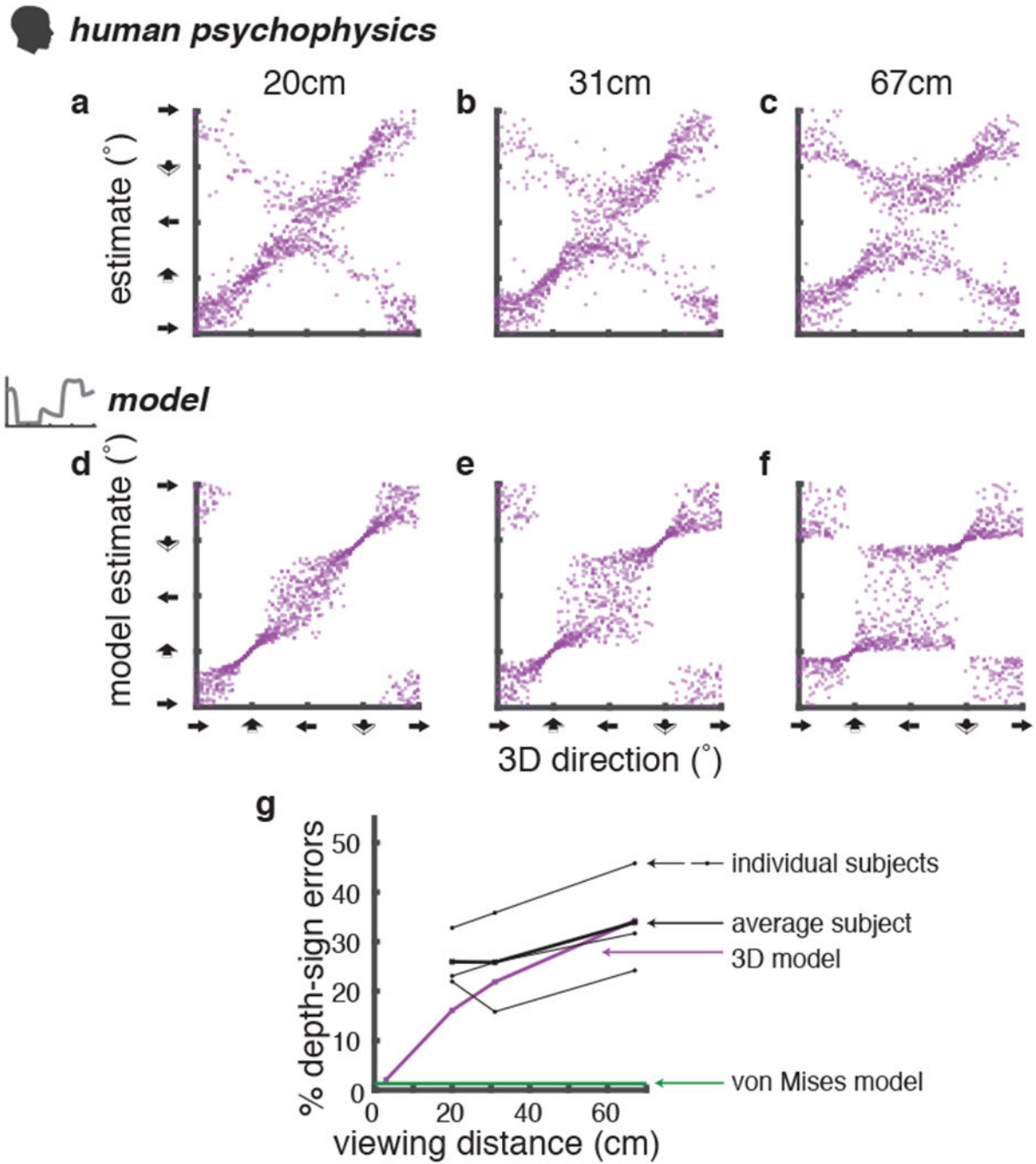


Figure 6: Human performance on a 3D motion direction estimation task matches model observer performance

a-c. Results from a human psychophysics experiment. Three observers were shown dot motion clouds moving in one direction and asked to estimate the 3D motion direction. **a.** 3D motion direction estimation performance collapsed across 3 human observers at a 20 cm viewing distance. Each dot represents an estimate from a single trial (n=15 per 72 directions tested). Data points are rendered semi-transparently in order to make visible the density of estimates. **b.** 3D motion direction estimation performance collapsed across 3 human

observers at a 31 cm viewing distance. **c.** 3D motion direction estimation performance collapsed across 3 human observers at a 67 cm viewing distance. **d-f.** 3D model performance estimating motion direction in the same conditions as the human observers in a-c. Notice that with the increased viewing distance there is an increase in the number of depth sign errors and a bias away from frontoparallel motion for both the model and the human observers. **g.** The percentage of depth sign errors as a function of viewing distance for the two models and 3 human observers, demonstrating that there is a categorical difference between the predictions made by the 3D model and the von Mises model. Human observers are clearly better matched by the 3D model.

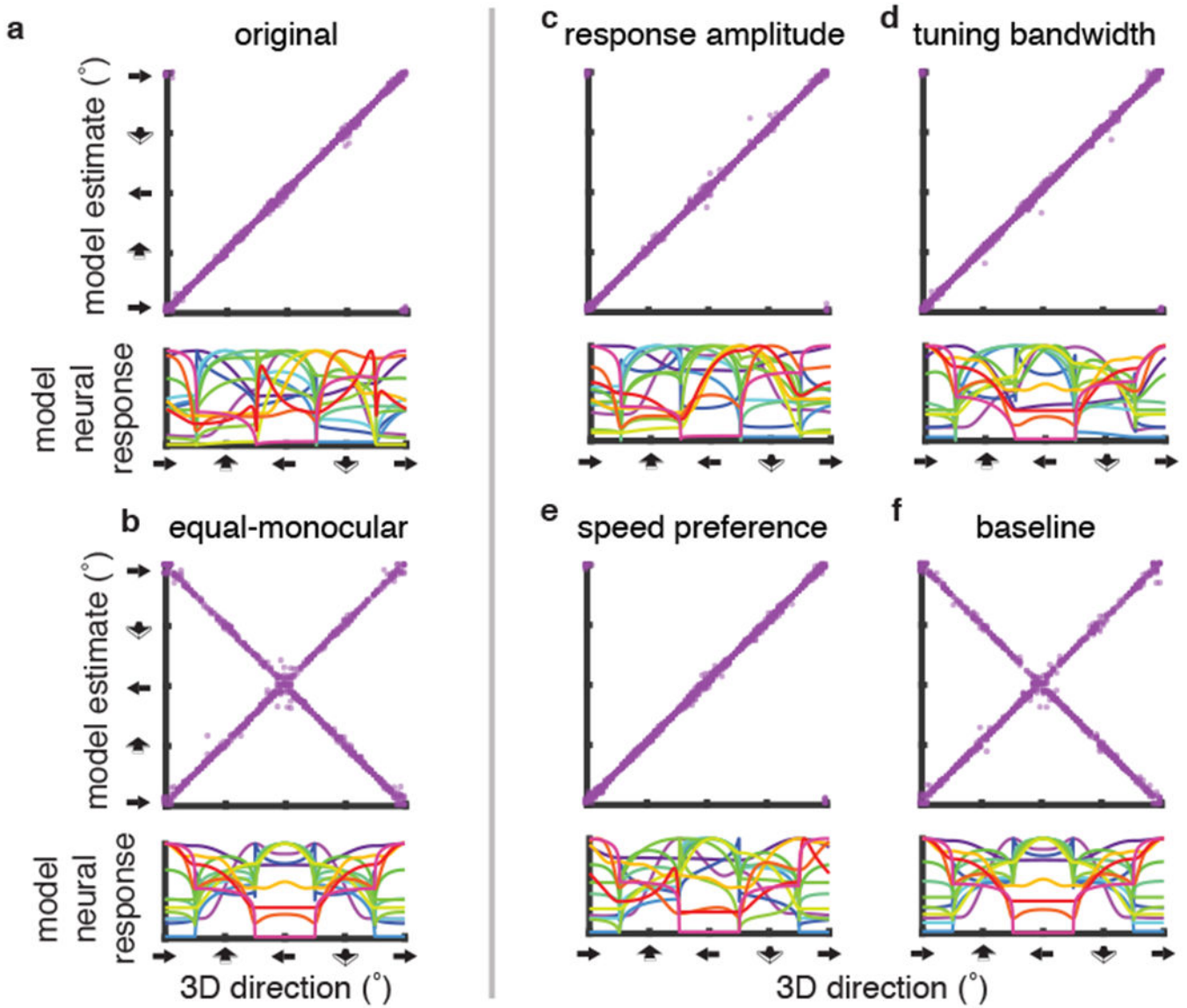


Figure 7: Subtle tuning differences across the two eyes enable the toward-vs-away aspect of decoding for 3D motion direction.
 Each lettered panel shows the performance of a decoder (upper), based upon a particular simulated neural population (lower) at a simulated viewing distance of $\frac{ipd}{2}$ (as in Figure 3), given a particular set of tuning characteristics: **a.** the original tuning measured in this paper (slightly different across the two eyes for all parameters) **b.** equal monocular inputs from the two eyes **c.** differs across the two eyes only in response amplitude **d.** differs only in tuning bandwidth **e.** differs only in speed preference or **f.** differs only in baseline firing rate