nature neuroscience

ARTICLES

https://doi.org/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 B. Czuba[1,2], Jake A. Whritner[2], Adam Kohn[4], Alexander C. Huk[1,2,5] and Lawrence K. Cormack[1,2]

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. However, 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 the 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 on the left and right retinae. Then subsequent decoding processes must allow the animal to interact with the dynamic, three-dimensional (3D) world. It is therefore not the retinal motion that is ultimately important, but rather inferring the 3D environmental motion that gave rise to the retinal stimulation and subsequent cortical activity.

In some cases, the stimulation on the sensory epithelium is a fairly direct proxy for the stimulus in the environment. Some tactile perception works this way (that is, 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). Therefore, for the visual system to work outside of the context of a frontoparallel computer screen in a laboratory, decoding of the 3D environment requires additional computation to infer the properties of the world that gave rise to the stimulation on 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 area MT of the 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 that 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), a well-established phenomenon that has until now lacked any clear functionality. Together, this framework shows that even visual perception, long taken as a model system owing to its apparently simple stages of image formation and transduction, involves idiosyncratic encoding that is shaped by geometric projection at the earliest stages of stimulation on the sensory epithelium. Visual perception therefore requires corresponding non-canonical decoding mechanisms downstream to reconstruct the 3D environment well enough to inform perception and guide action.

Results

We developed a computational model of MT responses to motion that 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 started with environmental representations of object motion, worked through the projective geometry on both retinae, and then took into account known responses to monocular velocities and binocular combination. The model predicted 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.

[1]Institute for Neuroscience, The University of Texas, Austin, TX, USA. [2]Center for Perceptual Systems, Department of Psychology, The University of Texas, Austin, TX, USA. [3]Center for Neural Science, New York University, New York, NY, USA. [4]Dominick 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. [5]Department of Neuroscience, The University of Texas, Austin, TX, USA. *e-mail: kathryn.bonnen@nyu.edu

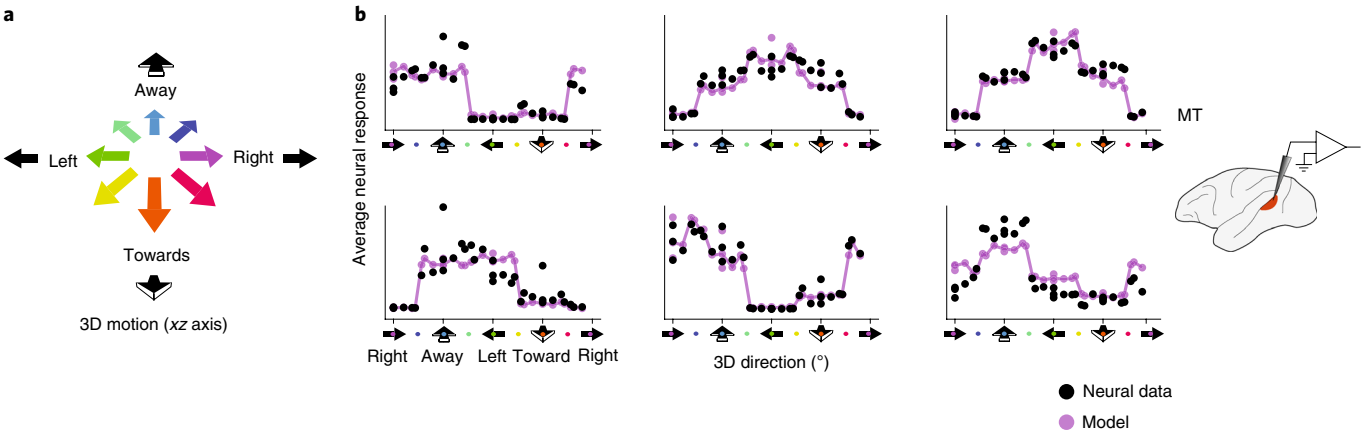

Fig. 1 | MT neurons exhibit an atypical 'terraced' tuning structure for environmental velocities in 3D. a, For the purposes of this study, 3D motion refers to velocities that fall on the *xz* plane. This allows us to unwrap the direction of motion onto a linear axis (as is typically done with frontoparallel motion): right, away, left, towards, right. **b**, Average neural response to 3D (*xz*) direction of motion for six example neurons in the macaque MT[4]. Each panel depicts the average response of a single example neuron to the presentation of different 3D directions of motion (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° s^{-1}, −2° s^{-1}, −1° s^{-1}, 1° s^{-1}, 2° s^{-1} and 10° s^{-1}). This results in motions presented in 28 unique directions (of varying environmental speeds), with each of the three cardinal directions (right, away, left, towards) repeated at three different speeds. These motion stimuli were presented at six different grating orientations (0°, 30°, 60°, 90°, 120° and 150°), all drifting orthogonal to the 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 does not have a proper binocular velocity signal. Additional details about these experiments can be found in the original paper[4].

Highly atypical tuning structure for 3D environmental velocities in the macaque MT. Recent work across electrophysiology and functional magnetic resonance imaging (fMRI) has implicated the MT in the processing of motion off the frontoparallel plane[3–5]. Here we performed a closer examination of neural recordings in the macaque MT to characterize the functional form of tuning for 3D motion direction (specifically *xz* directions, Fig. 1a). The black points in Fig. 1b show the measured tuning curves (that is, the average neural response) to the presentation of different 3D directions of motion (that is, motion on the *xz* plane) for six example neurons. The firing rate of these neurons was modulated by changes in the 3D direction of motion. Notably, the tuning curves were characterized by steep transitions in four locations on the motion direction axis (roughly between each pair of adjacent cardinal directions: right, away, left, towards) with relatively little change in firing everywhere else. Given that the vast majority of tuning to simple sensory features takes on a Gaussian form[6–8], including MT responses to frontoparallel directions of motion, at first glance this seems 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 that fall on the retina.

Atypical tuning structure for 3D environmental velocities is predicted by a model that incorporates environment-to-retina geometry. To understand the non-canonical 3D tuning curves, we developed a model to encode 3D motion that incorporated the projective geometry from the environment onto the two retinae. It then applied the canonical log-Gaussian tuning of MT neurons to the pair of retinal velocities that correspond to a particular 3D direction[9,10], and took the linear combination of those monocular responses. When a particular 3D direction of motion was presented on the *xz* plane at a given viewing distance, the geometric projection onto the retinae resulted in separate left and right eye retinal velocities (Fig. 2a,b). The direction and speed of the retinal velocities were dependent on the environmental velocity of the 3D motion as well as its distance to the eyes. Correspondingly, any egocentric representation of the 3D direction of motion along the *xz* plane must consider the locations of the eyes, as well as the viewing distance.

Here, we used an egocentric coordinate system in which the frontoparallel rightward–leftward and 3D towards–away motion axes of the *xz* plane were always anchored to the cardinal axes (0° and 180°, and 270° and 90°, respectively). In the first part of the paper (including Figs. 1–3), we used a scaling that made the ocular axes of the left and right eye (that is, motions directly towards or away from either eye) orthogonal to one another, placing each midway between the 3D and frontoparallel axes. In effect, this also divided the space equally into regions with the same or oppositely signed motion in the two eyes. We began with this representation because it matched that of previous work[4,11,12] and because the layout made it very easy to examine the relationship between the motion in the environment and the motion that falls on the retinae. This coordinate system can be interpreted in environmental terms as having an implausibly short effective viewing distance, equal to half the average inter-pupillary distance (~3.25 cm in humans and ~1.63 cm in macaques, Fig. 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 ref. [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 produced a neural response derived directly from the monocular tuning curve (Fig. 2c–e, left panels); both of these monocular velocity responses could then be replotted as functions of the 3D direction of motion (Fig. 2c–e, middle panels). The predicted binocular response was a linear combination of the corresponding monocular responses (Fig. 2c–e, right panels; Online Methods equation (1)). Combining binocular projective geometry and canonical tuning for retinal stimulation within a simple linear model resulted in tuning curves with abrupt discontinuities, characterized by multiple plateaus separated by steep cliffs (for example, Fig. 2c–e, right panels). This shape deviated substantially from the classic smooth unimodal (that is, bell-shaped) tuning observed across almost all sensory features

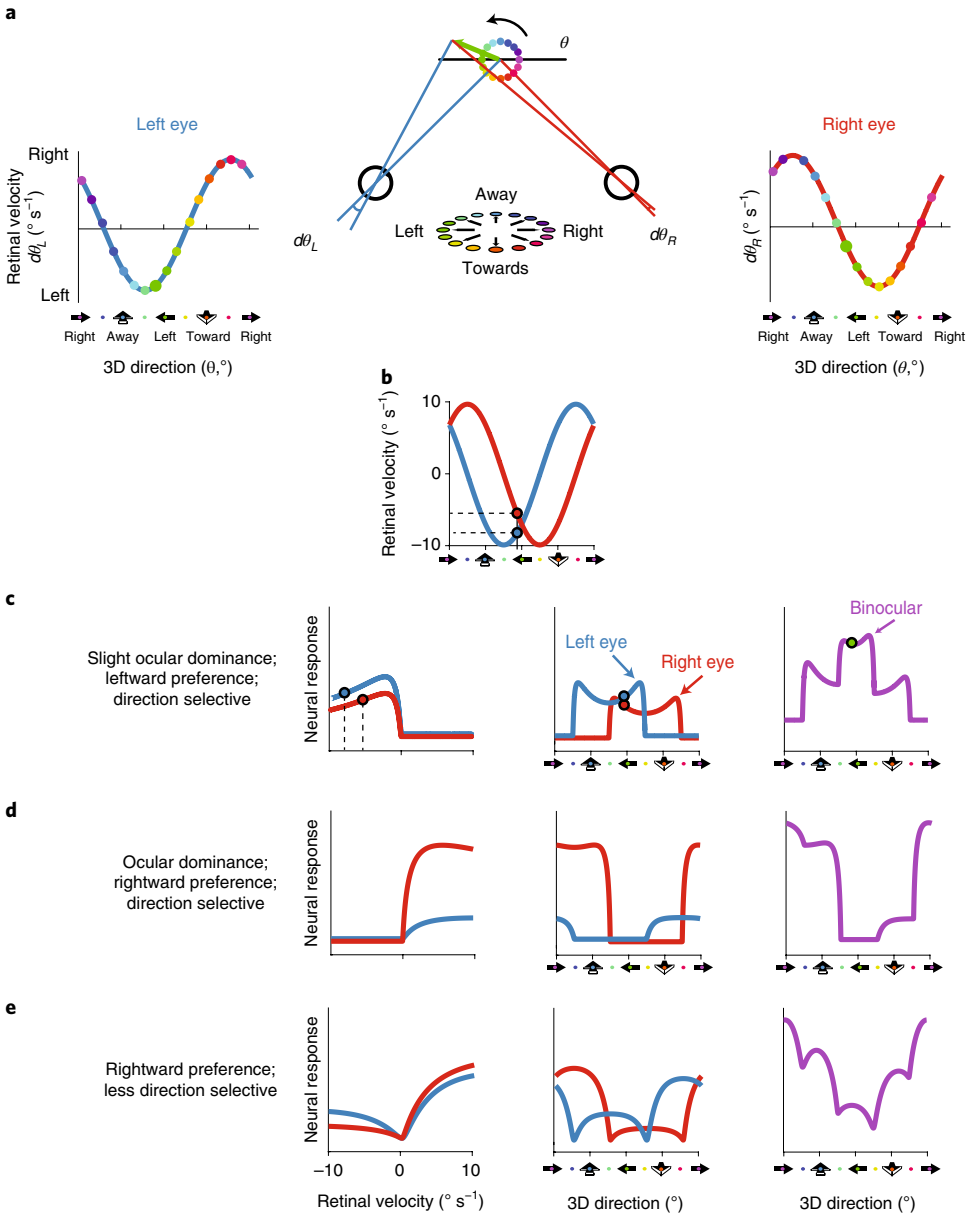

Fig. 2 | An encoding model that incorporates the environment-to-retina geometry of 3D motion predicts atypical structures for binocular 3D motion tuning curves. a, Diagram of the projection of 3D motion (confined to the *xz* plane; middle panel) onto the left eye (blue; left panel) and the right eye (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 (that is, leftward motion in the world is also plotted as leftward in the retinal velocity panels). The assumption that the ocular axes are 90° apart results in an effective viewing distance of $\frac{ipd}{2}$, where *ipd* is the interpupillary distance. **b**, Left and right eye retinal velocities as a function of the direction of 3D motion. 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 three neurons found in ref. [4]. **c**, A 3D model neuron that exhibits slight ocular dominance, leftward preference, and is direction selective. Left: monocular retinal velocity tuning curves for the left and right eye. Middle: monocular neural responses as a function of the direction of 3D motion, built from the composition of the functions depicted in **b** and the left panel. Right: 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 than the neurons in **c** and **d**.

and systems[6–8]. Although atypical in appearance, the model tuning curves bore a striking resemblance to MT responses to binocular 3D motion stimulation, capturing the qualitative deviations from bell-shaped tuning curves[4] (for example, purple vs black dots in Fig. 1, discussed further below).

Given the qualitative success of this model, we further quantified its ability to describe a full electrophysiological data set ($n = 236$ neurons, 4,500 responses collected per neuron) collected in the macaque MT[4] (Online Methods). We predicted the binocular response to the 3D direction of motion 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 the 3D direction of motion. Relying solely on the geometric transformations from environment to retinae (and the assumption of interocular additivity), the model accounted for 76% of the variance in the data (187 of

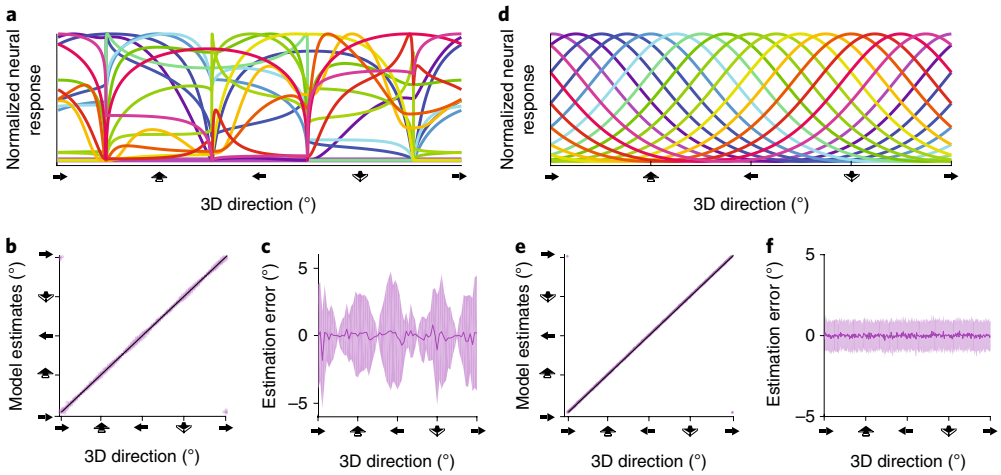

Fig. 3 | A 3D model decoder successfully estimates the 3D direction of motion, but the resulting pattern of estimates is distinct from an idealized Gaussian (von Mises) model. a, Binocular tuning curves from the computational model for decoding the direction of 3D motion, assuming a viewing distance of $\frac{ipd}{2}$, where ipd is interpupillary distance. These 16 example 3D direction tuning curves were chosen because their preferred direction (as calculated by the vector average) was closest to tiling 3D direction with 16 evenly spaced values in the *xz* plane (0°, 22.5°, 45°, …, 337.5°). **b**, The decoder successfully estimates the direction of 3D motion; estimates (dots) fall on the unity line (black line). **c**, The mean estimation error (purple line) and standard deviation (purple cloud) are plotted as a function of 3D direction ($n = 36{,}000$, 100 independent estimates per 360 directions tested). The standard deviation of the estimates (purple cloud) varies cyclically as a function of the direction of motion 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°, 22.5°, 45°, …, 337.5°); 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 the direction of 3D motion; estimates (purple dots) fall on the unity line (black line). **f**, The mean estimation error (purple line) and standard deviation of estimates (purple cloud) are plotted as a function of 3D direction ($n = 36{,}000$, 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 part **c**).

236 units with 50% or more of the variance explained, median root mean-squared error 7.0 spikes per 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) resulted in a modest improvement, accounting for 82% of the variance in the data (190 of 236 units with 50% or more of the variance explained, median root mean-squared error 4.6 spikes per s).

In the sample tuning curves shown in Fig. 1, the purple curves depict the predictions of our model using the fitted combination coefficients (purple dots and line). Supplementary Figure 1 shows additional example fits for neurons that were well fitted and poorly fitted by this model. The von Mises tuning curve (that is, circular normal) is the canonical tuning curve used to describe 2D direction tuning curves in the MT. For comparison to our principled 3D model, we also fitted von Mises tuning curves to the middle temporal area 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% or more of the variance explained, median root mean-squared error 4.9 spikes per s). A direct statistical model comparison using Akaike information criterion (AIC) and Bayesian information criterion (BIC) analyses (Online Methods) further supported the conclusion that the 3D encoding model performed better than the canonical von Mises (3D model, $\Delta AIC = 274$, 95% confidence interval (CI) 197–348, $\Delta BIC = 173$, 95% CI 173–320; von Mises model, $\Delta AIC = 380$, 95% CI 302–482, $\Delta BIC = 427$, 95% CI 345–530; noting that, for example, a difference in ΔBIC greater than ten 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 was not uniform over the different directions of motion presented. This is related to the failure of the von Mises model to capture the qualitative shape (cliffs and plateaus) of many of the neurons that we observed. In particular, Supplementary Fig. 2 shows that the 3D model was better than the von Mises model at capturing the responses to towards 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 the two models were largely qualitative in nature. The 3D encoding model could capture the qualitative shapes of many of the neurons that we observed, including the abrupt cliffs and the long plateaus in the tuning curves. Furthermore, the 3D model directly implemented binocular combination and suggested explicit mechanisms for the construction of 3D motion direction tuning. By contrast, the von Mises tuning model was purely descriptive. Therefore the 3D encoding model was a better model than the von Mises 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 that are not explained by a von Mises tuning model.

Estimating the 3D direction of motion from this atypical tuning structure reveals a sufficient but idiosyncratic encoding. The success of our model in describing neural responses to 3D motion in MT raised the question of whether a population with such idiosyncratic tuning curves could be used to estimate 3D direction (Fig. 3a). To investigate this, we built a population based on model fits to the neurons recorded in ref. [4], assuming Poisson output noise (Online Methods equations (1)–(3)). We simulated population responses to motion (5 cm s^{-1}) around the *xz* plane of 3D directions sampled at 1° intervals, at a viewing distance of $\frac{ipd}{2}$ (where *ipd* is interpupillary distance), and used a standard maximum log-likelihood decoder to estimate the 3D velocity (direction and speed) from the resulting population response (for example, ref. [13] and Online Methods equations (11) and (12)). Despite the unconventional encoding of 3D directions, this decoder successfully recovered the 3D direction of motion (Fig. 3b, estimates near the unity line).

The unusual structure of 3D direction encoding has important ramifications because the underlying tuning curves do not represent changes in 3D direction with equal fidelity. This was evident in the way in which decoding performance varied as a function of the true direction of motion (Fig. 3c), with regions of higher precision near the steepest portions of the tuning curves. These regions corresponded to what we deem the ocular axes, which are the directions for which the retinal velocities flip sign when the 3D direction changes. When the 3D direction of an object is very close to moving directly towards one of the eyes, small changes in 3D direction can correspond to categorical (direction) changes on 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 was notably distinct from decoding based on canonical tuning (Fig. 3d–f and Supplementary Fig. 3), which predicted 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 resulted 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 that has also been ignored in standard 'retinocentric' models of the MT and direction selectivity. Because our model included viewing distance as a parameter, we could simulate and decode at multiple realistic viewing distances using the same model population, Poisson output noise, and a maximum log-likelihood decoder. Figure 4 shows the pronounced effect of viewing distance: viewing distance changed the retinal projections, which affected the shape of individual tuning curves (Fig. 4c compared to Fig. 3a), and markedly changed the model decoding performance (Fig. 4d compared to 3b). At a further (and more perceptually realistic) viewing distance, the systematic biases and errors of the model estimation results revealed two notable features: a coarse-scale 'X' pattern, indicating errors that are orthogonal to the line of unity (which delineates perfectly accurate estimation), and square structures in the clouds of points, which reflect finer-scale deviations from unity (see refs. [14,15]). These patterns can be thought of as depth-sign errors and a bounded bias away from frontoparallel motion, respectively (Fig. 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 (that is, there is a reduced phase shift in the environment-to-retinal velocity mappings between the two eyes; compare Fig. 4a at a 67 cm viewing distance to Fig. 2b, left panel, at $\frac{ipd}{2}$ or a 3.25 cm viewing distance for a human). For a fixed environmental velocity, any single tuning curve is dependent on the resulting retinal velocities, and therefore on viewing distance. At larger viewing distances, the steep transitions of binocular tuning curves shifted closer to the towards and away directions of environmental motion (Fig. 4b,c), resulting in a more symmetrical tuning curve (see symmetry line in Fig. 4b). The depth-sign errors were due to this increasing symmetry across the neural representation in the presence of noise. Note the approximate mirror symmetry in Fig. 4c, with lines of symmetry at left (←) and right (→).

In addition to the depth-sign errors, a subtler but equally telling idiosyncrasy was 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 Fig. 4d: when a leftward direction was presented (middle of the x axis), estimates (y axis) were repulsed from frontoparallel, but could not be mistaken for motion containing a rightward component (that is, the decoder did not make x-axis sign-flip errors). Thus, the estimates are bounded at the towards 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.

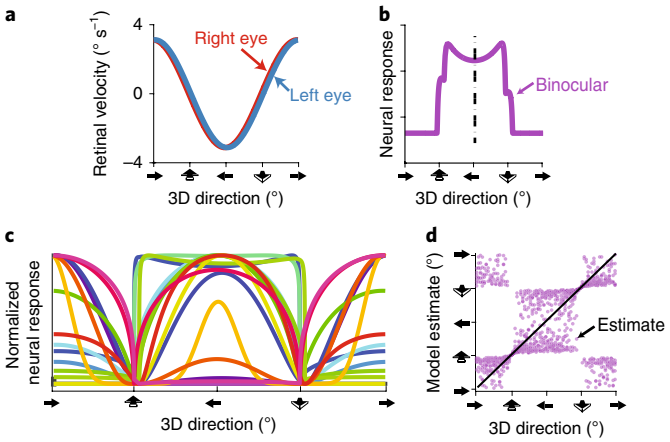

Fig. 4 | Model estimates of the direction of 3D motion change with viewing distance, resulting in surprising model errors at far viewing distances. a, At a larger (67 cm) viewing distance, the retinal velocities are smaller in magnitude, and the difference between the left and right eye retinal velocities is markedly reduced. **b**, The effect of increased viewing distance on individual tuning curves is a convergence of steep transitions on the towards or away motion directions. This results in a relatively symmetrical function except close to the towards and away directions. This symmetry is present across the whole population (because it is a consequence of binocular projective geometry (for example, **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 67 cm. These 16 3D direction tuning curves are the same example units as those shown in Fig. 3a. **d**, Model estimates of the direction of 3D motion for a viewing distance of 67 cm ($n = 15$ per 72 directions tested). A pattern of biases and depth-sign errors emerges, forming an 'X' pattern of results.

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

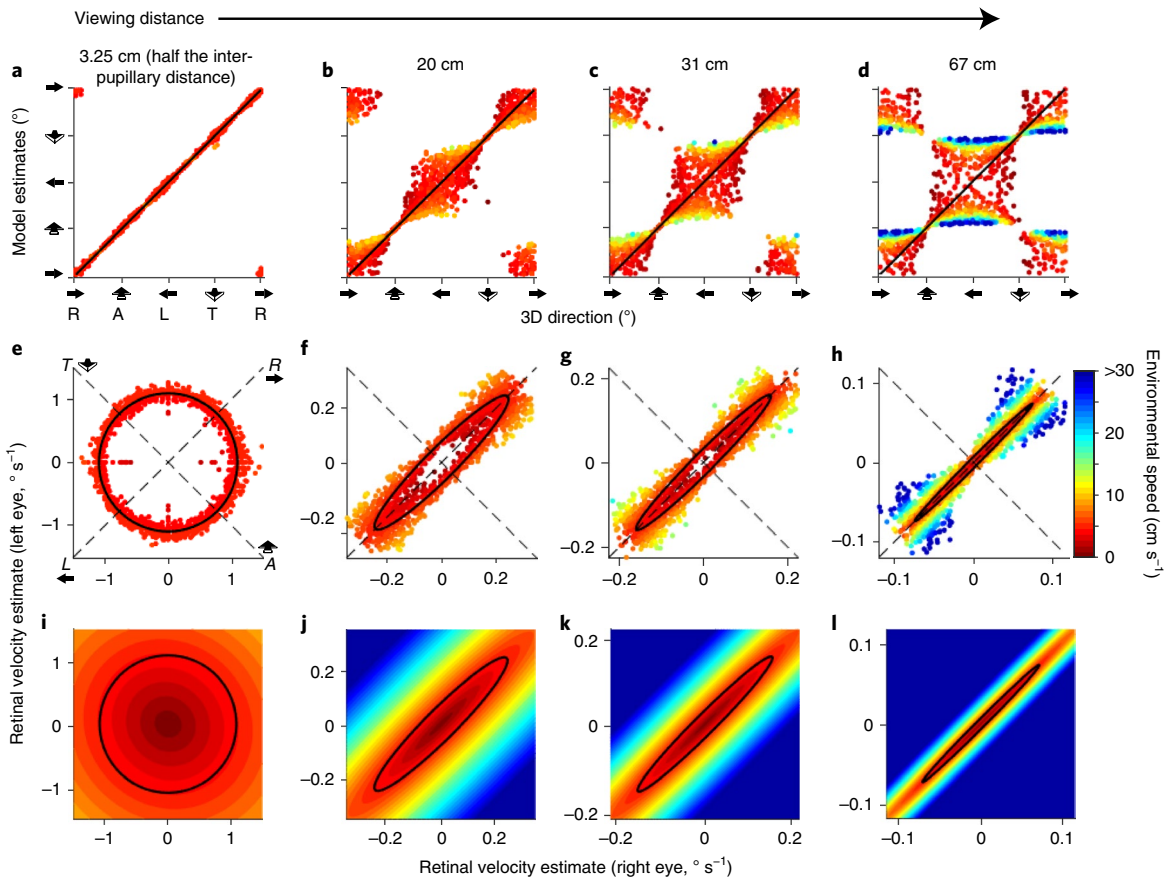

Fig. 5 | Systematic biases for towards or away motion emerge with increased viewing distances. a–d, Model performance for the estimation of the direction of motion for a single environmental speed (5 cm s^{-1}) at four different viewing distances (3.25 cm, 20 cm, 31 cm and 67 cm). Colors indicate model estimates of environmental speed. The unity line (black) marks the presented directions of motion. **e–h**, The same model and estimates as **a–d**, but plotted as a function of the corresponding left and right eye retinal velocities. The thick black line represents the presented motion. The dashed lines indicate the axes of towards or away motion and left or 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. The thick black line represents the presented motion. A, away; L, left; R, right; T, towards.

Although this explication of the model builds intuition for these errors in the performance of the decoder, 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 also show that these perceptual distortions emerge and obey the quantitative functional dependence on position in the environment (that is, 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 the estimation of 3D direction should be a function of both the direction of motion and viewing distance. We designed a perceptual experiment to examine human estimation of 3D direction of motion at several viewing distances. Observers estimated the 3D direction of motion of random dots within a spherical volume (5° in frontoparallel diameter, at 5% contrast, rendered with looming and expansion cues, motion direction at 0°, 5°, …, or 355° on the *xz* plane, with a motion speed of 5 cm s^{-1}) at three different viewing distances (20 cm, 31 cm, or 67 cm). These 3D motion volumes are analogous to the 2D motion apertures found in classic studies of 2D motion. Motion was presented for 1 s and observers reported their estimate of the 3D direction of motion of the dots using a knob to adjust the angle of a stereoscopically rendered indicator on the screen. Supplementary Videos 1 and 2 provide high-contrast examples of the motion stimuli.

Figure 6 shows the combined estimation performance for three observers at three different viewing distances and model performance at the same three viewing distances for comparison. Human observers exhibited depth-sign errors and biases for towards or away motion that fully emerge as a function of viewing distance (performance of individual subjects is shown in Supplementary Fig. 4). Figure 6g illustrates the increase in depth-sign errors with increased viewing distance and compares performance to the predictions of the 3D model and the von Mises model. Although 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 towards-versus-away aspect of decoding for 3D direction. By separately manipulating parameters of the simulated population

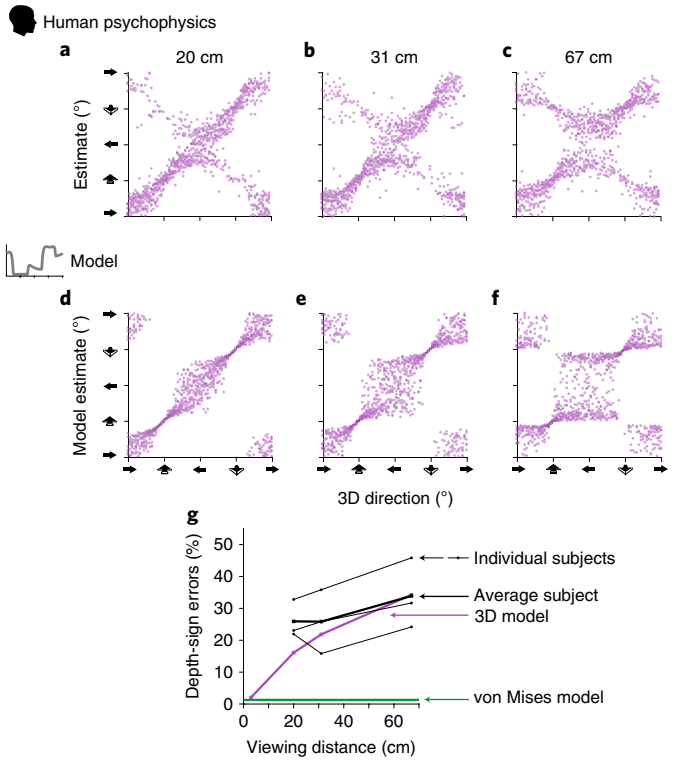

Fig. 6 | Human performance on a 3D motion direction estimation task matches model observer performance. a–c, Results from a human psychophysics experiment. Three observers were shown dot motion clouds moving in one direction and asked to estimate the direction of 3D motion. **a**, 3D motion direction estimation performance combined for three 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 to make the density of estimates visible. **b**, 3D motion direction estimation performance combined for three human observers at a 31 cm viewing distance. **c**, 3D motion direction estimation performance combined for three human observers at a 67 cm viewing distance. **d–f**, Performance of the 3D model in estimating the direction of motion in the same conditions as the human observers in **a–c**. Note 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 three 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.

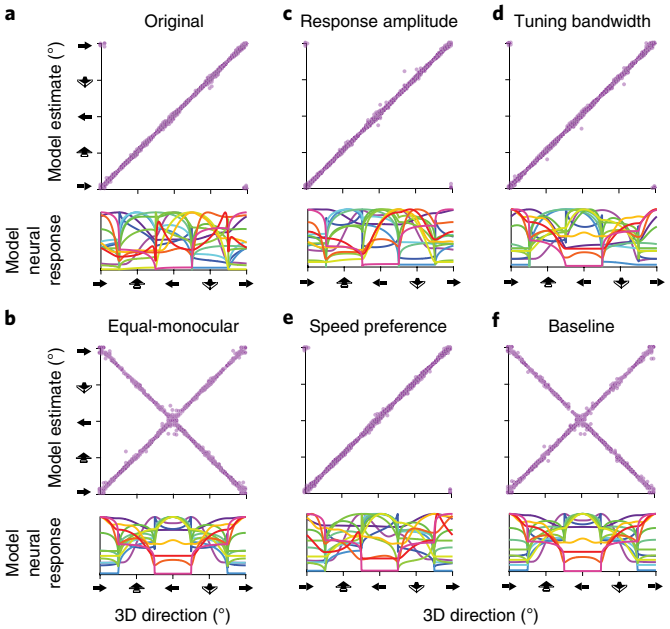

Fig. 7 | Subtle tuning differences across the two eyes enable the towards-versus-away aspect of decoding for the direction of 3D motion. Each lettered panel shows the performance of a decoder (top), based on a particular simulated neural population (bottom) at a simulated viewing distance of $\frac{ipd}{2}$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

(Online Methods equations (1)–(3)), we were able to examine which aspects of neural tuning in MT neurons affect the estimation of the 3D direction of motion. For example, a population with identical monocular tuning parameters (that is, the same speed preference, tuning bandwidth, response amplitude, and baseline firing rate in the two eyes), correctly identified the x (frontoparallel) component of 3D motion. However, in the absence of any implicit eye of origin signatures playing out in such parameters, this 'equal-monocular' encoding could not recover the direction for the depth component above chance levels because there was no differentiating information for towards-versus-away motion components (Fig. 7b).

However, merely incorporating subtly differential monocular tuning (at the levels measured in ref. [4]) revealed that small, seemingly trivial differences in response amplitude, tuning bandwidth, or speed preference between the two eyes are each, in principle, sufficient to represent the 3D direction of motion (Fig. 7c–e, respectively, and Fig. 7a for comparison). Differences in untuned components (such as the baseline firing rate from the two monocular response components) did not provide differential towards or away information (Fig. 7f). Therefore, 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 the visual cortex[17,18], but has rarely been posited as a scheme for carrying information[19]. This theoretical finding indicates that these subtle ocular imbalances may have an important role 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 the 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 the 3D direction of motion. Furthermore, a decoding analysis predicted 3D motion direction estimation performance that varies as a function of the direction of motion and viewing distance, which we showed is consistent with human perceptual judgements and is in stark contrast with default (Gaussian or von Mises) tuning models that have homogeneous

sensitivity across all 3D directions. Therefore, 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 studies in the perceptual literature have reported the frontoparallel bias and depth-sign errors that we observe to be prevalent at longer viewing distances[16,20]. 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[14,15,21]. The use of binocular velocities for the discrimination or estimation of the 3D direction of motion was also proposed by Beverley and Regan[22], with supporting psychophysical experiments that tested the discrimination of the 3D direction of motion 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 on 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 estimation of the 3D direction of motion (that is, 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 non-linear binocular summation[4]. They concluded that these non-linearities were probably critical for sensitivity to 3D motion. Despite the fact that the binocular combination included in our model is purely linear and does not take into account these non-linearities, our model accounts for more than half of the variance in most neurons. We found that some neurons are not well fit by our model. This is at least partly due to non-linearities in binocular combination, which probably sharpen sensitivity to 3D motion. However, the theoretical exercise described here 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 that fall on the two eyes. The field has named this binocular information about 3D motion interocular velocity differences[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 previously). The mechanism representing this type of information does not engage in any differencing per se, although 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 (for example, ref. [25]). Such experiments will also provide an important opportunity to further characterize the monocular and binocular tuning characteristics of neurons, as well as potential dependencies on viewing distance, to test and refine the model proposed here.

In conclusion, our findings emphasize the importance of recognizing the ultimate need of the nervous system 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 the 3D direction of motion as an example, demonstrating that a geometrically constrained encoding model for the 3D direction of motion 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 the estimation of the 3D direction of motion. The geometric framework presented here can be applied to other visual features; for example, slanted and tilted patterns project differential patterns of orientation on 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.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information, details of author contributions and competing interests, and statements of code and data availability are available at https://doi.org/10.1038/s41593-019-0544-7.

Received: 11 March 2019; Accepted: 22 October 2019;
Published online: 2 December 2019

References

1. von Helmholtz, H. *Treatise on Physiological Optics*. Translated by James P. C. Southall (Dover Publications, 1867).
2. Gibson, J. J. *The Perception of the visual World*. (Houghton Mifflin, 1950).
3. Rokers, B., Cormack, L. K. & Huk, A. C. Disparity- and velocity-based signals for three-dimensional motion perception in human MT+. *Nat. Neurosci.* **12**, 1050–1055 (2009).
4. Czuba, T. B., Huk, A. C., Cormack, L. K. & Kohn, A. Area MT encodes three-dimensional motion. *J. Neurosci.* **34**, 15522–15533 (2014).
5. Sanada, T. M. & DeAngelis, G. C. Neural representation of motion-in-depth in area MT. *J. Neurosci.* **34**, 15508–15521 (2014).
6. Hubel, D. H. & Wiesel, T. N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* **148**, 574–591 (1959).
7. Bacon, J. P. & Murphey, R. K. Receptive fields of cricket giant interneurones are related to their dendritic structure. *J. Physiol.* **352**, 601–623 (1984).
8. Jacobs, G. A. & Theunissen, F. E. Functional organization of a neural map in the cricket cercal sensory system. *J. Neurosci.* **16**, 769–784 (1996).
9. Maunsell, J. H. & Van Essen, D. C. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J. Neurophysiol.* **49**, 1127–1147 (1983).
10. Nover, H., Anderson, C. H. & DeAngelis, G. C. A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *J. Neurosci.* **25**, 10049–10060 (2005).
11. Beverley, K. I. & Regan, D. Evidence for the existence of neural mechanisms selectively sensitive to the direction of movement in space. *J. Physiol.* **235**, 17–29 (1973).
12. Cynader, M. & Regan, D. Neurones in cat parastriate cortex sensitive to the direction of motion in three-dimensional space. *J. Physiol.* **274**, 549–569 (1978).
13. Graf, A. B. A., Kohn A., Jazayeri, M. & Movshon, J. A. Decoding the activity of neuronal populations in macaque primary visual cortex. *Nat. Neurosci.* **14**, 239–245 (2011).
14. Cooper, E. A., van Ginkel, M. & Rokers, B. Sensitivity and bias in the discrimination of two-dimensional and three-dimensional motion direction. *J. Vis.* **16**, 5–11 (2016).
15. Rokers, B., Fulvio, J. M., Pillow, J. W. & Cooper, E. A. Systematic misperceptions of 3-D motion explained by Bayesian inference. *J. Vis.* **18**, 23–23 (2018).
16. Fulvio, J. M., Rosen, M. L. & Rokers, B. Sensory uncertainty leads to systematic misperception of the direction of motion in depth. *Atten. Percept. Psychophys.* **77**, 1685–1696 (2015).
17. 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).
18. Katz, L. C. & Crowley, J. C. Development of cortical circuits: lessons from ocular dominance columns. *Nat. Rev. Neurosci.* **3**, 34–42 (2002).

19. Lehky, S. R. Unmixing binocular signals. *Front. Hum. Neurosci.* **5**, 78 (2011).
20. Welchman, A. E., Tuck, V. L. & Harris, J. M. Human observers are biased in judging the angular approach of a projectile. *Vis. Res.* **44**, 2027–2042 (2004).
21. Welchman, A. E., Lam, J. M. & Bülthoff, H. H. Bayesian motion estimation accounts for a surprising bias in 3D vision. *Proc. Natl Acad. Sci. USA* **105**, 12087–12092 (2008).
22. Beverley, K. I. & Regan, D. The relation between discrimination and sensitivity in the perception of motion in depth. *J. Physiol.* **249**, 387–398 (1975).
23. Cumming, B. G. & Parker, A. J. Binocular mechanisms for detecting motion-in-depth. *Vis. Res.* **34**, 483–495 (1994).
24. Harris, J. M., Nefs, H. T. & Grafton, C. E. Binocular vision and motion-in-depth. *Spat. Vis.* **21**, 531–547 (2008).
25. Salzman, C. D., Bitten, K. H. & Newsome, W. T. Cortical microstimulation influences perceptual judgements of motion direction. *Nature* **346**, 174–177 (1990).

Methods

Electrophysiological data. Several analyses performed in this paper rely on an electrophysiological data set ($n=236$) collected in MT of two adult male macaques (*Macaca fascicularis*, aged 3 years and 4 years) under anesthesia in ref. [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: $-10°\ s^{-1}$, $-2°\ s^{-1}$, $-1°\ s^{-1}$, $1°\ s^{-1}$, $2°\ s^{-1}$, and $10°\ s^{-1}$), as well as the responses to the corresponding monocular velocities. The stimulus was constructed using drifting gratings at six different orientations (0°, 30°, 60°, 90°, 120°, and 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 does not have a proper binocular velocity signal. Additional details about these experiments can be found in the original paper[4].

Computational model. *Encoding.* Here we describe the single-neuron encoding model used to generate the model predictions for responses to the 3D direction of motion (for example, Figs. 3–7). Here 3D motion refers to motion on the *xz* plane (Fig. 1a). A velocity on this plane is specified by (θ,m), where θ is the *xz* direction (°) and m is the magnitude of the motion (cm s^{-1}). The binocular response function for 3D motion, $f_B(\theta,m)$ (for example, Fig. 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:

$$f_B(\theta,m) = c_L \cdot f_L(\theta,m) + c_R \cdot f_R(\theta,m) \tag{1}$$

where $f_L(\theta,m)$ and $f_R(\theta,m)$ are the monocular responses (spike rate, for example, Fig. 2c–e) to the corresponding left and right eye retinal velocities (Fig. 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 fitted by log-Gaussian functions[10] and we therefore parameterized the monocular response functions ($f_L(\theta,m)$, $f_R(\theta,m)$) using log-Gaussian curves (for example, Fig. 2c–e). 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 fitted to both directions with coefficients to modulate the relative amplitude of the neural response:

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

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

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; *and* $d\theta_L(\theta,m)$ and $d\theta_R(\theta,m)$ are functions that give the retinal velocities for the left and right eyes, respectively (see below, see also Fig. 2a), given the *xz* velocity (θ,m).

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

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

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

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 location of the motion (for schematic, see Supplementary Fig. 5):

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

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

To find the velocity for the right eye, take the derivative (that is, $d\theta_r$; note that $\frac{d}{dx}tan^{-1}(f(x)) = \frac{f\prime(x)}{1+f(x)^2}$):

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

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

$$d\theta_r = \frac{z\cdot dx - dz\cdot (x-\frac{ipd}{2})}{(x-\frac{ipd}{2})^2+z^2} \tag{10}$$

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

Decoding. The estimation of the 3D direction of motion 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:

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

$$= \sum_{i=1}^{N} log(f_{B_i}(\theta,m))\mathbf{r}_i - \sum_{i=1}^{N} f_{B_i}(\theta,m) - \sum_{i=1}^{N} log(\mathbf{r}_i!) \tag{12}$$

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 ref. [13]). The direction and magnitude of the motion were jointly estimated by maximizing the log-likelihood: $argmax_{\theta,m}\ logL(\theta,m)$.

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°. This is the model used in the comparison in Fig. 3 and Supplementary Figs. 1–3. The response function (f_{von}) is given by the following equation:

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

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*) direction of motion).

Model comparison. The AIC and BIC were calculated for fits of both the von Mises and 3D encoding model to the binocular response data from ref. [4]. AIC and BIC are designed for comparisons of models with differing numbers of parameters. The 3D model is a two-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 five-parameter model shown in equation (13), but is actually a 25-parameter model because a different five 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, aged 20–28 years, one 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 (ProPixx, 120 Hz per eye, 74.5 cm × 132.5 cm) and a Screen Tech ST-PRO-DCF black acrylic glass screen (Screen Tech). We designed a rail system for mounting both the screen and projector that can be easily adjusted to viewing distances from 20 cm to 120 cm without moving the subject. Supplementary Fig. 6 shows a schematic of this system.

Stimulus. The stimuli were fixed spherical dot motion volumes analogous to the dot motion apertures in the classic (frontoparallel) motion literature. Supplementary Videos 1 and 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 eight motion epochs (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°). Spherical motion volumes were 5° (frontoparallel, 1.78 cm, 2.75 cm, and 5.95 cm at the three viewing distances) in diameter at 5° in eccentricity left or right from fixation. 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 the background luminance), moving at one of three speeds (5 cm s^{-1}, 7.75 cm s^{-1}, or 16.75 cm s^{-1}), in one of 72 directions (0°, 5°, 10°, …, 350°, 355°), at one of three viewing distances (20 cm, 31 cm, or 67 cm), rendered with looming and expansion cues.

Procedure. Each trial consisted of a motion epoch lasting 1 s. Subjects reported the direction of motion 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. The figure below the motion cloud indicates the direction of motion of the dot cloud. The experiment was completed in blocks. Each block consisted of 72 trials at a single viewing distance, with pseudorandomly interleaved trials of different speeds and directions (the fastest speed, 16.75 cm s^{-1}, was presented at only the farthest viewing distance). Each condition (direction, speed, viewing distance combination) was repeated five times. The experiment was conducted in 35 blocks for a total of 2,520 trials.

Statistics. No statistical methods were used to predetermine sample sizes but our sample sizes were similar to those reported in previous publications[16,20]. 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 5 cm s^{-1} trials, since that speed was presented at all three viewing distances. Otherwise, no data were excluded. We did not explicitly test for normality.

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

Data and code availability

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

References

26. Barendregt, M., Dumoulin, S. O. & Rokers, B. Stereomotion scotomas occur after binocular combination. *Vis. Res.* **105**, 92–99 (2014).

Acknowledgements

This research was funded by the National Eye Institute at the National Institutes of Health (EY020592, to L.K.C., A.C.H., and A.K.), the National Science Foundation (DGE-1110007, to K.B.), the Harrington Fellowship program (to K.B.), and the National Institutes of Health (T32 EY21462-6, to T.B.C. and J.A.W.). Special thanks to J. Fulvio (University of Wisconsin) and B. Rokers (New York University, Abu Dhabi) for many insightful conversations and for sharing their psychophysical data early on in this project. We thank J. Kaas (Vanderbilt University) for dispensation in referring to 'MT' instead of 'middle temporal area'.

Author 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.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-019-0544-7.

Correspondence and requests for materials should be addressed to K.B.

Peer review information *Nature Neuroscience* thanks Gregory DeAngelis, Anthony Norcia, and Jenny Read for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Kathryn Bonnen

Last updated by author(s): Oct 9, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All visual experiments were performed using custom routines that depend on MatLab (2016b) and Psychtoolbox v3.0.12.
Data analysis	All data analysis and modeling was performed using custom routines that depended on MatLab (2016b) and Psychtoolbox v3.0.12.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The study is quantitative including experimental data and simulations.
Research sample	3 subjects (2 male, 1 female), ranging from 20-28 years of age, with normal or corrected-to-normal vision. This included one subject who was naive to the purpose and hypotheses of the study.
Sampling strategy	No statistical methods were used to pre-determine sample size, but our sample sizes are similar to those reported in previous publications (refs 15, 20).
Data collection	Data collection was performed using a computer and adjustable knob which subjects manipulated to indicate their response. Participants performed the experiment in blocks. For the first 3 blocks the researcher was present to instruct the participant and answer questions. For the remaining 69 blocks the participant collected data on their own. The researcher was not blind to the study hypothesis. Of the three participants, only the naive participant (S3) was completely blind the study hypothesis.
Timing	November 2017 - May 2018
Data exclusions	No data were excluded from the analyses.
Non-participation	No participants dropped out or declined participation
Randomization	Participants were not allocated into experimental groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Extracellular recordings were performed in two anesthetized, adult male macaque monkeys (macaca fascicularis; 3 and 4 years of age).
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	For the electrophysiological data we reanalyzed in this paper, all procedures were approved by the Institutional Animal Care and Use Committee of the Albert Einstein College of Medicine at Yeshiva University and were in compliance with the guidelines set forth in the National Institutes of Health Guide for the care and use of laboratory animals.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	3 subjects (2 male, 1 female), ranging from 22-29 years of age, with normal or corrected-to-normal vision.
Recruitment	Participants were existing lab members, including 1 naive subject.

Ethics oversight	University of Texas at Austin IRB board.

Note that full information on the approval of the study protocol must also be provided in the manuscript.