Supplementary Figure S1

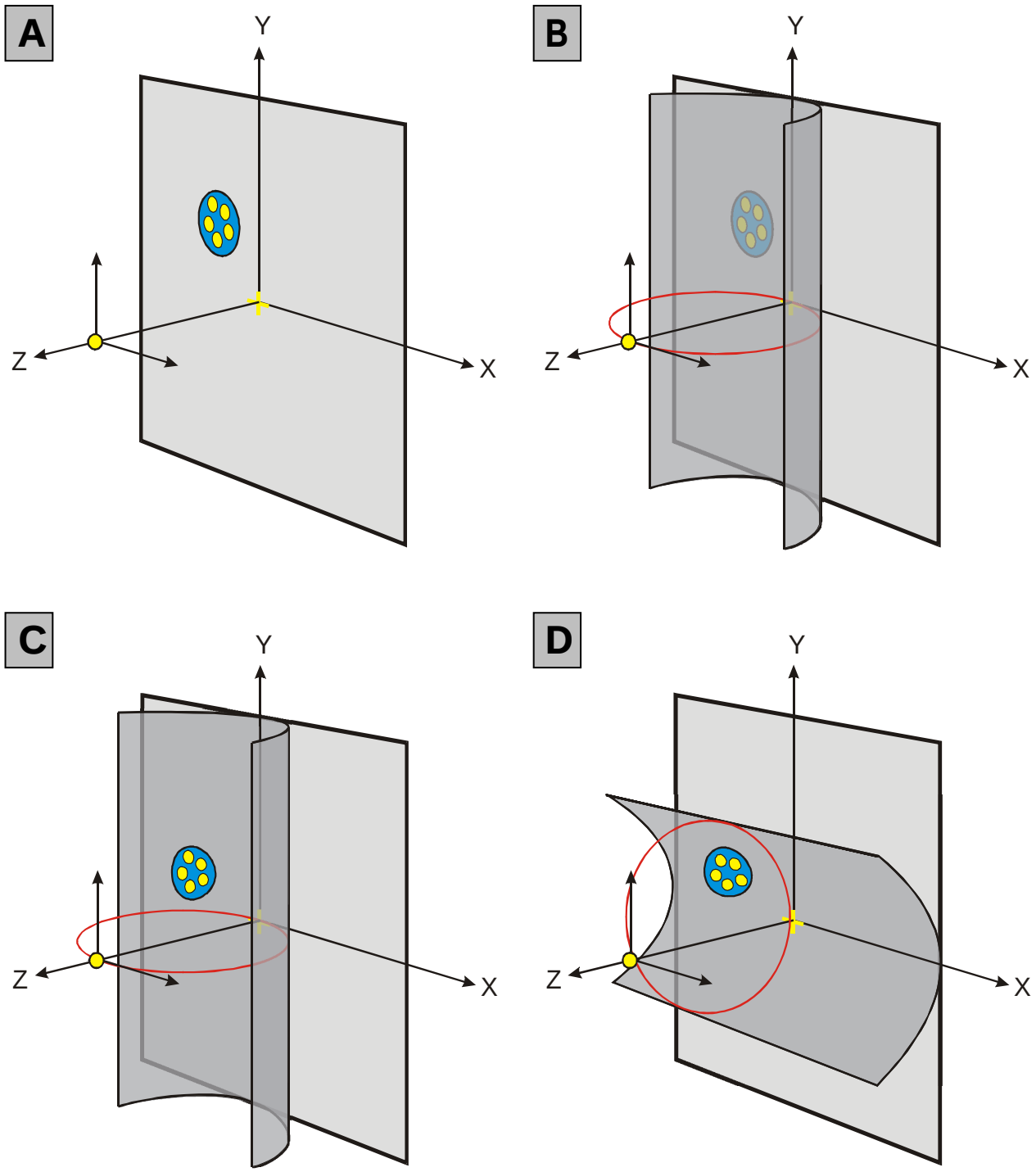

Fig. S1. Explanation of stimulus generation, positioning, and scaling. All visual stimuli were rendered using custom software programmed in C++ and OpenGL. Circular patches of random dots, placed within a virtual workspace, were generated the same way for both stimulus conditions (MP and RM).

A: Initially the center location and size of the random-dot stimulus was identified in retinal coordinates (degrees) by both hand mapping and quantitative mapping of the neuron's receptive field. Those coordinates

were then converted into a physical location on the frontoparallel plane of the screen (shown here in blue for illustration purposes). A random–dot pattern (shown in yellow) was created by choosing random locations within the dimensions of the receptive field, which resulted in an even dot density across the random-dot patch in the image plane (dot size = 0.21 degrees, density = 0.47 dots/degree^2).

B: To locate the random-dot patch in depth, it is insufficient to simply shift the patch along the z-axis. Points will have the same equivalent disparities on the retina if they lie on the theoretical horizontal horopter (the Vieth-Müller circle, VM). When considering the geometry of binocular disparity, the VM circle passes through the fixation point and the nodal point of each eye. For horizontal translations of the observer's eye, the set of points that have a constant equivalent disparity forms a vertically-oriented cylinder that has the VM circle (shown in red) as its cross-section. In this figure, the patch of dots still lies in the plane of the screen. Note that the entire plane of the screen is actually 'far', with the exception of points along the vertical meridian (including the fixation point) which have zero equivalent disparity. To present a stimulus that has a particular simulated depth (equivalent disparity), we therefore must project the image points from the plane of the display screen onto a vertical cylinder of a particular radius. Different equivalent disparities correspond to cylinders having different radii (only the cylinder corresponding to 0° equivalent disparity is shown here).

C: In this figure, our desired set of random-dot locations (in screen coordinates) has been projected onto the cylinder of appropriate radius through a ray tracing procedure. A ray was generated through the cyclopean eye to each individual dot, and the dot was translated along this ray such that it remained at a fixed retinal location (in degrees of retinal angle), but has moved in depth onto the cylinder having the appropriate VM circle as cross-section (for a top-down view, see Fig. 2a of the main text). This resulted in a projection of dots onto vertically-oriented cylinders, the radii of which were smaller for near patches and larger for far patches. By this procedure, the retinal image of the random-dot patch remains circular, but the patch appears as a concave surface in the virtual workspace, as though it were painted on a cylinder. This procedure assures that the patch size, location, and dot density are identical in the retinal image while the simulated depth varies. As the simulated depth increases away from the point of fixation (both near and far), the speed of motion of the dots will increase on the retina. Ideally, a stimulus presented at 0° of equivalent disparity should remain completely stationary on the retina as the observer is translated and maintains fixation. However, in practice the 0° equivalent disparity stimulus will produce very small retinal image motion due to the fact that the animal is translated along a fronto-parallel axis rather than along the VM circle, the fact that there may be small (millimeter scale) errors in positioning the animal at the exact viewing distance used in the calculations, and the fact that pursuit of the fixation target is not perfect (discussed further below).

D: The above depictions have assumed horizontal translation of the observer such that the relevant iso-disparity surface is a vertical cylinder. In our experiments, however, displacement of the observer was always along an axis in the fronto-parallel plane that was aligned with the preferred direction of the neuron under study. In this case, the set of points that have a constant equivalent disparity forms a cylinder that has been rotated about the z-axis so that the axis of movement is in the same plane as the cross-section of the reoriented cylinder. For example, if the neuron preferred dots moving upward, the animal would be translated along the y-axis, and dots would be projected on a horizontally-oriented cylinder as shown here. Since the stimuli were always viewed monocularly in the MP and RM conditions, the same basic geometry applies for any rotation about the z-axis.

Supplementary Figure S2

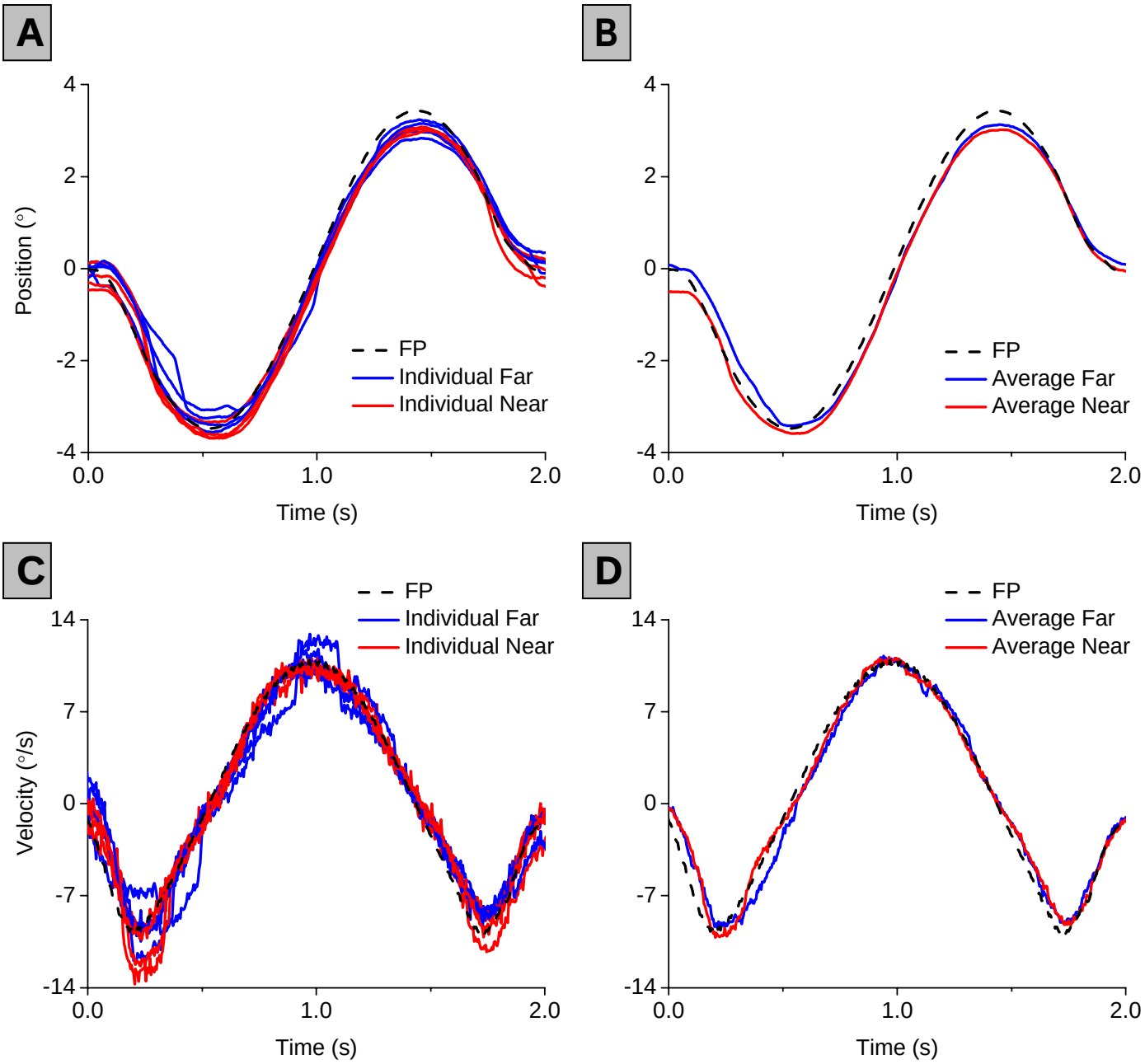

Fig. S2. Eye traces measured during recordings from an MT neuron that was tuned for depth from motion parallax (DSDI in the MP condition = -0.72). In the Motion Parallax condition the virtual world remains fixed while the animal is moved. Dashed curves represent the idealized eye movements necessary to keep the fixation point on the fovea. **A, C**: Solid lines show eye position (**A**) and velocity (**C**) traces for individual trials in which the simulated depth was near (red, equivalent disparity = -2°) or far (blue, equivalent disparity = +2°). **B, D:** Solid lines show average eye position (**B**) and velocity (**D**) traces across 5 repetitions of each simulated depth stimulus (near, red; far blue). In general, pursuit is quite good, and there are no substantial differences between near and far stimuli that would explain the large effect of depth sign on the neuron's firing rates.

Supplementary Figure S3

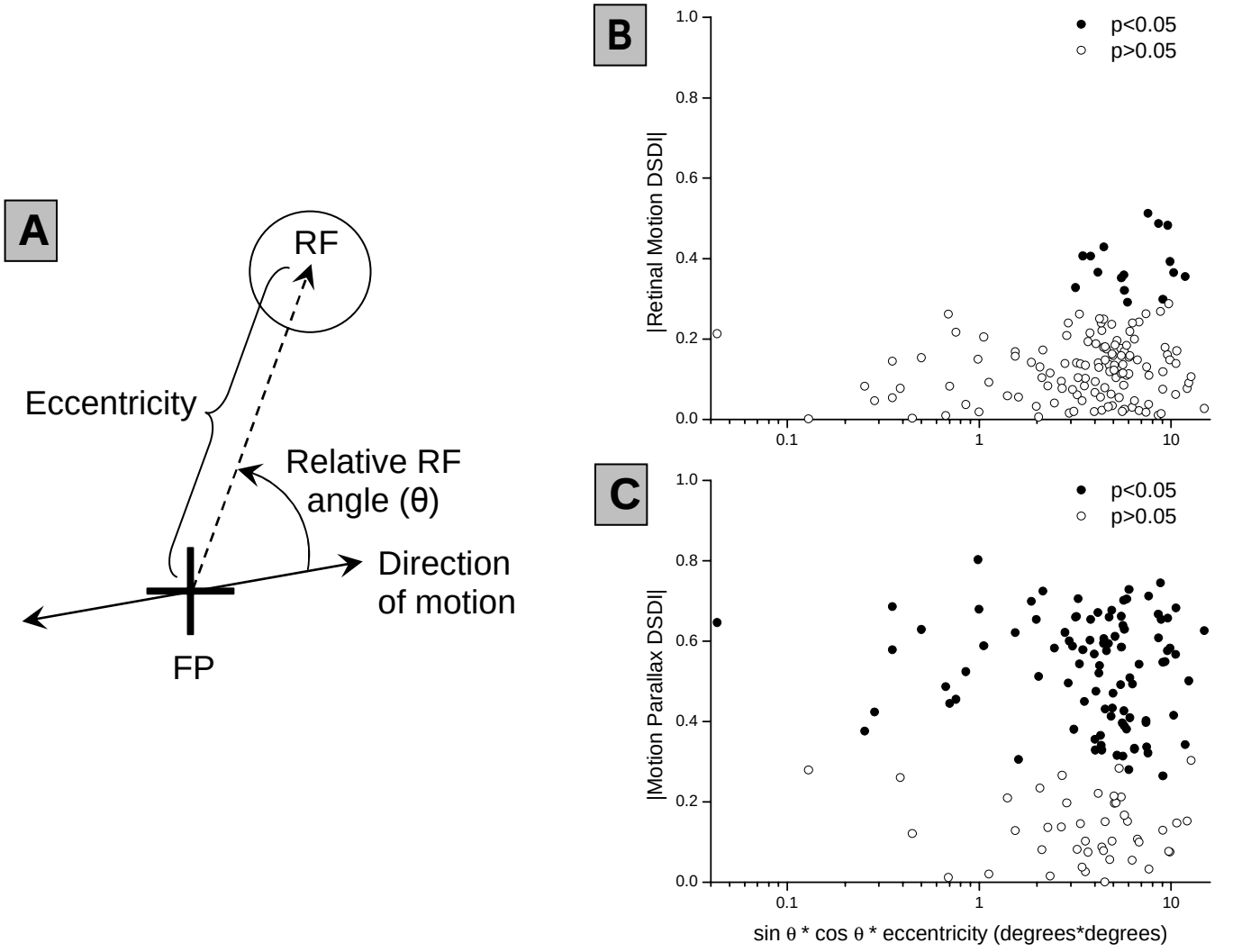

Fig. S3. A motion parallax equivalent of vertical disparities. For horizontal movements, objects located away from the horizontal and vertical meridia will have one component of motion associated with simulated depth (illustrated in Fig. 1 of the main text) and another that is tied to the direction of observer movement and is independent of simulated depth. Assume an object in the upper right visual field is viewed during horizontal observer motion. During observer motion to the right, a near patch will move to the left in the image, and a far patch will move to the right. However, both patches, independent of their simulated depth will rise slightly in the image due to the perspective change that occurs when the eye gets closer to the patch. When the observer moves back to the left, the depth-dependent motion reverses and the patches also sink slightly in the image. When these two components of motion are combined, the patch appears to move somewhat diagonally. The angle of this excursion is symmetric around the direction of observer motion, but it is different between near and far simulated depths. If the direction tuning curve of the neuron is symmetric, and we have accurately aligned our direction of motion with the preferred direction of the neuron, we may never see the effect of this additional motion component on our measurements. However, this motion parallax equivalent of vertical disparities is unavoidable given that we are accurately simulating motion in 3D space. Since this effect would be reflected in both the Motion Parallax and Retinal Motion conditions, it should not affect our conclusions. Nevertheless, we examined how this effect depends on receptive field (RF) location and eccentricity to validate

our explanation. **A:** A schematic illustrating how 'relative RF angle' (θ) is calculated. RF locations that are most susceptible to the motion parallax equivalent of vertical disparities will be eccentric and located obliquely relative to the direction of observer motion (θ ~ 45°). When θ is close to 45°, sin(θ)*cos(θ) will be maximal. Thus, the vertical disparity effect should only be substantial for cells having large values of sin(θ)*cos(θ)*eccentricity (and only the subset of those for which the direction preference was not estimated accurately or for which the tuning curve is asymmetric). **B**: The absolute value of DSDI in the Retinal Motion condition is plotted as a function of sin(θ)*cos(θ)*eccentricity. Reassuringly, the cells with significant DSDIs in the Retinal Motion condition (filled symbols, 17/144) all have large values on the abscissa. Moreover, there is a significant correlation ($R=0.22$, $p<0.02$, Spearman rank correlation) between |DSDI| and sin(θ)*cos(θ)*eccentricity in the Retinal Motion condition. **C**: |DSDI| from the Motion Parallax condition is plotted as a function of sin(θ)*cos(θ)*eccentricity. In this case, there is no correlation between the two variables ($R=-0.02$, $p>0.8$, Spearman rank correlation). This analysis shows that the motion parallax equivalent of vertical disparity causes some weakly significant effects in the Retinal Motion condition for oblique, eccentric RFs, but it cannot account for the much larger effects that we have observed in the Motion Parallax condition.

Supplementary Figure S4

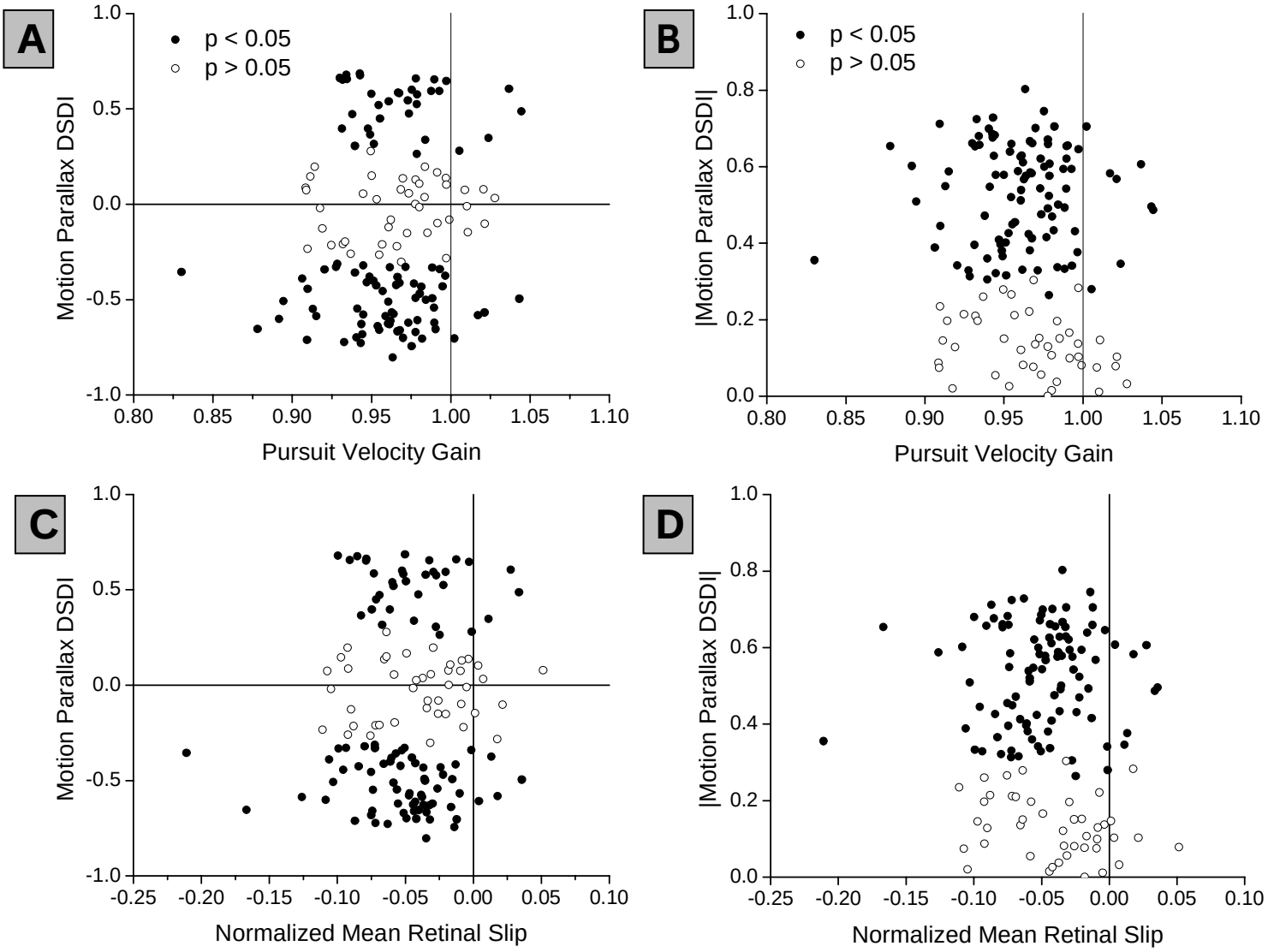

Fig. S4. Relationship between pursuit performance and depth-sign selectivity. Inaccurate pursuit could potentially be problematic because our Retinal Motion condition is only truly depth-sign ambiguous when the animal is precisely tracking the fixation point. Pursuit gain (PG) is one measure of how accurately the animal tracks the motion of the fixation target during translation. To calculate PG, we used Fourier analysis to compute the amplitude of the 0.5Hz component of the average measured eye velocity, and we divided this by the amplitude of the 0.5Hz component of the target (FP) velocity (0.5 Hz is the fundamental frequency). PG values < 1.0 reflect under-pursuit and values > 1.0 indicate over-pursuit. **A**: DSDI values from the Motion Parallax condition are plotted as a function of PG; each datum represents one recording session. Filled symbols denote neurons with DSDI values significantly different from zero ($p < 0.05$). Although many sessions have a PG < 1, there is no significant correlation between PG and DSDI ($N = 144$, $r = 0.12$, $p = 0.15$). **B**: The absolute value of DSDI is plotted as a function of PG. If significant depth-sign tuning in MT were driven by inaccurate pursuit, then one would expect a significant negative correlation in this plot. However, there is no significant correlation between PG and |DSDI| ($N = 144$, $r = -0.08$, $p = 0.33$). Notably, there are many neurons with significant DSDI values for which PG is very close to 1.0. Thus, there is no evidence that inadequate pursuit gain accounts for the observed depth-sign tuning in MT. **C**: Pursuit gain does not take into account the phase lag between eye movement and target (FP) movement. As seen in the example in the main text (Figure 1d), it

was common for eye velocity to lag behind target velocity. In fact, this lag was consistent across sessions (mean phase lag = 5.5 degrees; range 3.3-8.3 degrees). We found no correlation between phase lag and DSDI ($N = 144$, $r = 0.08$, $p = 0.32$) or between phase lag and |DSDI| ($N = 144$, $r = -0.12$, $p = 0.15$). Nevertheless we measured the mean normalized retinal slip (MNRS), which considers phase lag in addition to gain, for each session to see if it could account for depth-sign tuning. To compute MNRS, eye velocity traces were averaged together separately for each starting stimulus phase in a session. At each point in time we calculated normalized retinal slip—the difference between eye velocity and target velocity relative to the target velocity (excluding time points where target velocity was very close to zero). This produces negative values when the eye was slower than the target and positive values when the eye was faster than the target. Considering that all sessions showed some degree of phase lag, eye traces commonly had periods of time in which the eye was moving slower than the target velocity (e.g. 0.5-1s in Fig. 1d), but also had periods in which the eye was moving slightly faster than the target (e.g. 1-1.5s in Fig. 1d). MNRS was then computed by averaging the normalized slip across the two second stimulus period. The scatter plot shows that there is no significant correlation between MNRS and DSDI ($N = 144$, $r = 0.07$, $p = 0.41$). **D**: The absolute value of DSDI is plotted as a function of MNRS and there is again no significant correlation ($N = 144$, $r = -0.09$, $p = 0.27$). The MNRS measure captures the net retinal slip caused by variations in eye movement (both gain and phase), and the lack of significant correlations between it and our DSDI metric confirms that inaccurate pursuit cannot account for the observed depth-sign tuning in MT.

Supplementary Figure S5

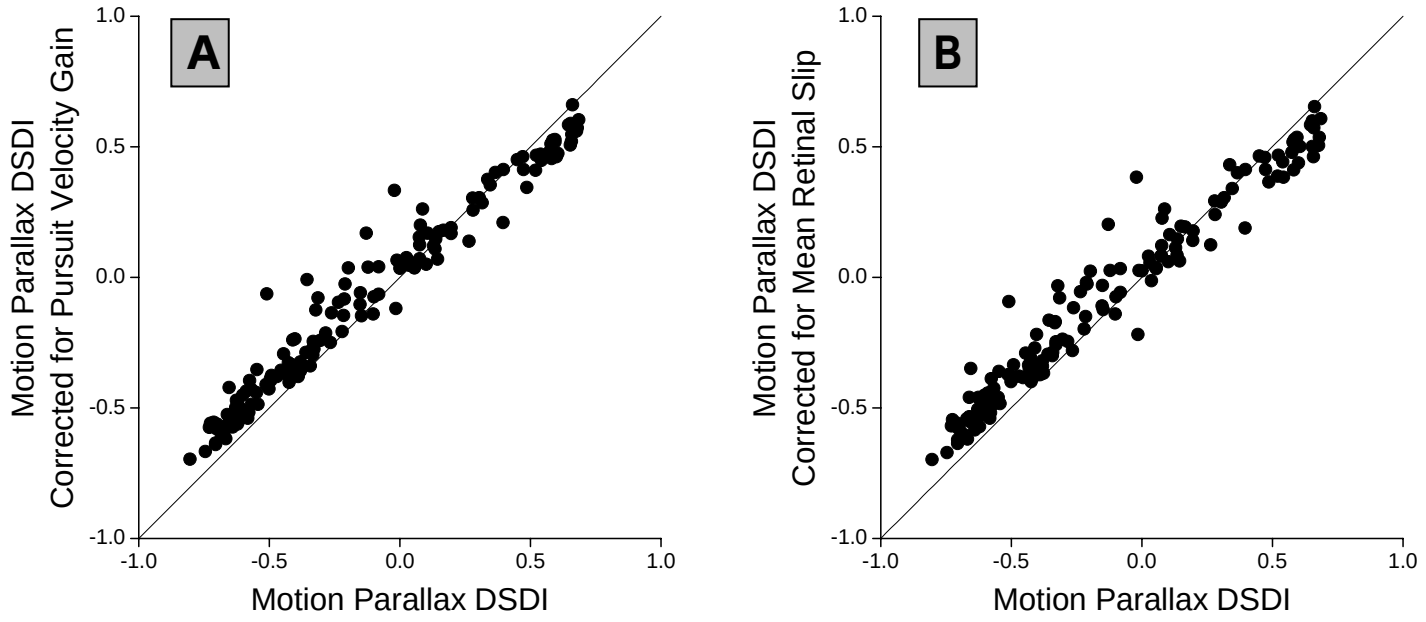

Fig. S5. Population DSDI data corrected for small inaccuracies in pursuit. When a monkey underpursues the fixation target, far patches move more slowly on the retina and near patches move faster. This effect is separate from the fact that patches with greater simulated distances from the fixation plane (near OR far) will have higher retinal velocities. It is possible to relate the effect that magnitude of equivalent disparity has on retinal velocity to the effect that pursuit gain has on retinal velocity. The following expression computes the change in equivalent disparity that corresponds to a particular pursuit gain,

$$\Delta d = \frac{PG - 1}{\left(\frac{\pi \times V}{180 \times I}\right)}$$

where d = equivalent disparity (in degrees), PG = pursuit gain, V = viewing distance (32cm) , and I = interocular distance (Monkey 1 = 3.1cm, Monkey 2 = 3.5cm). If PG = 0.95, for example, a far patch with an equivalent disparity of 1.0° will move on the retina as though it had an equivalent disparity of 0.72° (and PG = 1.0). Likewise, if PG = 0.95, a near patch with an equivalent disparity of -1.0° will move on the retina as though it had an equivalent disparity of -1.28°. Notice that a pursuit gain of 1.0 is associated with no change in equivalent disparity (the numerator of the equation will be zero). Using this expression, we are able to shift our measured tuning curves and recalculate DSDI values that take into account under (or over) pursuit. **A**: As shown in the scatter plot, correcting for pursuit gain slightly reduces the magnitude of depth sign tuning (points lie above the unity slope diagonal on the left, and below the unity slope diagonal on the right). Prior to this correction, the median |DSDI| value in the Motion Parallax condition was 0.42. After correction, the median |DSDI| is 0.36, which is still highly significantly different from the median |DSDI| in the Retinal Motion condition ($p << 0.001$; paired t-test). **B:** An analogous correction can be made based on the mean normalized retinal slip (MNRS), which also takes into account any phase lag in pursuit. In this case, the numerator of the above equation was replaced by the MNRS. Again, the scatterplot shows that DSDIs are slightly reduced, but the central effect is not eliminated. The median |DSDI| under this correction is 0.35, which is still highly significantly different from the median |DSDI| in the Retinal Motion condition ($p << 0.001$; paired t-test). Thus, we conclude that deviations from perfect pursuit have only very modest consequences for our measurements of depth-sign selectivity.

Supplementary Figure S6

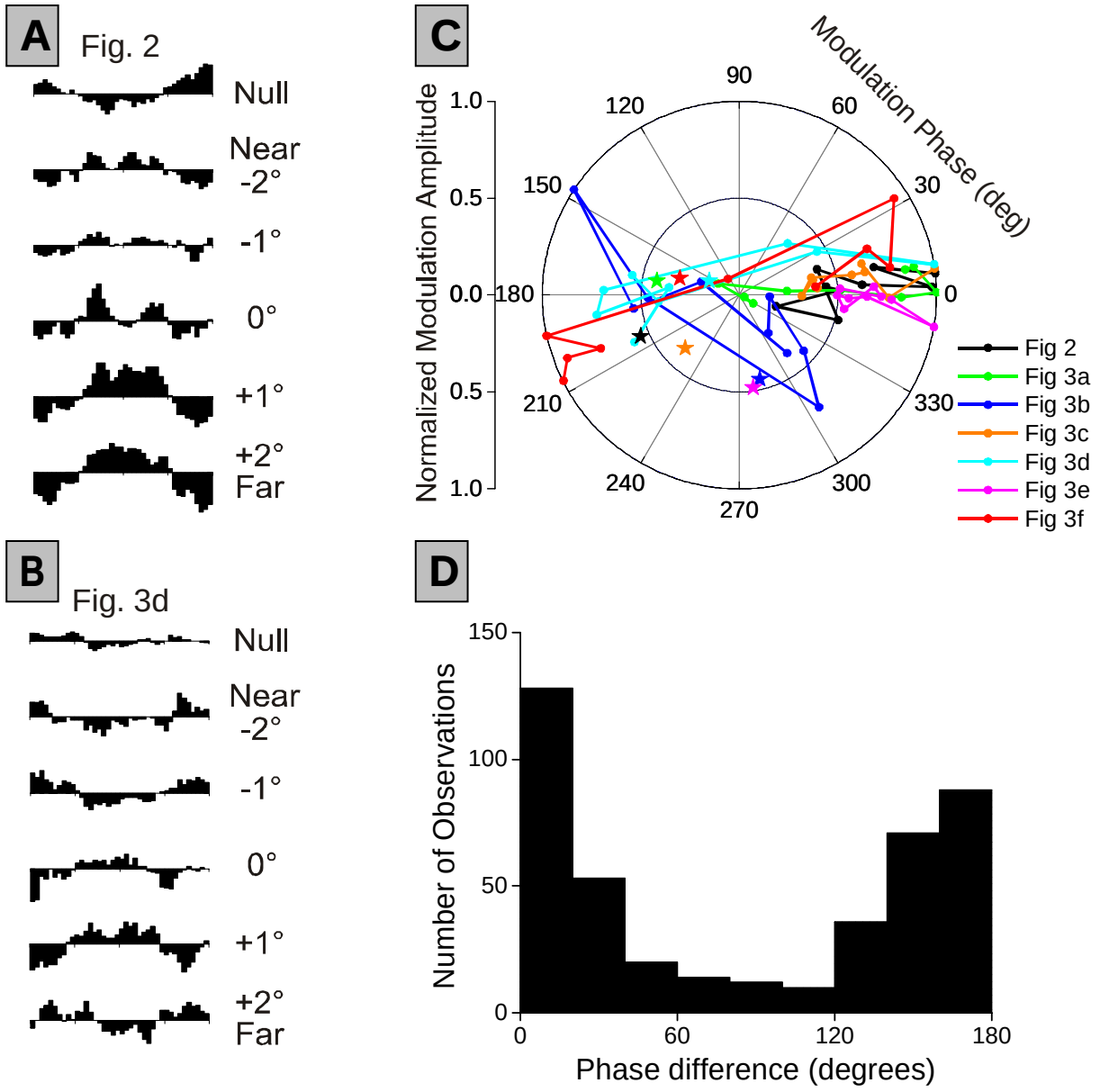

Fig. S6. Responses in the MP condition are not simply an additive combination of visual motion responses and an extra-retinal signal. **A**: PSTHs show the difference in neural responses between the MP and RM conditions for the example neuron of Fig. 2. To obtain these data, a single composite sinusoidal modulation was first calculated for each depth in the RM condition by subtracting the PSTHs for one starting phase (right column, Fig. 2a) from the PSTHs for the opposite starting phase (left column, Fig. 2a). Similar composite modulations were calculated for each depth in the MP condition by subtracting the two columns of PSTHs in Fig. 2b. The difference between these two sets of composite sinusoidal modulations reflects the extra-retinal signals that are added to the RM responses to produce the responses observed in the MP condition. Note that the top PSTH here, labeled 'Null', represents the differential response modulation between the MP and RM conditions when no dots were presented in the MT receptive field. If the responses in the MP condition were simply an additive combination of visual motion responses and an extra-retinal signal related to head or eye movement, then we would expect the MP-RM modulations to be constant across simulated depths (-2° to +2°) and very similar to

the modulation seen in the null condition. While such an arrangement could produce depth-sign tuning, it clearly does not hold for this example neuron. The MP-RM differences for this cell are clearly larger for far depths than near depths. Even more strikingly, the MP-RM modulations have a phase opposite to that seen in the 'null' MP-RM difference (top row). Thus, the interaction for this example neuron is non-linear and has a sign inversion. **B:** The difference in response modulations between the MP and RM conditions for another example neuron (neuron from Fig. 3d). Again the MP-RM differences are not consistent across simulated depths. In this case, both the amplitude and the phase of the modulation vary considerably across simulated depths. **C:** This polar plot describes the MP-RM differences for the seven example neurons of Figs. 2 and 3. Each colored curve represents data from a single cell and shows the amplitude and phase of the MP-RM differences for each of the nine simulated depths tested. Data were calculated by using Fourier analysis to compute the amplitude and phase of the 0.5Hz component, and amplitudes have been normalized to the maximum amplitude for each cell. Colored stars represent the amplitude and phase of modulations from the 'null' trials. Both the amplitude and the phase of the MP-RM modulation vary widely across simulated depths within single cells, and show no consistent relationship to the amplitude and phase of the null modulations. Note that the null trial modulation amplitudes are typically about half as large as the modulation at the maximal depth. We cannot rule out the possibility that some of the response in the null trials is visually driven by retinal slip because the field of view of the monkey was not completely dark (due to the black level of the projector). **D:** Histogram showing the distribution of phase differences (in degrees) between the MP-RM modulations for individual simulated depths and the null modulation for each cell. Only neurons with DSDI values significantly different from zero in the MP condition are included (100/144 cells). Also, phase differences are only reported for pairs of MP-RM modulations in which both the null MP-RM modulation and the MP-RM modulation for that particular depth were significantly different from zero ($p<0.05$; permutation test). Seventy-five of the 100 tuned cells are represented by 1-9 pairs each, for a total of 432 observations. The distribution is clearly bimodal, such that roughly half of the modulation pairs are out of phase with each other. This demonstrates that the responses in the MP condition are not simply an additive combination of the RM responses and an extra-retinal signal.

Supplementary Figure S7

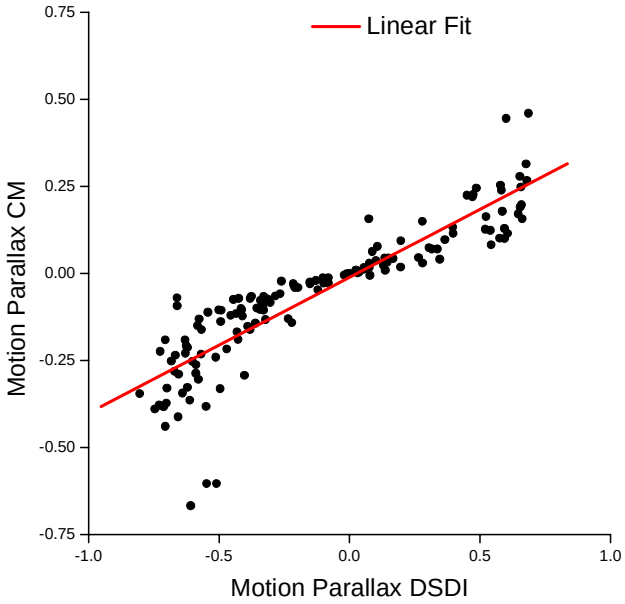

Fig. S7. Alternative metric of depth sign selectivity. To be sure that our conclusions were not peculiar to our particular metric of selectivity, we compared the Depth Sign Discrimination Index (DSDI) to a standard contrast measure (CM) in the Motion Parallax condition. The contrast measure is a conventional tuning metric given by the equation:

$$CM = \frac{1}{4}\sum_{i=1}^{4}\frac{R_{far(i)} - R_{near(i)}}{R_{far(i)} + R_{near(i)}}$$

Note that the DSDI includes a measure of response variability whereas the CM does not. Nevertheless, we find that the CM correlates well ($r=0.89$, $p<0.0001$) with the DSDI. The relationship between the two variables is non-linear but monotonic.