Misperception of motion in depth originates from an incomplete transformation of retinal signals

T. Scott Murdison
Centre for Neuroscience Studies, Queen's University, Kingston, Ontario, Canada
Canadian Action and Perception Network (CAPnet), Toronto, Ontario, Canada
Association for Canadian Neuroinformatics and Computational Neuroscience (CNCN), Kingston, Ontario, Canada

Guillaume Leclercq
ICTEAM and Institute for Neuroscience (IoNS), Université catholique de Louvain, Louvain-La-Neuve, Belgium

Philippe Lefèvre
ICTEAM and Institute for Neuroscience (IoNS), Université catholique de Louvain, Louvain-La-Neuve, Belgium

Gunnar Blohm
Centre for Neuroscience Studies, Queen's University, Kingston, Ontario, Canada
Canadian Action and Perception Network (CAPnet), Toronto, Ontario, Canada
Association for Canadian Neuroinformatics and Computational Neuroscience (CNCN), Kingston, Ontario, Canada

Depth perception requires the use of an internal model of the eye-head geometry to infer distance from binocular retinal images and extraretinal 3D eye-head information, particularly ocular vergence. Similarly, for motion in depth perception, gaze angle is required to correctly interpret the spatial direction of motion from retinal images; however, it is unknown whether the brain can make adequate use of extraretinal version and vergence information to correctly transform binocular retinal motion into 3D spatial coordinates. Here we tested this hypothesis by asking participants to reconstruct the spatial trajectory of an isolated disparity stimulus moving in depth either peri-foveally or peripherally while participants' gaze was oriented at different vergence and version angles. We found large systematic errors in the perceived motion trajectory that reflected an intermediate reference frame between a purely retinal interpretation of binocular retinal motion (not accounting for veridical vergence and version) and the spatially correct motion. We quantify these errors with a 3D reference frame model accounting for target, eye, and head position upon motion percept encoding. This model could capture the behavior well, revealing that participants tended to underestimate their version by up to 17%, overestimate their vergence by up to 22%, and underestimate the overall change in retinal disparity by up to 64%, and that the use of extraretinal information depended on retinal eccentricity. Since such large perceptual errors are not observed in everyday viewing, we suggest that *both* monocular retinal cues *and* binocular extraretinal signals are *required* for accurate real-world motion in depth perception.

Introduction

Stereoscopic vision is crucial for perceiving and acting on objects moving around us in three-dimensional (3D) space. Consider a batter in baseball: To accurately swing at an approaching pitch, the visuo-

Citation: Murdison, T. S., Leclercq, G., Lefèvre, P., & Blohm, G. (2019). Misperception of motion in depth originates from an incomplete transformation of retinal signals. *Journal of Vision*, *19*(12):21, 1–15, https://doi.org/10.1167/19.12.21.

https://doi.org/10.1167/19.12.21 Received May 28, 2018; published October 24, 2019 ISSN 1534-7362

motor system must first estimate the 3D spatial motion of the ball in space from two 2D retinal projections (Batista, Buneo, Snyder, & Andersen, 1999; Blohm & Crawford, 2007; Blohm, Khan, Ren, Schreiber, & Crawford, 2008; Chang, Papadimitriou, & Snyder, 2009). That means the brain has the difficult task of assigning coordinating points on each retina to the moving object and using an internal model of the eye-head geometry to accurately compute its 3D egocentric distance (Blohm et al., 2008; Harris, 2006; Welchman, Harris, & Brenner, 2009). However, exactly which signals are used to extract motion-in-depth from binocular images is unclear.

Part of the confusion comes from an overabundance of available depth cues in typical viewing. Motion-in-depth cues can arise from both retinal and extraretinal sources and can be monocular or binocular. Monocular cues include retinal image features (e.g., shading, texture, defocus blur, perspective, optical expansion, kinetic depth cues, motion parallax, etc.; Guan & Banks, 2016; Held, Cooper, & Banks, 2012; Zannoli, Love, Narain, & Banks, 2016; Zannoli & Mamassian, 2011), and ocular accommodation (Guan & Banks, 2016; Mon-Williams & Tresilian, 2000). Binocular cues include retinal disparity, inter-ocular velocity differences (Nefs & Harris, 2010; Nefs, O'Hare, & Harris, 2010), change in disparity over time (Nefs & Harris, 2010; Nefs et al., 2010), ocular vergence (Mon-Williams & Tresilian, 1999; Mon-Williams, Tresilian, & Roberts, 2000) and version angles (Backus, Banks, Van Ee, & Crowell, 1999; Banks & Backus, 1998). Ultimately, however, because retinal disparity varies nonuniformly with 3D eye-in-head orientation (Blohm et al., 2008), retinal signals alone are insufficient to estimate motion-in-depth; rather, the visual system must account for the full 3D geometry of the eye and head (Blohm et al., 2008; Harris, 2006; Welchman et al., 2009). Indeed, Blohm et al. (2008) demonstrated that the visual system accounts for 3D eye-in-head orientation to accurately reach to static objects in depth, but how this finding extends to *moving* objects in depth is unclear. Harris (2006) and Welchman and colleagues (2009) found psychophysical evidence supporting the use of binocular extraretinal signals (both static and dynamic) for motion in depth perception, but their relative contributions to the spatial 3D percept remain unclear. Here, we attempt to answer this question by asking participants to reconstruct motion-in-depth trajectories from only binocular depth cues across various vergence and horizontal version angles, then use 3D geometric modeling to compare these reconstructions directly to motion-in-depth perception predicted by relative disparity.

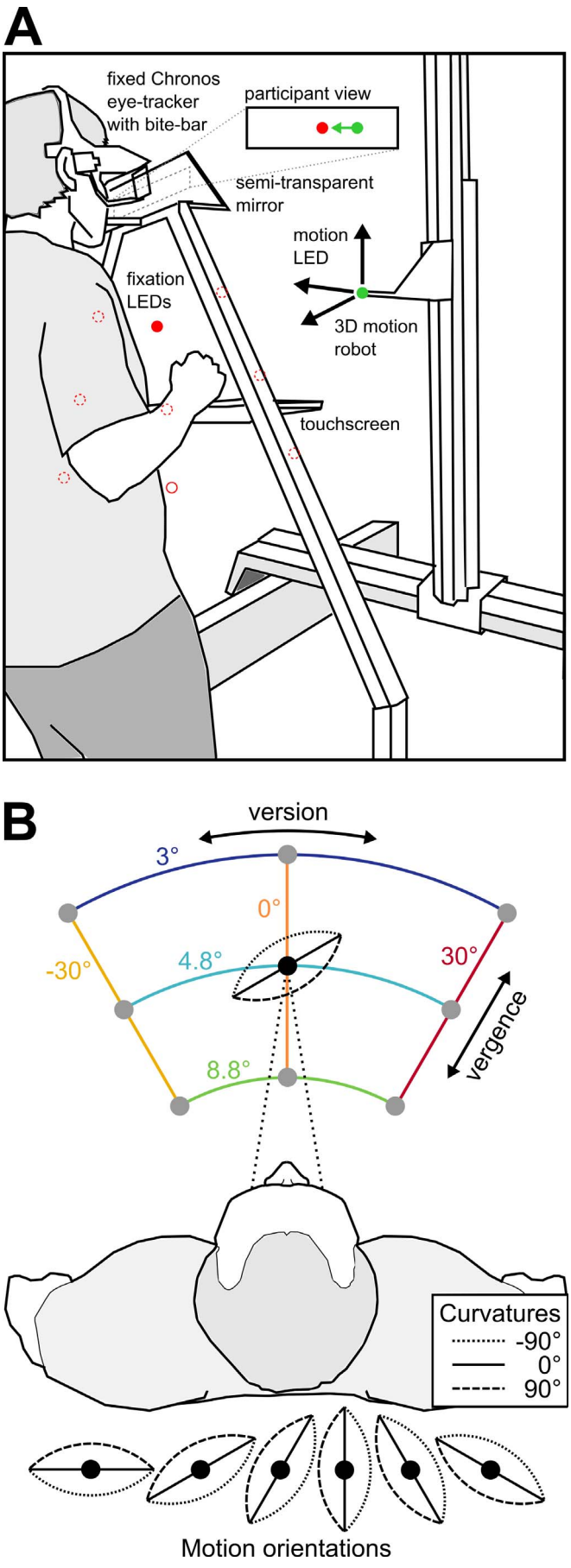
Another open question is how motion-in-depth perception depends on retinal eccentricity. Although the magnitude of relative disparity increases with retinal eccentricity for a given fixation (Blohm et al., 2008), many of the observed disparity-selective cortical cells are tuned for small-magnitude disparities (DeAngelis & Uka, 2003), hinting that binocular signals may play a large role for depth perception near the fovea but not in the periphery. Convincing work from Held et al. (2012) found that position-in-depth is extracted in a complementary way: using mostly binocular disparity signals at the fovea and using mostly defocus blur in the periphery. Rokers et al. (2018) found that a Bayesian model weighting retinal motion signals according purely to sensory noise and 3D viewing geometry could explain a lateral bias (resulting in a compression-like effect) in motion in depth perception for trajectories presented off the midsaggital line (Fulvio, Rosen, & Rokers, 2015). Therefore, understanding the interaction between available motion in depth cues (e.g., changing disparity over time, defocus blur), retinal eccentricity, and binocular geometry, and their contributions to motion in depth perception presents a logical next step.

In this study, we asked participants to reproduce the perceived horizontal depth spatial trajectory of an isolated disparity stimulus observed either foveally or peripherally under different vergence and version angles, with the goal of understanding the relative contribution of binocular eye orientation signals to motion in depth perception for motion at different retinal eccentricities. Using this motion reconstruction task, we found large systematic errors in the perceived motion trajectory. These errors seemed to reflect those predicted by a partial transformation of retinal motion into an intermediate (i.e., spatially incorrect) reference frame resulting from an incomplete accounting of binocular eye orientation signals into the percept. A simple geometric model could capture the behavior well and allowed us to quantify the relative contribution of horizontal version, vergence, and change in retinal disparity to motion-in-depth perception. This model revealed that participants tended to underestimate their version, overestimate their vergence, and underestimate the overall change in retinal disparity in a nonuniform way across the retina—resulting an encoding of 3D spatial motion in neither retinal, nor spatial, but *intermediate* coordinates irrespective of retinal eccentricity. Extended to real world viewing, we infer that motion-in-depth estimation is an eccentricity-dependent process that explicitly requires the use of *both* binocular and monocular depth cues for accuracy.

Materials and methods

Participants

In total, 13 participants (age 22–35 years, 10 male three female) were recruited for two experiments after informed



consent was obtained. Twelve of 13 participants were right-handed and all participants were naïve as to the purpose of the experiment. All participants had normal or corrected-to-normal vision and did not have any known neurological, oculomotor, or visual disorders. We also evaluated participants' stereoscopic vision using the following tests: Bagolini striated glasses test (passed by all participants), Worth's four dot test (passed by all participants), and TNO stereo test. All but participants 1 and 4 could detect disparities ≤60 s of arc. Participant 1's and 4's detection thresholds were 240 and 120 s of arc, respectively; however, despite their higher stereoacuity thresholds, these participants did not behave in any qualitatively different way relative to the rest of the participants during the study. All procedures were approved by the Queen's University Ethics Committee in compliance with the Declaration of Helsinki.

Experimental paradigm

We used a novel 3D motion paradigm to determine how motion-in-depth is perceived across different horizontal version and vergence angles in complete darkness. This paradigm is illustrated in Figure 1. In panel A, we show the physical setup with the array of red light-emitting diodes (LEDs) representing possible fixation targets (FTs; filled red circle represents the sample trial's illuminated FT) and the green LED (filled green circle) representing the motion target (MT), which was attached to the arm of a custom 3D gantry system (Sidac Automated Systems, North York, ON) that was positioned at the same elevation as the eyes and moved within the lateral depth (x-z) plane. All LEDs had a physical diameter of 5 mm. At the end of target motion, participants were instructed to reconstruct the motion of this target using a stylus on the touchscreen in front of them. On each trial, the FT was reflected through a mirror oriented at 45° and positioned at the level of the eyes, such that the participant perceived the FT as located in the same lateral depth plane as the MT. Other key elements in the physical setup included a stationary Chronos C-ETD 3D video-based eye tracker

←

Figure 1. Apparatus and virtual setup. (A) Experimental apparatus, including 3D motion robot with attached MT LED (green), fronto-parallel arc-array of 9 FT LEDs (red), 45° oriented semitransparent mirror, fixed Chronos eye tracker, and touchscreen. For a given trial, one of the FT LEDs is illuminated and reflected at eye-level using the semitransparent mirror. Meanwhile, the motion robot moves the MT LED in the horizontal depth plane also at eye level, creating the participant view shown in the inset. (B) Virtual setup created by the experimental apparatus and tested motion trajectories, with six orientations (30° steps from 0° to 150°) and three curvatures (−90°, 0°, and 90°), with version angles of −30°, 0°, and 30°, and vergence angles of approximately 8.8°, 4.8°, and 3°.

→

(Chronos Vision, Berlin, Germany) with an attached bite-bar for head stabilization to ensure stable fixation on the FT during target motion. This physical arrangement allowed us to present FTs in the MT plane while avoiding physical collisions (panel B) with FTs positioned at nine different locations, spaced in a polar grid. The FTs' spatial positions corresponded to three horizontal version angles, −30°, 0°, and 30°. The FTs were also positioned across three metric distances corresponding to three vergence angles (assuming a nominal interocular distance of 6.5 cm): 42 cm (~8.8° vergence angle), 78 cm (~4.8° vergence angle), and 124 cm (~3° vergence angle). The MTs moved around these points according to 18 different motion trajectories (six orientations spaced equally from 0° to 180°, with three possible curvatures) purely in the lateral depth plane. At the chosen depths, FT LEDs subtended a maximum visual angle of 4 arcmin and a minimum of 1.3 arcmin, while the MT LED subtended a maximum visual angle of 5.5 arcmin and a minimum of 1.1 arcmin. The trajectories of the MTs were scaled with distance such that they subtended the same overall retinal angle, traversing a maximum of 22 cm in depth when centered on the near FTs, 40 cm when centered on the middistance FTs and 65 cm when centered on the far FTs.

Procedure

Participants knelt, supported by the custom apparatus, in complete darkness. Each trial was defined by three phases: (a) fixation, (b) motion observation, and (c) reporting. During the fixation phase (0–500 ms), participants fixated a randomly selected, illuminated FT from the array of nine LEDs. During the motion observation phase (1,500–3,200 ms), participants maintained fixation on the FT while the MT was displaced by the robot. That MT displacement either occurred in the immediate space around the FT (foveal condition) or around the central (nonilluminated) LED while the participant maintained fixation on the FT (peripheral condition). The eccentricity of target motion was random and counterbalanced across trials such that participants could not predict the upcoming eccentricity. Participants were asked to memorize its trajectory in the lateral depth plane. During the reporting phase (3,200 ms through trial end), participants were asked to remove their head from the bite-bar and trace the perceived spatial trajectory using a stylus on a touchscreen, illuminated using a single bright LED for this trial phase only. The light remained on until a response was recorded, and participants were free to restart their trace at any time. They touched the lower right corner of the screen in order to end the current trial, triggering the start of the next trial.

Trajectories were generated with the goal in mind of covering a wide range of potential retinal motion signals at various eccentricities. For each FT, we produced six different orientations of movement, indicating the orientation of the axis along which the target would travel, including 0°, 30°, 60°, 90°, 120°, and 150°. For each of these motion orientations, the direction of motion (toward or away) was chosen randomly and only one motion direction was sampled per participant. Each motion orientation also had three potential curvatures, between −90°, 0°, or +90°, measured as the angle between the apex of the curve and the motion axis at its halfway point. Finally, motion was either presented peri-foveally around the FT (referred to as the "foveal" condition throughout this report) or presented peripherally around any of the eight other possible FTs (referred to as the "peripheral" condition throughout this report).

Trial selection

We recorded a total of (9 fixation targets × 3 curvatures × 6 orientations × 2 motion location conditions) = 324 trials for each participant (324 trials × 13 participants = 4,212 total trials). Each trial type was randomly interleaved throughout 10 blocks, with the same trial order for each participant. Each participant performed all 10 blocks. This allowed us to pool the responses together across conditions and participants for graphical purposes, as there were no within-participant trial repetitions (model fits were performed on individual trajectories). Upon offline analysis, we discovered that one participant consistently failed to perform the reconstruction portion of the task as instructed: The participant drew the motion backwards and we therefore excluded that data from the analysis, leaving 12 participants (3,888 total trials). Although such an inversion of the motion trajectory could be a result of increased sensory noise for this particular participant (Fulvio et al., 2015; Rokers et al., 2018), the behavior was consistent across most, but not all, MTs and motion conditions, thus making their data impossible to interpret using reference frame analysis. Of the remaining 3,888 trials, we examined recorded eye movement data and removed trials containing eye movements or blinks during the motion phase of each trial, leaving 3,869 valid trials for analysis. For offline comparisons between reconstructed trajectories, we normalized all drawn trajectories according to the distance between the start and endpoint of the target motion. We then used linear interpolation to subsample exactly 1,000 points for each drawn trace, equalizing each trial in length.

3D binocular kinematic model

We developed a 3D model of the binocular retina-eye-head geometry to predict how behavioral motion reconstructions might vary across version and ver-

gence angles (Blohm et al., 2008). This model consisted of three primary stages: retinal motion encoding, inverse modeling, and spatial motion decoding. We used the dual quaternion algebraic formulation previously described elsewhere to estimate the complete relationship between retinal geometry, eye geometry, head geometry, and space (Blohm & Crawford, 2007; Blohm et al., 2008; Leclercq, Blohm, & Lefèvre, 2013; Leclercq, Lefèvre, & Blohm, 2013). Except for when otherwise denoted, when we present standard quaternion equations below, we present only the relevant portion of the dual quaternion (rotational or translational), while the other portion of the quaternion represented the null operation. First, we computed the binocular retinal projections of the motion stimulus, given the current eye and head orientations (retinal motion encoding stage). We carried out this step assuming an interocular distance of 6.5 cm and accounting for the binocular orientation of each eye's Listing's plane (Listing's law extended) according to the following set of quaternion equations (Blohm & Crawford, 2007; Leclercq, Lefèvre, et al., 2013):

$$Q_{LP} = \begin{bmatrix} 0 \\ 0 \\ \cos\alpha \\ -\sin\alpha \end{bmatrix} \quad (1)$$

Because we used an assumed, and not measured, interocular distance to compute the retinal model predictions, we carried out additional testing to ensure that they did not change with an over- or underestimation of interocular distance for our tested vergence angles. We found that our retinal model predictions were stable for participants within the expected anatomical ranges, given our testing population (Dodgson, 2004). With the pitch of Listing's plane (α) assumed to be 5° (Blohm & Crawford, 2007), we implemented the binocular extension of Listing's Law using the following relationship between vergence angle and the "saloon door"-like tilt around the vertical axis of each eye's Listing's plane (Blohm et al., 2008):

$$Q_{L2,L} = Q_{LP}\begin{bmatrix} \cos\frac{\theta_v}{2} \\ 0 \\ 0 \\ \sin\frac{\theta_v}{2} \end{bmatrix} \quad (2A)$$

$$Q_{L2,R} = Q_{LP}\begin{bmatrix} \cos\frac{-\theta_v}{2} \\ 0 \\ 0 \\ \sin\frac{-\theta_v}{2} \end{bmatrix} \quad (2B)$$

and

$$\theta_v = \mu_v \times vergence$$

where θ_v represents the tilt around the vertical axis and the Listing's plane tilt gain μ_v is assumed to be 0.25 (Blohm et al., 2008). We then computed the shortest rotation from primary position Q_{L2} to the normalized current eye-centered, head-fixed gaze position G_{EH}:

$$Q_{EH,L} = \left(-\frac{G_{EH,L}}{\|G_{EH,L}\|} \times Q_{L2,L}\right)^{\frac{1}{2}} \quad (3A)$$

$$Q_{EH,R} = \left(-\frac{G_{EH,R}}{\|G_{EH,R}\|} \times Q_{L2,R}\right)^{\frac{1}{2}} \quad (3B)$$

These basic equations allowed us to compute the binocular retinal projections of the spatial motion given their position relative to the eyes. Note that for this experiment we ignored the effects of ocular counter-roll and the torsional VOR on the orientation of Listing's plane, as the head was always maintained upright (Blohm & Lefèvre, 2010). The projection from a cyclopean centered, head-fixed MT p_{CH} with corresponding quaternion representation $Q_{p,CH}$ to each eye was given by the following translation (based on interocular distance, Q^T_{iod}), followed by rotation by the current eye orientation (Q^R_{EH}), resulting in an eye-centered, eye-fixed, target representation Q_p:

$$Q_{p,L} = {Q^R_{EH,L}}^{-1}\left({Q^T_{iod,L}}^{-1} Q_{p,CH} Q^T_{iod,L}\right) Q^R_{EH,L} \quad (4A)$$

$$Q_{p,R} = {Q^R_{EH,R}}^{-1}\left({Q^T_{iod,R}}^{-1} Q_{p,CH} Q^T_{iod,R}\right) Q^R_{EH,R} \quad (4B)$$

We then computed the retinal projection angles from the vector part of each eye's quaternion, p_L and p_R, using the Fick convention (Blohm et al., 2008). This gave us the eye-centered, eye-fixed, target direction of the MT, $D_{MT,EE}$:

$$D_{MT,EE,L} = \begin{bmatrix} \sin(\theta_L)\cos(\phi_L) \\ \cos(\theta_L)\cos(\phi_L) \\ \sin(\phi_L) \end{bmatrix} \quad (5A)$$

$$D_{MT,EE,R} = \begin{bmatrix} \sin(\theta_R)\cos(\phi_R) \\ \cos(\theta_R)\cos(\phi_R) \\ \sin(\phi_R) \end{bmatrix} \quad (5B)$$

where θ and ϕ refer to the horizontal and vertical target projection angles, respectively.

Using this procedure, we computed the retinal projection of the target motion, given some proportion of version and vergence based on version gain (g_{vs}) and vergence gain (g_{vg}). This operation represented the motion trajectory after the eyes had moved to the inverse estimates of version and vergence angles during the encoding of retinal motion (inverse

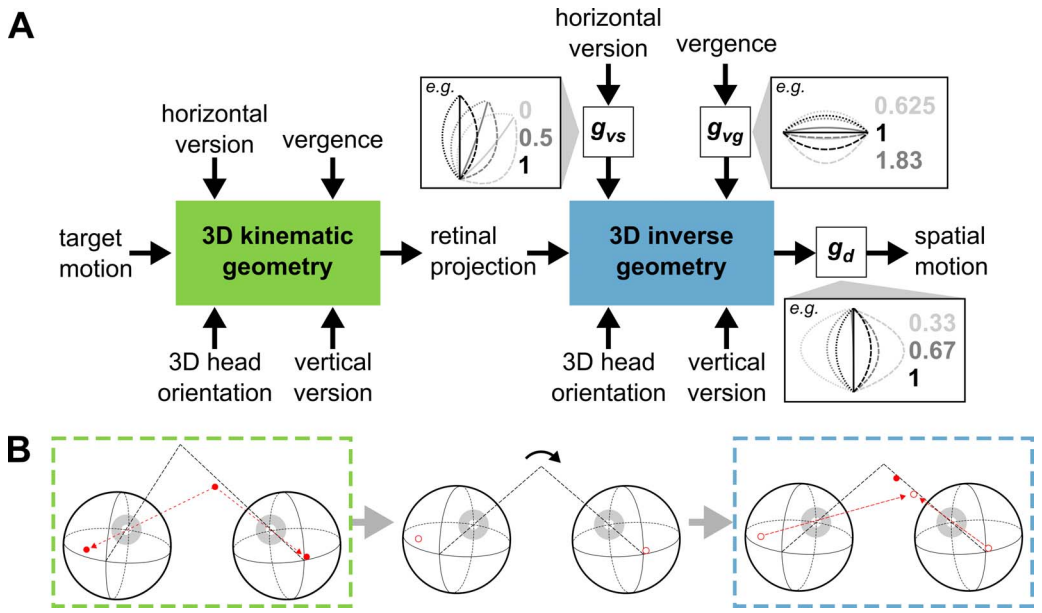


Figure 2. 3D geometrical model. (A) Retinal projection and inverse modeling stages of the 3D binocular kinematic model to generate retinal and partial model predictions. Insets show individual parameter effects on reconstructed traces. The effects of version gain are shown for a fixation version angle of 30; the effects of vergence gain are shown for a fixation vergence of 4.8; the effects of depth gain are shown for trajectories with an orientation of 90. (B) Sample geometrical schematic showing a sample nonspatial prediction for a single point within the motion trajectory. Color-matching dashed boxes represent the retinal projection and inverse geometry stages of the model, respectively. In this example, $g_{vs} = 0$, $g_{vg} = 1$, and $g_d = 1$.

modeling stage). We then back-projected the retinal motion trajectory points into space for each eye by computing the cyclopean-centered, head-fixed transformation using the left and right eye rotation and translation quaternions representing Listing's law and interocular distance, respectively, as described above. We computed the 3D location of the rays' intersection, representing the decoded depth (spatial decoding stage). At this stage, we applied the depth gain (g_d) to the depth component only, representing any lateral compression effects (Fulvio et al., 2015; Rokers et al., 2018). We present this modeling framework in Figure 2 from left to right. In panel A, first we represent the retinal motion encoding stage, which we computed based on the actual geometry of the eyes and head.

This modeling framework allowed us to describe the reconstructed trajectories by varying the contributions of version (g_{vs}) and vergence (g_{vg}) to the inverse model, and motion purely in depth (g_d). Each parameter accounted for a different aspect of the trajectory (shown in Figure 2 insets). To produce the retinal prediction, we set the version gain to 0 and used a constant vergence gain of 1. Importantly, this retinal prediction arbitrarily assumes that vergence is 100% accounted for. Note that, because our model computes the spatial intersection of the binocular back-projections, vergence gains had to be greater than 0 (otherwise the back-projections would be parallel).

Second, we represent the inverse modeling stage where we varied the contributions of extraretinal signals. Third, we represent the spatial motion output stage, where we project the retinal motion, transformed by some proportion of extraretinal signals (depending on the values of g_{vs}, g_{vg}, and g_d), back into space. In Figure 2B, we show a sample transformation for the retinal case where version is unaccounted for, but vergence and depth are accounted for (i.e., $g_{vs} = 0$, $g_{vg} = 1$, and $g_d = 1$).

For each participant, we carried out an 8,000-point nonparametric grid search to minimize the RMSE between the reconstructed trajectories separately for foveal and peripheral motion and the model output, we tested over the full plausible range of parameters (20 linearly spaced values for each parameter). After an initial coarse search to find the expected range of parameters, we used a search space with version gain ranging from 0.5 to 1 with step sizes of 0.03, vergence gain ranging from 0.25 to 1.75 with step sizes of 0.08, and depth gain ranging from 0.2 to 0.8 with step sizes of 0.03. This was followed by a second 512 point least-squares fine fitting method within a ±10% range for g_{vs}, g_{vg}, and g_d around the initialized parameters (eight linearly spaced values for each parameter). We performed this exact optimization procedure separately for each vergence angle to avoid confounding vergence effects. In total, we computed the fits of $3 \times (8{,}000 + 512) = 25{,}536$ total parameter combinations. Given the

Participant	R^2 spatial	R^2 retinal	Version g_{vs}	Vergence g_{vg}	Depth g_d	R^2 model	F (retinal, model)	F (spatial, model)
1								
Foveal	0.33	0.16	0.66	1.51	0.37	0.52	**2.95**	**1.51**
Peripheral	0.28	0.17	0.97	0.97	0.26	0.49	**2.34**	**1.21**
2								
Foveal	0.67	0.29	0.81	1.24	0.74	0.67	**3.51**	0.87
Peripheral	0.60	0.20	0.97	1.04	0.57	0.59	**4.56**	**1.07**
3								
Foveal	0.48	0.27	0.66	1.38	0.63	0.54	**2.77**	**1.02**
Peripheral	0.50	0.25	0.93	1.03	0.49	0.53	**3.38**	1.00*
4								
Foveal	0.49	0.33	0.55	1.30	0.53	0.62	**4.95**	**1.55**
Peripheral	0.42	0.23	1.00	1.03	0.36	0.52	**4.76**	**2.06**
5								
Foveal	0.65	0.23	0.94	1.24	0.63	0.67	**3.68**	**1.04**
Peripheral	0.50	0.25	0.99	0.97	0.44	0.53	**6.50**	0.79
6								
Foveal	0.64	0.27	0.76	1.38	0.81	0.59	**3.34**	**1.25**
Peripheral	0.55	0.24	0.91	0.97	0.56	0.55	**4.08**	**1.10**
7								
Foveal	0.67	0.27	0.84	1.24	0.74	0.66	**5.06**	0.80
Peripheral	0.66	0.23	0.97	1.11	0.61	0.62	**5.27**	0.88
8								
Foveal	0.53	0.33	0.69	1.11	0.64	0.59	**2.92**	**1.10**
Peripheral	0.55	0.22	0.93	1.03	0.56	0.57	**2.88**	**1.37**
9								
Foveal	0.53	0.33	0.63	1.50	0.79	0.64	**3.12**	**1.32**
Peripheral	0.47	0.25	0.80	1.04	0.53	0.56	**3.89**	**1.27**
10								
Foveal	0.47	0.27	0.69	1.51	0.59	0.55	**4.40**	**1.96**
Peripheral	0.33	0.22	0.96	0.97	0.27	0.53	**4.47**	**1.32**
11								
Foveal	0.49	0.20	0.87	1.77	0.71	0.55	**2.81**	**1.34**
Peripheral	0.34	0.17	0.96	0.97	0.31	0.47	**5.43**	**1.09**
12								
Foveal	0.57	0.28	0.69	1.30	0.76	0.63	**2.03**	**1.05**
Peripheral	0.51	0.25	0.97	0.97	0.51	0.51	**3.13**	**1.13**

Table 1. Model parameters and goodness of fit comparisons. *Notes*: Bold *F* statistics represent significantly decreased residuals for the model fit. Asterisk (*) represents a null two-tailed *F*-test result, given a null one-tailed *F*-test result.

number of potential parameter combinations and the complex interactions with the 3D geometry, our optimization strategy was more computationally tractable than more sophisticated error gradient-based methods. This optimization provided parameter estimates that consistently accounted for behavioral variability across participants and motion conditions (see Table 1). To ensure we did not artificially introduce effects between motion conditions, we also carried out this fitting procedure with the full datasets (merged across foveal and peripheral conditions) and found qualitatively identical reference frame results. Custom Matlab scripts for generating model predictions and all data are available on the Open Science Framework website, and can be found via this link: https://osf.io/pvz97/.

Statistical analyses

Group-level statistical tests primarily consisted of two-tailed student t tests. We also performed paired t tests when appropriate for comparing parameters across conditions. We also tested for differences in the residuals for different model outputs using one- and two-tailed F tests for equal variance. The rest of the statistical treatment of the data consisted primarily of computing correlation coefficients and regression analyses.

Results

We sought to determine how visual perception accounts for binocular eye orientation when reconstructing motion in depth from disparity signals. To do this, we designed a novel paradigm in which participants reconstructed motion of an LED in the horizontal depth plane presented either foveally or peripherally on the retina, while fixated in one of nine randomly selected version and vergence orientations. After observing the motion, we then instructed participants to transform their fronto-parallel view into a top-down spatial representation by asking them to reconstruct the motion using a touchscreen positioned in the coronal plane directly in front of them. We then analyzed these reconstructed trajectories to determine how they varied across eye orientation and motion condition. To generate model predictions for the reconstructed signals across changes in version and vergence angles, we developed a 3D model of the binocular eye-head geometry (Figure 2; see Materials and methods for model details). This model allowed us to characterize the eye orientation signals accounted for by the perceptual system.

Reconstructed trajectories deviated from both the spatial (physical) and retinal (see Materials and methods) predictions for both foveal and peripheral motion across all vergence angles. In Figure 3, we show reconstructed motion trajectories for all participants for a single horizontal version angle (30° to the left), a single motion stimulus (orientation 60°, curvature 0°), and for both foveal and peripheral motion conditions. Also in Figure 3, we show the retinal prediction for the motion in each case. These trajectories reveal a qualitative compression effect for motion presented in the retinal periphery (Figure 3, right column) compared to that presented near the fovea (Figure 3, left column). Such qualitative differences are not as readily apparent between vergence angles (Figure 3, rows) given interparticipant variability.

Similarly, we can observe how trajectories change and compare to the retinal and spatial hypotheses for different version angles. We show these trajectories alongside their predictions for three representative motion orientations in Figure 4, averaged across all participants and vergence angles for foveal (top row) and peripheral motion (bottom row). Note that, although these are average trajectories, error bars have been omitted for viewing clarity. These comparisons revealed both an angular rotation between the trajectories during nonzero version as well as a compression of the behavioral traces in the depth dimension across motion orientations; however, these patterns were not consistent for both motion conditions, as only the compression effect was obvious in the peripheral case.

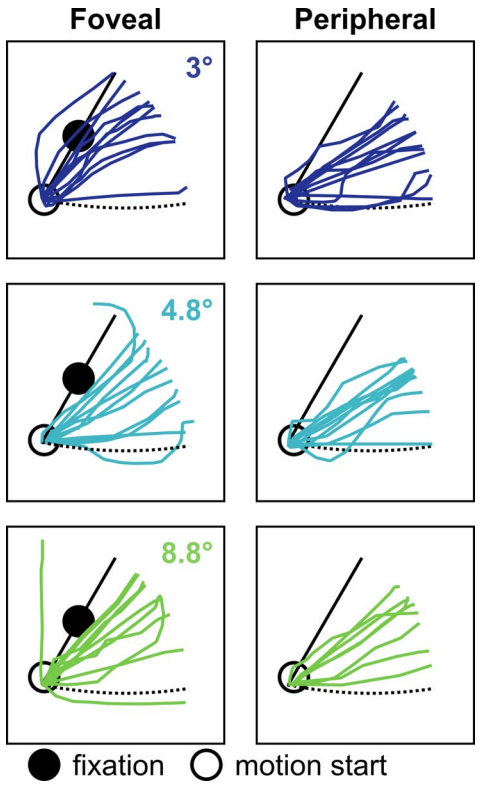


Figure 3. Reconstructed trajectories across vergence, motion conditions. Also shown are the spatial motion prediction (black solid lines) and retinal motion prediction (black dashed lines). Open disks represent motion start, black disks represent fixation relative to the spatial motion (foveal condition only), different colors represent different vergence angles, shown in left column. Motion orientation = 60°, motion curvature = 0°, version angle = −30° (i.e., eyes to the left).

The reconstructed trajectories appear to match neither the spatial nor retinal hypotheses, suggesting that the perceptual system only partially transformed the retinal MT trajectories into spatial coordinates. Importantly, participants were required to transform retinal disparity using the full binocular geometry (and not simply horizontal version, which may be interpreted from Figure 4) in order to correctly perceive the spatial motion. We hypothesize that a failure to fully account for this geometry led to the systematic errors we observe in Figure 4. To capture the extent to which the perceptual system accounted for binocular eye orientations when estimating motion in the lateral depth dimension, we used a two-step nonparametric grid search to optimize the g_{vs}, g_{vg}, and g_d inverse model parameters for the behavioral trajectories, doing separate optimizations for foveally and peripherally presented motion to observe how the parameters depend on retinal eccentricity (see Materials and methods for detailed explanation of optimization procedure). The results of this optimization are shown in Figure 5 at both the single participant level (Figure

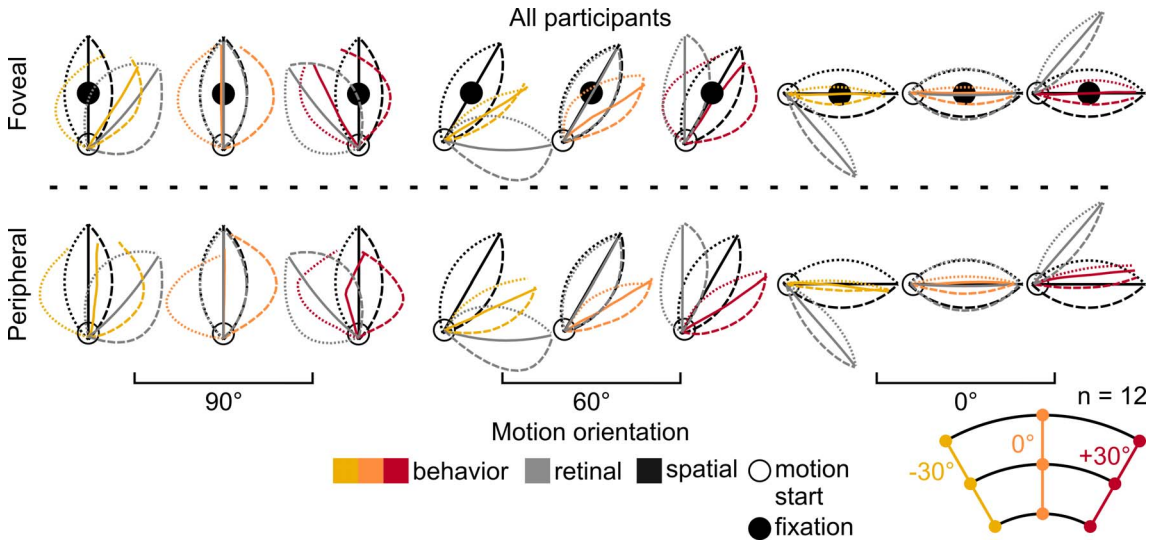


Figure 4. Average across-participant ($n = 12$) foveal and peripheral reconstructed motion, compared with spatial (black) and retinal (gray) predictions for three motion orientations (90°, 60°, and 0°) and all three curvatures (0°, 90°, and −90°) across horizontal version, averaged across vergence. Visible are the more pronounced lateral compression effects and lack of version effects for peripheral motion compared to foveal motion. Note that reconstructed traces were normalized in amplitude to the spatial and retinal predictions, and error bars are omitted for viewing clarity.

5A) and group level (Figure 5B) for both the foveal and peripheral motion conditions.

The parameters optimized for foveal and peripheral motion were distinct, suggesting that motion-in-depth perception varies with retinal eccentricity. For version gain, we found that participants accounted for 83% ± 13% (mean ± *SD*) of horizontal version during foveal motion, compared to 96% ± 10% during peripheral motion: paired *t* test, $t(35) = -5.22$, $p < 0.01$. Given that version compensation during foveal motion was incomplete, the apparent full compensation during peripheral motion could have been the result of the visual system using the retinal location of the stimulus as a cue for current horizontal eye orientation, effectively bypassing an explicit need for extraretinal signals. Next, we found that the foveal depth gains accounted for 54% ± 13% of depth speed and was significantly greater than that for peripheral motion at 36% ± 14%: paired *t* test, $t(35) = 8.70$, $p < 0.01$, indicating that motion in depth was perceived to be faster when foveal. Finally, participants used a foveal vergence gain of 1.22 ± 0.18. In contrast, participants used a significantly smaller (and more accurate) peripheral vergence gain of 0.98 ± 0.15: paired *t* test, $t(35) = 6.30$, $p < 0.0$.

We present the parameter values and the model R^2 values in Table 1. In this table, the computed R^2 values represent the variance accounted for under the spatial and retinal hypotheses for each participant in each motion condition (foveal and peripheral). Also shown are the optimized model gain parameters (g_{vs}, g_{vg}, g_d), the corresponding model R^2 values, and the results of a one-tailed *F* test for equal variance (comparing the model residuals to those for the retinal and spatial hypotheses). In general, the model fit provided good fit to the data, yielding a larger R^2 value than either the retinal or spatial predictions in 41 of 48 possible comparisons. Statistically, the model fit provided a better fit to the data in all retinal comparisons (all *F* statistics significantly greater than 1), and in 19 of 24 spatial comparisons, though in one case we did not detect a significant difference in residual variability (participant 3, peripheral motion condition, spatial hypothesis). To be sure we did not bias these results by splitting motion conditions into the foveal and peripheral conditions, we also carried out the model optimization with the full datasets, merged across motion conditions, and observed qualitatively identical results. These findings suggest that an underestimation of 3D eye orientation signals during the transformation from retinal to spatial coordinates is responsible for observed distortions to motion-in-depth perception.

Taken together, these three fit parameters allowed us to characterize the extent to which the transformation from retinal to spatial coordinates occurred for each participant. We computed a transformation index, I_T, represented by Equation 6:

$$I_T = (D_R - D_S)/(D_R + D_S) \quad (6)$$

where D_R and D_S are the Euclidian distances of each set of gain parameters from the retinal and spatial hypotheses, respectively. For example, a purely spatial set of gain parameters would be represented by $[g_{vs}\ g_{vg}\ g_d] = [1\ 1\ 1]$, corresponding to a $D_S = 0$ and a $D_R = 1$;

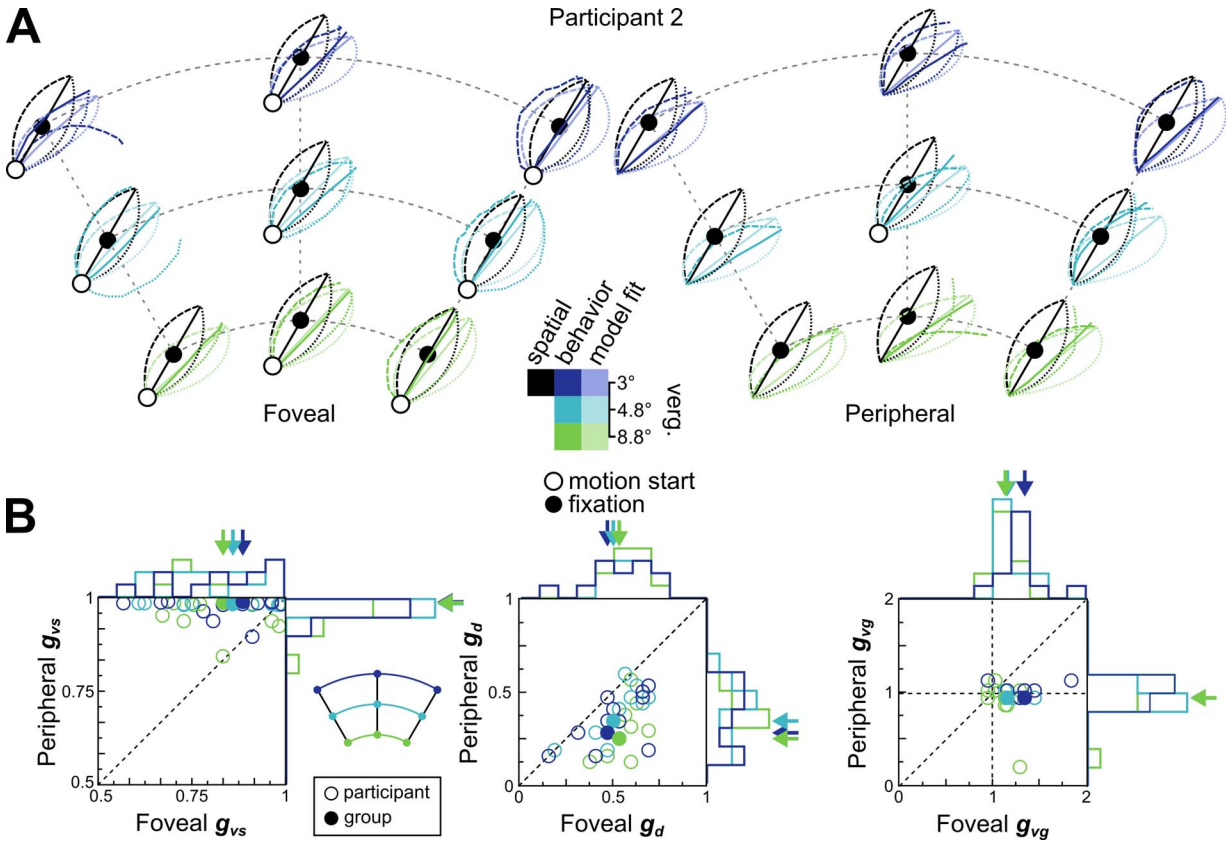


Figure 5. Results of model optimization. (A) Comparison of model outputs (light colored traces) and spatial predictions (black traces) with actual reconstructed trajectories for a single motion orientation (60°) and all curvatures after fitting version gain (g_{vs}), depth gain (g_d), and vergence gain (g_{vg}) parameters separately for each vergence distance and for foveal (left) and peripheral (right) motion conditions, for single participant (#2). Also shown is the spatial motion start position in each motion condition; Note that the foveal condition has fixation and motion start at every FT, but the peripheral condition has only motion start at the central FT, even though participants fixated all nine FTs. (B) Group-level scatter plots showing peripheral versus foveal motion parameter fits for version gain (g_{vs}, left), depth gain (g_d, middle), and vergence gain (g_{vg}, right). Open disks represent participant parameters and solid disks represent group-level parameters fit on all the data. Arrows above histograms represent group-level fit parameter locations along a given axis.

subsequently, $I_T = (1-0)/(1+0) = 1$. By the same logic, for a purely retinal set of gains, $I_T = -1$. We present the distributions of these gain parameters for each participant separated for foveal and peripheral motion, merged across vergence fits, in Figure 6.

I_T was significantly greater than 0 for both foveal: mean ± *SD*, 0.28 ± 0.13; $t(35) = 13.19$, $p < 0.01$, and peripheral motion: mean ± *SD*, 0.29 ± 0.09; $t(35) = 19.13$, $p < 0.01$, suggesting that, in both cases, the average reconstructed trajectory was coded according to an intermediate coordinate frame that was more similar to spatial than to retinal.

Discussion

We asked participants to estimate the motion-in-depth of an isolated disparity stimulus and found large systematic errors that differed depending on viewing eccentricity, viewing angle, and motion trajectory (i.e., orientation, curvature). We found that a simple model of the 3D eye-in-head geometry that used inverse estimates of the ocular version angle, vergence angle, and speed-in-depth could capture the reconstructed trajectories well. For foveal motion, a model that overestimated ocular vergence angle, underestimated ocular version, and target speed fit the perceived trajectories. For natural viewing, this result suggests that additional monocular cues are necessary to accurately estimate foveal motion in depth. Contrastingly, for peripheral motion, a model that accurately estimated eye orientation signals fit the perceived reconstructions, but this model also severely underestimated the change in disparity over time—more than during foveal motion. In this condition, binocular eye orientations may have been inferred using eccentricity. This finding assumes that the visual system has access to the implicit mapping between retinal eccentricity and eye orientation, and that it can use this mapping to

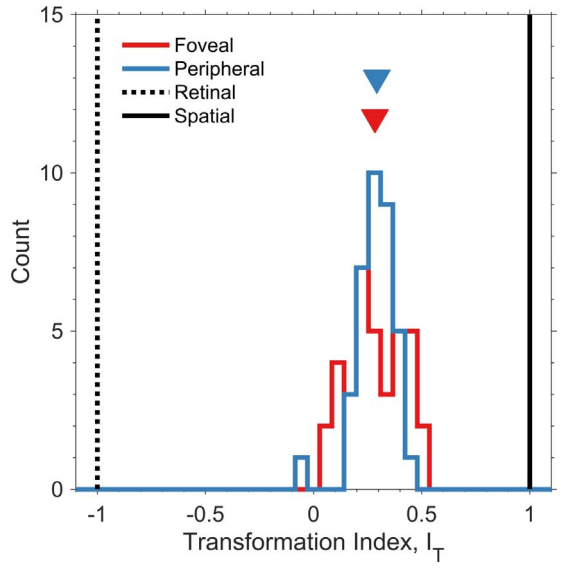


Figure 6. Transformation indices (I_T) for foveal and peripheral motion. Also shown are the retinal (dashed) and spatial (solid) predictions, with means for foveal (red) and peripheral (blue) motion represented by color-matched arrows. Despite their distinct model parameters, a similar intermediate reference frame could represent perceived motion at both retinal eccentricities.

correct for version-generated shifts of the retinal projection. Using a simple transformation index computed from the inverse model fits, we found that spatial misperceptions corresponded to a partial transformation of retinal motion into spatial coordinates, regardless of retinal eccentricity. This evidence supports an explicit requirement for *both* monocular retinal cues and extraretinal signals for an accurate perception of 3D motion, as neither retinal (Blohm et al., 2008) nor extraretinal signals alone can support geometrically correct stereoscopic perception. This claim extends earlier work (Blohm et al., 2008; Harris, 2006; Welchman et al., 2009) with a model which directly quantifies the extraretinal contribution to motion in depth perception.

Such a reliance on monocular retinal cues and/or contextual depth cues at the fovea is understandable given their abundance in natural vision; however, absolute depth cannot be extracted from monocular cues alone (Blohm et al., 2008). Of course, this incapacity rarely impacts typical naturalistic 3D perception when multiple relative depth cues are available. An additional reason why the visual system might rely on monocular cues, despite often being geometrically inaccurate, potentially comes from the idea that estimates of binocular eye orientation are unreliable (Blohm et al., 2008; McGuire & Sabes, 2009) compared to monocular retinal cues (Welchman et al., 2009), and these estimates might become even more variable due to stochastic reference frame transformations (Alikhanian, Carvalho, & Blohm, 2015). The head-to-world-centered coordinate transformation required to extract depth from disparity could be stochastic (i.e., adding uncertainty to the final depth estimate) and the fovea's high spatial acuity means that monocular motion cues could be quite reliable for stereopsis (e.g., Ponce & Born, 2008) This interpretation is consistent with various lines of evidence showing a propensity of the visuomotor system to optimally account for perceptual (Burns, Nashed, & Blohm, 2011) and motor uncertainty (Burns & Blohm, 2010; Schlicht & Schrater, 2007; Sober & Sabes, 2003) resulting from stochastic reference frame transformations (Alikhanian et al., 2015), and is consistent with behavioral evidence from visuomotor updating work (Fiehler, Rösler, & Henriques, 2010; Henriques, Klier, Smith, Lowy, & Crawford, 1998; Medendorp, Goltz, Vilis, & Crawford, 2003; Murdison, Paré-Bingley, & Blohm, 2013). Our setup included an explicit requirement for a reference frame transformation, by design, in asking participants to draw their perception of the top-down motion on a touch screen in front of them. Although this paradigm allowed us to assess the perceptual reference frame, it is likely that this forced transformation resulted in added stochasticity to drawn trajectories (Alikhanian et al., 2015; Burns & Blohm, 2010; Schlicht & Schrater, 2007). Ultimately, our setup did not inform whether this added stochasticity altered the reliability (and therefore the utility) of different sensory and motor signals for perception.

The peripheral motion case presents an apparently paradoxical finding: Eye-in-head orientation can be accurately estimated (likely using retinal eccentricity) whereas target change in disparity over time is significantly underestimated relative to both its spatial motion and its foveal motion. However, we provide *only* a disparity stimulus to the observer regardless of retinal location, and the relative contribution of disparity to depth perception decreases with eccentricity (Held et al., 2012). The observed percept of compressed motion in the periphery is in line with the idea of a lower weighted contribution of disparity cues (Held et al., 2012), while changes in defocus blur of the point stimulus were likely negligible. As well, the tendency of disparity-tuned neurons to disproportionately prefer disparities $<1°$ (DeAngelis & Uka, 2003) is another clue that motion-in-depth at the fovea is represented differently in the visual system than motion-in-depth in the periphery, where disparity magnitudes are much larger (Blohm et al., 2008). In agreement with this idea, psychophysical findings reveal that such a lateral compression could be captured using Bayesian probabilities for visual target motion (Rokers, Fulvio, Pillow, & Cooper, 2018; Welchman, Lam, & Bulthoff, 2008) and due to greater relative uncertainty in the estimate of the depth motion

component for motion in the periphery (Fulvio et al., 2015; Rokers et al., 2018). Determining whether motion-in-depth perception is based on such a statistically optimal combination of disparity, retinal defocus blur and extraretinal cues therefore represents a potential extension of this work.

To isolate for horizontal disparity as the primary cue for depth perception, we removed any contribution of visuomotor feedback by restricting movements of the eyes and head. We determined the role of *static* eye orientation signals in interpreting a *dynamic, moving* stimulus, although in natural viewing our eyes and head are often moving as well. Both disparity and eye movements contribute to depth perception but the precise nature of these contributions, and how they might depend on one another, is unclear. For example, vergence angle corresponds to perceived depth during the kinetic depth effect (Ringach, Hawken, & Shapley, 1996), but artificially inducing disparity changes between correlated (and anticorrelated) random-dot stimuli can cause the eyes to rapidly converge (or diverge) without any perception of depth (Masson, Bussettini, & Miles, 1997). On the neural level, disparity is coded in V1 without a necessary perception of depth (Cumming & Parker, 1997). Psychophysics work has shown that vergence eye movements are beneficial, though not sufficient, for judging the relative depth (Foley & Richards, 1972) and depth motion of stimuli (Harris, 2006; Welchman et al., 2009), but to our knowledge, before this report, no one had quantified the 3D geometric extent to which these signals are used to form a continuous perception of motion in depth.

In addition, by restricting the orientations of the eyes and head we removed feedback due to motion parallax and changes in vertical disparity. Importantly, providing such dynamic visual feedback has been shown to improve motion-in-depth perception in virtual reality (Fulvio & Rokers, 2017). Although vertical disparity naturally varies during normal ocular orienting, we designed our task to keep vertical disparity constant for a given gaze location. This manipulation not only removed vertical disparities due to changes in cyclovergence, but also vertical disparities due to changes in head orientation (Blohm et al., 2008). These natural changes in vertical disparity during eye and head movements likely serve as another informative dynamic cue for judging motion-in-depth under normal viewing contexts. For the above reasons, presenting participants with a dynamic, motion-tracked version of our task could therefore represent an important extension of this work.

From an evolutionary perspective, it is unclear why the visual system would underestimate binocular cues when estimating motion in depth with static gaze. Indeed, in an enriched visual environment, there are often sufficient monocular cues available to the visual system to be able to judge relative depth. During everyday viewing in natural contexts, this is often the case, especially for self-generated motion in depth. On the other hand, our findings suggest that in some special cases without an enriched viewing context such a monocular strategy fails. How would such a monocular strategy work for natural situations in which monocular cues are sparse? And how can we reconcile our findings that binocular gaze is not fully accounted for with these special cases? We posit that spatially correct depth motion in these cases must be a result of motor adaptation, which can come from repeated exposures to the same object physics resulting from the same motor commands or from continuous visual feedback providing a dynamic, predictive cue in the event of changing visual and motor input. To illustrate our point, consider two edge cases: juggling and firefly catching. Expert jugglers learn to keep mostly static fixation near the apex of the balls' trajectories or even sometimes not fixate the balls at all (especially when combining juggling with other tasks, such as balancing). This strategy presumably takes advantage of a learned internal model of the balls' ballistic trajectory (resulting from manual motor commands) combined with various monocular motion cues to intercept each ball. Alternatively, one can consider the case of attempting to catch a firefly in darkness: Fixating while attempting this is intuitively a bad idea because the flight of a firefly is largely unpredictable. Instead, to catch the fly, a better strategy might be to visually track its motion as, for example, amphibians do (Borghuis & Leonardo, 2015). Such a strategy would allow for the use of consistent visuomotor feedback, allowing the construction of a predictive model of the fly's path (Borghuis & Leonardo, 2015). Thus, follow-up experiments investigating the interplay between (a) availability of monocular cues, (b) predictability of object physics, and (c) facilitation from visuomotor learning would be informative of how our brain constructs motion-in-depth percepts.

Conclusions

We quantified the extent to which visual perception accounts for the 3D geometry of the eyes and head when interpreting motion in depth under static viewing conditions. We found that participants underestimated 3D binocular eye orientations, leading to different spatial motion percepts for identical egocentric trajectories. To perceive and successfully navigate through the 3D world, our findings suggest that perception must supplement binocular disparity signals with binocular eye and head orientation estimates, monocular depth

cues, and dynamic visuomotor feedback. It remains to be seen, however, what the precise contributions and relative weightings of each of these cues might be.

Keywords: ocular vergence, horizontal version, depth perception, motion perception, reference frames

Acknowledgments

The authors want to thank our colleagues at the Centre for Neuroscience Studies (CNS) at Queen's University and the Institute of Neuroscience (IoNs) at Université catholique de Louvain for their helpful feedback on this project. This work was supported by NSERC (Canada), CFI (Canada), the Botterell Fund (Queen's University, Kingston, ON, Canada) and ORF (Canada). TSM was also supported by DAAD (Germany) and DFG (Germany). PL was supported by the Belgian Federal Science Policy Office, IAP VII/19 DYSCO, the European Space Agency (ESA), PRODEX C90232.

Commercial relationships: none.
Corresponding author: T. Scott Murdison.
Email: smurdison@fb.com.
Address: Facebook Reality Labs, Redmond, WA, USA.

References

Alikhanian, H., Carvalho, S. R., & Blohm, G. (2015). Quantifying effects of stochasticity in reference frame transformations on posterior distributions. *Frontiers in Computational Neuroscience*, *9*(July), 1–9, https://doi.org/10.3389/fncom.2015.00082.

Backus, B. T., Banks, M. S., Van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, *39*(6), 1143–1170, https://doi.org/10.1016/S0042-6989(98)00139-4.

Banks, M. S., & Backus, B. T. (1998). Extra-retinal and perspective cues cause the small range of the induced effect. *Vision Research*, *38*(2), 187–194, https://doi.org/10.1016/S0042-6989(97)00179-X.

Batista, A. P., Buneo, C. A., Snyder, L. H., & Andersen, R. A. (1999, September 1). Reach plans in eye-centered coordinates. *Science*, *285*(5425), 257–260. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10398603

Blohm, G., & Crawford, J. D. (2007). Computations for geometrically accurate visually guided reaching in 3-D space. *Journal of Vision*, *7*(5):4, 1–22, https://doi.org/10.1167/7.5.4. [PubMed] [Article]

Blohm, G., Khan, A. Z., Ren, L., Schreiber, K. M., & Crawford, J. D. (2008). Depth estimation from retinal disparity requires eye and head orientation signals. *Journal of Vision*, *8*(16):3, 1–23, https://doi.org/10.1167/8.16.3. [PubMed] [Article]

Blohm, G., & Lefèvre, P. (2010). Visuomotor velocity transformations for smooth pursuit eye movements. *Journal of Neurophysiology*, *104*(4), 2103–2115, https://doi.org/10.1152/jn.00728.2009. Smooth.

Borghuis, B. G., & Leonardo, A. (2015). The role of motion extrapolation in amphibian prey capture. *Journal of Neuroscience*, *35*(46), 15430–15441, https://doi.org/10.1523/jneurosci.3189-15.2015.

Burns, J. K., & Blohm, G. (2010). Multi-sensory weights depend on contextual noise in reference frame transformations. *Frontiers in Human Neuroscience*, *4*(December), 1–15, https://doi.org/10.3389/fnhum.2010.00221.

Burns, J. K., Nashed, J. Y., & Blohm, G. (2011). Head roll influences perceived hand position. *Journal of Vision*, *11*(9):3, 1–9, https://doi.org/10.1167/11.9.3. [PubMed] [Article]

Chang, S. W. C., Papadimitriou, C., & Snyder, L. H. (2009). Using a compound gain field to compute a reach plan. *Neuron*, *64*(5), 744–755, https://doi.org/10.1016/j.neuron.2009.11.005.

Cumming, B. G., & Parker, A. J. (1997, September 18). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, *389*(6648), 280–283, https://doi.org/10.1038/38487.

DeAngelis, G. C., & Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *Journal of Neurophysiology*, *89*(2), 1094–1111, https://doi.org/10.1152/jn.00717.2002.

Dodgson, N. A. (2004). Variation and extrema of human interpupillary distance. In A. J. Woods, J. O. Merritt, S. A. Benton, & M. T. Bolas (Eds.), *Stereoscopic displays and virtual reality systems XI* (Vol. 5291, pp. 36–46). Bellingham, WA: SPIE. Retrieved from https://doi.org/10.1117/12.529999

Fiehler, K., Rösler, F., & Henriques, D. Y. P. (2010). Interaction between gaze and visual and proprioceptive position judgements. *Experimental Brain Research*, *203*(3), 485–498, https://doi.org/10.1007/s00221-010-2251-1.

Foley, J. M., & Richards, W. (1972). Effects of voluntary eye movement and convergence on the binocular appreciation of depth. *Perception &*

Psychophysics, *11*(6), 423–427, https://doi.org/10.3758/BF03206284.

Fulvio, J. M., & Rokers, B. (2017). Use of cues in virtual reality depends on visual feedback. *Scientific Reports*, *7*(1):16009, https://doi.org/10.1038/s41598-017-16161-3.

Fulvio, J. M., Rosen, M. L., & Rokers, B. (2015). Sensory uncertainty leads to systematic misperception of the direction of motion in depth. *Attention, Perception, and Psychophysics*, *77*(5), 1685–1696, https://doi.org/10.3758/s13414-015-0881-x.

Guan, P., & Banks, M. S. (2016). Stereoscopic depth constancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(20150253), 1–15, https://doi.org/10.1098/rstb.2015.0253.

Harris, J. M. (2006). The interaction of eye movements and retinal signals during the perception of 3-D motion direction. *Journal of Vision*, *6*(8):2, 777–790, https://doi.org/10.1167/6.8.2. [PubMed] [Article]

Held, R. T., Cooper, E. A., & Banks, M. S. (2012). Blur and disparity are complementary cues to depth. *Current Biology*, *22*(5), 426–431, https://doi.org/10.1016/j.cub.2012.01.033.

Henriques, D. Y. P., Klier, E. M., Smith, M. A., Lowy, D., & Crawford, J. D. (1998). Gaze-centered remapping of remembered visual space in an open-loop pointing task. *The Journal of Neuroscience*, *18*(4), 1583–1594.

Leclercq, G., Blohm, G., & Lefèvre, P. (2013). Accounting for direction and speed of eye motion in planning visually guided manual tracking. *Journal of Neurophysiology*, *110*(8), 1945–1957, https://doi.org/10.1152/jn.00130.2013.

Leclercq, G., Lefèvre, P., & Blohm, G. (2013). 3D kinematics using dual quaternions: Theory and applications in neuroscience. *Frontiers in Behavioral Neuroscience*, *7*(February):7, https://doi.org/10.3389/fnbeh.2013.00007.

Masson, G. S., Bussettini, C., & Miles, F. A. (1997, September 18). Vergence eye movements in response to binocular disparity without the perception of depth. *Nature*, *389*, 283–286.

McGuire, L. M. M., & Sabes, P. N. (2009). Sensory transformations and the use of multiple reference frames for reach planning. *Nature Neuroscience*, *12*(8), 1056–1061, https://doi.org/10.1038/nn.2357. Sensory.

Medendorp, W. P., Goltz, H. C., Vilis, T., & Crawford, J. D. (2003). Eye-centered remapping of remembered visual space in human parietal cortex. *Journal of Vision*, *3*(9): 125, https://doi.org/10.1167/3.9.125. [Abstract]

Mon-Williams, M., & Tresilian, J. R. (1999). Some recent studies on the extraretinal contribution to distance perception. *Perception*, *28*(2), 167–181, https://doi.org/10.1068/p2737.

Mon-Williams, M., & Tresilian, J. R. (2000). Ordinal depth information from accommodation? *Ergonomics*, *43*(3), 391–404, https://doi.org/10.1080/001401300184486.

Mon-Williams, M., Tresilian, J. R., & Roberts, A. (2000). Vergence provides veridical depth perception from horizontal retinal image disparities. *Experimental Brain Research*, *133*(3), 407–413, https://doi.org/10.1007/s002210000410.

Murdison, T. S., Paré-Bingley, C. A., & Blohm, G. (2013). Evidence for a retinal velocity memory underlying the direction of anticipatory smooth pursuit eye movements. *Journal of Neurophysiology*, *110*, 732–747, https://doi.org/10.1152/jn.00991.2012.

Nefs, H. T., & Harris, J. M. (2010). What visual information is used for stereoscopic depth displacement discrimination? *Perception*, *39*(6), 727–744, https://doi.org/10.1068/p6284.

Nefs, H. T., O'Hare, L., & Harris, J. M. (2010). Two independent mechanisms for motion-in-depth perception: Evidence from individual differences. *Frontiers in Psychology*, *1*(October), 1–8, https://doi.org/10.3389/fpsyg.2010.00155.

Ponce, C. R., & Born, R. T. (2008). Stereopsis. *Current Biology: CB*, *18*(18), R845–R850, https://doi.org/10.1016/j.cub.2008.07.006.

Ringach, D. L., Hawken, M. J., & Shapley, R. (1996). Binocular eye movements caused by the perception of three-dimensional structure from motion. *Vision Research*, *36*(10), 1479–1492, https://doi.org/10.1016/0042-6989(95)00285-5.

Rokers, B., Fulvio, J. M., Pillow, J. W., & Cooper, E. A. (2018). Systematic misperceptions of 3-D motion explained by Bayesian inference. *Journal of Vision*, *18*(3):23, 1–23, https://doi.org/10.1167/18.3.23. [PubMed] [Article]

Schlicht, E. J., & Schrater, P. R. (2007). Impact of coordinate transformation uncertainty on human sensorimotor control. *Journal of Neurophysiology*, *97*, 4203–4214, https://doi.org/10.1152/jn.00160.2007.

Sober, S. J., & Sabes, P. N. (2003). Multisensory integration during motor planning. *Journal of Neuroscience*, *23*(18), 6982–6992.

Welchman, A. E., Harris, J. M., & Brenner, E. (2009). Extra-retinal signals support the estimation of 3D motion. *Vision Research*, *49*(7), 782–789, https://doi.org/10.1016/j.visres.2009.02.014.

Welchman, A. E., Lam, J. M., & Bulthoff, H. H. (2008). Bayesian motion estimation accounts for a surprising bias in 3D vision. *Proceedings of the National Academy of Sciences*, *USA*, *105*(33), 12087–12092, https://doi.org/10.1073/pnas.0804378105.

Zannoli, M., Love, G. D., Narain, R., & Banks, M. S. (2016). Blur and the perception of depth at occlusions. *Journal of Vision*, *16*(6):17, 1–25, https://doi.org/10.1167/16.6.17. [PubMed] [Article]

Zannoli, M., & Mamassian, P. (2011). The role of transparency in da Vinci stereopsis. *Vision Research*, *51*(20), 2186–2197, https://doi.org/10.1016/j.visres.2011.08.014.